Impact of counterfactual emotions on the experience of algorithm aversion

Ph.D. Thesis

Andrea Beretta

XXXII cycle

Department of Psychology and Cognitive Science, DiPSCo

University of Trento

Supervisor

Prof. Massimo Zancanaro

Co-Supervisor

Dr. Bruno Lepri

2016/2019

UNIVERSITY OF TRENTO - Italy

"We must design for the way people behave,

not for how we would wish them to behave."

— Donald A. Norman, Living with Complexity

Table of Contents

LIST OF FIGURES .... **VII**

LIST OF TABLES .... **IX**

ABSTRACT .... **XI**

GENERAL INTRODUCTION .... **XIII**

CHAPTER 1: LITERATURE REVIEW **19**

1. The pervasive role of algorithms 19
2. The importance of algorithm-based applications 19
3. Algorithm aversion: distrust in algorithmic suggestions 20
4. Algorithm appreciation: reliance on algorithms 21
5. Reliance on complex and opaque algorithms 22
6. Trust violations: blame in human-algorithm interaction 24
7. Emotions related to wrong decisions: Regret and Disappointment 26
8. Limitations of existing evidence 28

8.1. Negative Outcomes 28

8.2. The interest in maintaining a Human-In-The-Loop 28

CHAPTER 2: DO PEOPLE FEEL MORE REGRET FOLLOWING A WRONG SUGGESTION FROM AN ALGORITHM OR OTHER PEOPLE? **29**

1. Motivation and aims 29
2. Hypotheses 30
3. Equipment 31
4. Methods 31

4.1. Experimental design 32

4.2. Setting and Sample 32

4.3. Experimental Procedure 33

4.4. Materials 34

5. Results 36

5.1. Descriptive statistics 36

5.2. The role of trust on the suggestion on the dependent variables 40

5.3. Hierarchical analysis of variance for Linear Mixed-Effects Models 46

5.4. Summary of Linear Mixed-Effects Models analysis 51

6. Discussion and Limitations 52

CHAPTER 3: REGRET WITH INTELLIGENT SYSTEMS PREDICTIONS **56**

1. Motivation and aims 56

2. HYPOTHESES........57
3. EQUIPMENT........57
4. METHOD........58
4.1. Experimental Design........58
4.2. Sample........58
4.3. Experimental Procedure........59
4.4. Materials........60
5. RESULTS........61
5.1. Descriptive statistics........61
5.2. Removed outliers........71
5.3. Summary of results........79
6. DISCUSSION AND LIMITATIONS........79

CHAPTER 4: COUNTERFACTUAL EMOTIONS AND "ALGORITHM APPRECIATION."........81

1. MOTIVATION AND AIMS........81
2. HYPOTHESES........81
3. EQUIPMENT........82
4. METHOD........83
4.1. Experimental Design........83
4.2. Sample........83
4.3. Experimental procedure........84
4.4. Materials........85
5. RESULTS........87
5.1. Descriptive Statistics........87
5.2. Replication of Logg's results........94
5.3. Results of emotions on choice........95
5.4. Results of perception of suggestion........96
5.5. Results on acceptance of the suggestion........97
6. DISCUSSION AND LIMITATIONS........97

CHAPTER 5: FINAL DISCUSSION AND CONCLUSION........100

1. GENERAL DISCUSSION........100
2. SUMMARY OF RESULTS........103
3. IMPLICATIONS AND CONTRIBUTIONS........104
4. FUTURE DIRECTIONS........105
5. CONCLUSION........106

BIBLIOGRAPHY........108

List of Figures

Figure 1 Distribution of trust in the two conditions..38

Figure 2 Boxplots of the counterfactual emotions (regret on the left) and the two conditions (human and computer). Lower and upper box boundaries 25th and 75th percentiles........................39

Figure 3 Boxplots of the attribution of responsibility (internal on the left) and the two conditions (human and computer). Lower and upper box boundaries 25th and 75th percentiles........................40

Figure 4 The effect of suggestion on regret in the first task..42

Figure 5 The effect of suggestion on regret in the second task..42

Figure 6 The effect of suggestion on disappointment in the first task..43

Figure 7 The effect of suggestion on disappointment in the second task..43

Figure 8 The effect of suggestion on internal attribution in the first task..44

Figure 9 The effect of suggestion on internal attribution in the second task..44

Figure 10 The effect of suggestion on external attribution in the first task..45

Figure 11 The effect of suggestion on external attribution in the second task..45

Figure 12 Linear mixed models results of regret scores..48

Figure 13 Linear mixed models results of disappointment scores..49

Figure 14 Linear mixed models results of internal attribution scores..50

Figure 15 Linear mixed models results of external attribution scores..51

Figure 16 Total scores in the first task in human and computer condition by the dimensions of responsibility, guilt, regret, disappointment (N=70)..63

Figure 17 Total scores in the second task in human and computer condition by the dimensions of responsibility, guilt, regret, disappointment (N=70)..64

Figure 18 Scores in the first task in human and computer condition of the dimensions of regret and disappointment (N=70)..65

Figure 19 Scores in the second task in human and computer condition of the dimensions of regret and disappointment (N=70)..66

Figure 20 Scores in the first task in human and computer condition of the dimensions of responsibility and guilt (N=70)..67

Figure 21 Scores in the second task in human and computer condition of the dimensions of responsibility and guilt (N=70)..68

Figure 22 Scores in both tasks in human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on choice (N=70)..69

Figure 23 Scores in both tasks in human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on outcome (N=70)..70

Figure 24 Scores in the first task human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on choice (N=68)..71

Figure 25 Scores in the second task human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on choice 72
Figure 26 Scores in the first task human and computer condition of the dimensions of regret and disappointment............73
Figure 27 Scores in the first task human and computer condition of the dimensions of regret and disappointment............74
Figure 28 Scores in the first task human and computer condition of the dimensions of responsibility and guilt............75
Figure 29 Scores in the second task human and computer condition of the dimensions of responsibility and guilt............76
Figure 30 Total scores on choice in human and computer conditions of the dimensions of responsibility, guilt, regret, and disappointment............77
Figure 31 Total scores on outcome in human and computer conditions of the dimensions of responsibility, guilt, regret and disappointment............78
Figure 32 Boxplots of Logg's replicated results............94
Figure 33 Reliance on the suggestion between the human condition and computer condition95

List of Tables

Table 1 Means (and standard deviations) of the four dimensions in RDS scale scores and trust by gender, age group, locus of control, personality traits, objective numeracy score. 37

Table 2 distribution of trust in the two conditions with percentages 38

Table 3 Means and standard deviation on regret between conditions 39

Table 4 Means and standard deviation on disappointment between conditions 39

Table 5 Means and standard deviation on internal attribution between conditions 40

Table 6 Means and standard deviation on external attribution between conditions 40

Table 7 Compared linear mixed models explanatory analysis models of regret 48

Table 8 Compared linear mixed models explanatory analysis models of disappointment 49

Table 9 Compared linear mixed models explanatory analysis models of internal attribution....50

Table 10 Compared linear mixed models explanatory analysis models of external attribution .51

Table 11 Means (and standard deviations) of the four dimensions in RDS scale scores and trust by gender, age group, locus of control, personality traits, objective numeracy score. 62

Table 12 Summary descriptive statistics on dependent variables in the first task in human and computer conditions (N=70) 63

Table 13 Summary descriptive statistics on dependent variables in the second task in human and computer conditions (N=70) 64

Table 14 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the first task in human and computer conditions (N=70) 65

Table 15 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the second task in human and computer conditions (N=70) 66

Table 16 Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the first task in human and computer conditions (N=70) 67

Table 17 Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the second task in human and computer conditions (N=70) 68

Table 18 Summary descriptive statistics of DVs on choice in human and computer conditions (N=70) 69

Table 19 Summary descriptive statistics of DVs on outcome in human and computer conditions (N=70) 70

Table 20 Summary descriptive statistics on dependent variables in the first task in human and computer conditions 71

Table 21 Summary descriptive statistics on dependent variables in the second task in human and computer conditions 72

Table 22 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the first task in human and computer conditions 73

Table 23 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the second task in human and computer conditions (N=67)....74

Table 24 Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the first task in human and computer conditions....75

Table 25: Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the second task in human and computer conditions (N=70)....76

Table 26 Summary descriptive statistics of DVs on choice in human and computer conditions (N=70)....77

Table 27 Summary descriptive statistics of DVs on outcome in human and computer conditions (N=70)....78

Table 28 Pearson's Product-Moment Correlation Table of variables of experienced emotions on outcome....88

Table 29 Pearson's Product-Moment Correlation Table of variables of experienced emotions on the source of suggestion....89

Table 30 Pearson's Product-Moment Correlation Table of variables of experienced emotions on the source of suggestion and participants' emotions on the wrong outcome....90

Table 31 Pearson's Product-Moment Correlation Table of Acceptance measures, confidence, reliance on suggestion and on self, objective numeracy scores and experienced emotions on outcome....91

Table 32 Pearson's Product-Moment Correlation Table of Acceptance measures, confidence, reliance on suggestion and on self, objective numeracy scores and experienced emotions on the source of suggestion....92

Table 33 Pearson's Product-Moment Correlation Table of Acceptance measures, confidence, reliance on suggestion and on self, objective numeracy scores....93

Abstract

Today more and more algorithms and their applications are entering into the everyday life of each of us. Algorithms can help people to make more effective choices through historical data analysis, generating predictions to present to the user in the form of advice and suggestions. Given the increasing popularity of these suggestions, a greater understanding of how people could increase their judgment through the suggestions presented is needed, in order to improve the interface design of these applications.

Since the envision of Artificial Intelligence (AI), technical progress has the intent of surpassing human performance and abilities (Crandall et al., 2018). Less consideration has been given to improve cooperative relationships between human agents and computer agents during decision tasks.

No study up to date has investigated the negative emotions that could arise from a bad outcome after following the suggestion given by an intelligent system, and how to cope with the potential distrust that could affect the long-term use of the system.

According to Zeelenberg et al. (Martinez & Zeelenberg, 2015; Martinez, Zeelenberg, & Rijsman, 2011a; Zeelenberg & Pieters, 1999), there are two emotions strongly related to wrong decisions, regret, and disappointment. The objective of this research is to understand the different effects of disappointment and regret on participants' behavioral responses to failed suggestions given by algorithm-based systems.

The research investigates how people deal with a computer suggestion that brings to a not satisfying result, compared to a human suggestion. To achieve this purpose, three different scenarios were tested in three different experiments.

In the first experiment, the comparison was amongst two wrong suggestions in a between-subjects design through the presentation of a flight ticket scenario with two tasks. The first study analyzed exploratory models that explain the involvement of the source of suggestion and the trust in the systems in the experience of counterfactual emotions and responsibility attribution.

The second experiment takes advantage of a typical purchase scenario, already used in the psychological literature, which had the aim to solve the issues found in the first study and test the algorithm aversion paradigm through the lenses of a classic study of regret literature. Results showed that, contrary to early predictions, people blame more the source of the suggestion when it comes from a human as compared with an intelligent computer suggestion.

The third study had the aim to understand the role of counterfactuals through a paradigmatic experiment from algorithm aversion literature. In this study, the main finding is about the reliance people have on the algorithmic suggestion, which is higher compared to the reliance they have with a human

suggestion. Nevertheless, people felt more guilt when they had a wrong outcome with a computer compared with a suggestion given by a person.

Results are relevant in order to better understand how people decide and trust algorithm-based systems after a wrong outcome. This thesis is the first attempt to understand this algorithm aversion from the experienced counterfactual emotions and their different behavioral consequences. However, some of these findings showed contradictory results in the three experiments; this could be due to the different scenarios and participants' thoughts and perceptions of artificial intelligence-based systems. From this work, three suggestions can be inferred to help designers of intelligent systems. The first regards the effective involvement of counterfactuals during the user interaction with a wrong outcome and the potential behavioral consequences that could affect the future use of the intelligent system. The second suggestion is the contribution to the importance of the context in which decisions are made, and the third guideline suggests the designer rethink about anthropomorphism as the best practice to present suggestions in the occurrence of potential wrong outcomes.

Future works will investigate, in a more detailed way the perceptions of users and test different scenarios and decision domains.

General Introduction

In everyday life, people make innumerable choices with the help of algorithms or algorithmic systems. These systems can help them decide on which place to eat, or inform what is the shortest road to go from a point A to a point B; or even more, suggest them what music to listen or what movie to watch (from Yelp, Google maps, to Spotify and Netflix respectively). To make decisions more effective, decision-makers rely more and more on algorithms or algorithm-based systems to inform their decisions.

A considerable amount of research has shown how algorithms produce more accurate predictions and better suggestions compared to human experts (Dawes, Faust, & Meehl, 1989; Dawes, 1979; Meehl, 1954), and how people decide with an algorithm (Kleinberg et al., 2017). However, human decision-makers are often reluctant to use these recommendations and prefer human advice instead of algorithms' (Önkal, Goodwin, Thomson, Gönül, & Pollock, 2009). This phenomenon is what is called *algorithm aversion (*Dietvorst, Simmons, & Massey*, 2015).*

This discounting of the advice from an "intelligent machine," according to the literature (Dawes, 1979; Dietvorst et al., 2014), can arise from the expectation of perfection that people have towards algorithms and intelligent machines (Madhavan & Wiegmann, 2007). Nevertheless, no algorithm is perfect due to the probabilistic nature of the suggestion based on data, and people's intolerance of algorithm imperfection could lead to the abandonment of these systems. This mismatch between the user's expectation of perfection and the actual outcomes is one of the reasons for algorithm aversion (Dietvorst et al., 2015). The core idea of this research is that the distrust after seeing an algorithm errs can be the result of the emotional aspects related to a non-optimal decision outcome.

In particular, this work focused on the emotional experience of "it would have been better if…" or what is called "counterfactual thinking" (Zeelenberg, Van Dijk, Van Der Pligt, et al., 1998). Counterfactual thinking defines the comparison between the actual outcome and the outcome one could have "if only" she/he chose differently (Byrne, 2016). The related cognitive-based counterfactual emotions that stem from these comparisons are regret and disappointment (Giorgetta et al., 2013). These feelings are different for the attribution of responsibility for a wrong decision, or whom people blame for the wrong outcome, and their behavioral consequences (Bougie, Pieters, & Zeelenberg, 2003; Giorgetta, Zeelenberg, Ferlazzo, & D'Olimpio, 2012; Zeelenberg & Pieters, 1999). Disappointment is more related to external attribution of responsibility and does not have aversive behavioral consequences(Martinez & Zeelenberg, 2015). Regret, instead, is connected to self-attribution, and it results in switching the behavior that led to a wrong outcome (i.e., do not trust again in a suggestion that makes us made something wrong) (Zeelenberg & Pieters, 1999). Due to the consequences of these emotions, it is of paramount relevance to understanding how counterfactual thoughts can affect the so-called algorithm aversion.

This behavioral change can be due to the emotions people feel after a wrong outcome following a machine suggestion compared to human advice. While if one follows a suggestion given by another person, the decision-maker can blame the other for the adverse outcome. The core idea of this research is that it is more difficult to blame a machine then the self, and this could lead to the experience of regret and the following behavioral switch. In short, the presence of regret can clarify how and why decision-makers discount the advice from an algorithm when they experience that is not perfect.

Even if some researchers studied the different reliance of people in algorithmic suggestion compared to human suggestion (Logg, 2018; Prahl & Van Swol, 2017), less attention has been given to the emotional aspects, the attribution of responsibility and the counterfactual thoughts one can experience.

The initial curiosity that inspired this work needs to be found in the idea of relating the behavioral switch caused by regret to the phenomenon of algorithm aversion and the reaction users have after seen an algorithm errs.

This research investigates the users' emotional reactions on the bad outcome after following a wrong suggestion given by an algorithm compared to the wrong suggestion given by other humans. The primary purpose is to understand how counterfactual thoughts can affect the future use of intelligent systems. The main research question that leads this dissertation is: do counterfactual emotions explain algorithm aversion? This document proposes to explore that question at several depths.

The contribution of this research is to investigate and analyze the fundamental role of counterfactual emotions in decision-making, comparing a human advisor to an algorithm advisor when the suggestion leads to a negative or unexpected outcome. The study of counterfactuals during a decision with a computer has usually been used as a control condition (e.g., Giorgetta et al., 2012; Zeelenberg, Van Dijk, & Manstead, 1998) while in this work it is the treatment condition through the assessment of regret and disappointment during the use of an intelligent system that gives a wrong suggestion.

Three experiments examined how people feel after a wrong suggestion given from an algorithm compared to a wrong suggestion given by a human. All three studies manipulate the source of suggestion (human suggestion vs. computer suggestion). All experiments manipulate the outcome of the decision, which is always wrong, to observe the experience of counterfactual emotions.

As the purpose of this work is to merge psychological aspects to decision with algorithms, these three studies are organized into two main categories. The first two experiments have their basis in the psychological literature of regret (e.g., Giorgetta et al., 2012), presenting a scenario to the participants related to the counterfactual experience. The third study is related to the algorithm aversion literature and Judge-Advisor System (JAS) (Logg, Minson, & Moore, 2019; Prahl & Van Swol, 2017), in these experiments the participants have to give two estimates, before and after a suggestion with implicit data on reliance on the suggestion and explicit measures to undercover the experienced user's emotions. Even if the core dependent variables of this thesis are the counterfactual emotions of regret and disappointment and the attribution of responsibility; these measures have been collected with different methods according to the framework and literature.

Due to the multidisciplinary purpose, the three experiments reported in this thesis have their basis in both psychological literature as well as in advice taking literature. The different perspectives taken are the main reason why the main variables of regret, disappointment, and responsibility have been collected with different methods.

In particular, in the first explorative experiment, it was used Regret and Disappointment Scale from Marcatto and Ferrante (2008) to assess counterfactuals with a validated scale.

In the second experiment from a psychological perspective, it was preferred to assess counterfactuals with explicit labels on the scale using the questionnaire from Giorgetta (2012, pp. 117–118). The third experiment had its basis in advice-taking literature and algorithm aversion studies; hence the measures of regret and disappointment have been collected following the practice commonly used in these two fields. Then, the assessment of counterfactuals was carried out by a list of positive and negative emotions constructed from items from previous advice literature (MacGeorge, Guntzviller, Hanasono, & Feng, 2013; Prahl & Van Swol, 2017).

Chapter 2: aimed at exploring the dimensions that affect regret, disappointment, internal and external attribution of responsibility. The task was a scenario repeated twice in which the participants had to choose whether to buy a flight ticket. They could follow the suggestion given (by a group of travel agents or by a machine learning system) and buy the ticket now or wait two weeks, hoping for a better price. The results showed that people felt more regret when they followed the suggestion given by a human compared to the experienced emotions when they had a wrong outcome with the computer machine. Besides that, linear mixed-effect models explored the relationship of the collected variables explaining regret, disappointment, internal and external attribution. It was found that other variables are involved in the experience of the dimensions reported above. Mainly, besides the source of suggestion, also trust in the suggestion, locus of control, the time of the experience, objective numeracy have a role in this scenario.

The primary limitations found are about the nature of the scenario that was perceived as not engaging enough, and the purpose of the study (work trip) indirectly could influence the participants and limited the experience of the counterfactuals. These issues can be seen through the scores that appear to be centered on the scale in both experimental conditions.

Chapter 3: investigates the specific experience when the user follows a suggestion given by an intelligent system that leads to a wrong outcome. The study tested two scenarios based on psychological literature (Giorgetta, 2012; Inman & Zeelenberg, 2002); it was run to test the emotional differences of responsibility, regret, and disappointment.

Experiment2 (Chapter 3:) found that there is a difference in terms of responsibility, even if this finding does not meet the initial hypothesis on regret and disappointment. People feel more responsible when they experience an adverse outcome after have followed a human suggestion compared to the wrong outcome after an algorithm suggestion.

The main limitations of these results are the scores of the questionnaires that have quite high values. These scores might be due to a bias in the participants' recruitment process in a crowdsourcing service, or this could due to some confounder not tested in this experiment as the locus of control, the difference in the perception of the system, and personality traits.

Chapter 4: has its basis in advice-taking and advice-discounting literature. Experiment 3 has its roots in a widely used paradigm in algorithm aversion literature (Logg et al., 2019) and tested a scenario in which participants had to make a forecast twice, and they have been presented with a suggestion from an algorithm or a human between the first and the second evaluation. This experiment assessed, besides regret and disappointment, the Weight of Advice (WOA) and the SHIFT measures, two measures related to the reliance or the discounting of the received advice.

Experiment 3 started with the idea that people prefer computer advice in the first instance, but they trusted less and change behavior when they see an error from the algorithm, compared to the human. This result can be related to the higher expectations people have in the algorithm that can lead to a higher experience of the counterfactual emotions an error can provoke. Hence, the primary purpose was to replicate results found in previous research on algorithm reliance (Logg et al., 2019) and relate the Weight of Advice (WOA) to counterfactual emotions. The experiment replicated the results of implicit reliance on algorithm suggestions through the WOA measure collected. Participants who had the algorithmic suggestion moved more towards the suggestion, while people who received human advice switched significantly less. Another interesting finding was about the positive correlation between how much people relied on the suggestion and the perceived guilt and self-blame after receiving the bad feedback after they followed advice from a human, while there was no relationship when people followed bad advice from an algorithm. From the previous study, it was not replicated the results of increased confidence in the participants during the use of an algorithm. This difference might be due to the different samples between Logg's design and the one presented here. The sample in this study was collected on Amazon MTurk from the US population, while the previous sample was collected on the same service without any specifications. These cultural differences might be one of the reasons for these results.

Chapter 5: summarizes the findings, the theoretical and practical implications as well as future directions of this research. The results shed light on questions about emotional reactions to adverse outcomes and the importance of these aspects to design better systems according to the different decisions a human decision-maker has to take.

These three studies observe in three different ways the topic of algorithm aversion and its relationship with the counterfactual emotion of regret. The first study has an exploratory nature and aims to model the relevance of several factors, such as the role of trust, locus of control, time of the experience, and objective numeracy. The experiment aims to understand the role of these factors during a decision-making task that results in a wrong outcome, and how they interact with the source of suggestion, being a human or an intelligent system.

The second study follows the first one in the comparison of the experience of counterfactual emotions in reaction to an adverse outcome, and it focuses more precisely on the role of perceived responsibility. Self-responsibility has been found to correlate with perceived regret (Frijda, Kuipers, & Ter Schure, 1989); therefore, this study is useful to grant a more detailed picture of the counterfactual emotions involved in a decision-making task based on a human or computer interaction. In particular, the findings highlight that the participants felt less responsible for the bad choices and outcomes when they have followed the computer's advice. The fact that following advice provided by a computer decreases the perceived responsibility is a relevant aspect to explore further since it has the possible implication of dehumanizing individuals' choices when taken following an algorithm. It could indicate a lower moral involvement in the decision-making process.

The third study belonging to the decision-making category and judge-advisor system, as described above, it investigates the role of the phenomenon of "algorithm appreciation." This phenomenon happens when individuals are more inclined to follow an algorithm-advice, compared to a human one, until they observe that the very same algorithm makes a mistake. This phenomenon is based on the assumption that an algorithm must always make more accurate suggestions compared to a human, given its higher calculation power. The assumption leads individuals to trust the algorithm, despite its non-human nature. The experience of observing an algorithm making a mistake, after the expectation of perception, makes people switch their behavior promoting higher reliance in humans than in algorithm suggestions, leading to the choice of discounting the algorithm-based one for the future uses. In this dissertation, the phenomenon of "algorithm aversion" is put into relation with the experience of counterfactual emotions, to observe a possible relationship. To author's knowledge, this is the first study attempting to observe a correlation between emotion and the specific behavior generated by the "algorithm appreciation" phenomenon.

All these experiments together grant the possibility to observe the presence of counterfactual emotions as result of a bad suggestion given by a human vs. by a computer, from three different complementary perspectives: investigating the plurality of factors interacting in the decision-making process, observing the impact on perceived responsibility during a decision-making process, and investigating the role of specific emotions that lead to discard an algorithm even in cases in which at first glance the algorithm was appreciated.

Despite the exploratory nature of this thesis, all three experiments together shed light on the crucial role of emotion during human-computer interactions, and grant a general perspective indicating that in order for algorithms to be more consistently used in everyday life the very human and emotional experience of the users must be taken into account. Despite the limitations found, this work shows the potential of this promising field of study that has the intent of maintaining the human at the center of the design of these intelligent systems that will be more and more spread in the next future.

In particular, the consideration of users' regret and disappointment during the interaction design phase with intelligent systems can significantly help go further algorithm aversion and prevent technology abandonment. What unifies these studies is the effort of finding a set of principles or guidelines aimed at guiding designer of intelligent systems. Specifically, the result of this work may suggest that more attention needs to be given to the understanding of counterfactual emotions during the interaction with intelligent systems to prevent the abandonment of the technology. A relevant point for designer is the importance of the context of decision and the presentation of the feedback after a wrong outcome. This dissertation is the first attempt to the author's knowledge to merge the study of counterfactuals with the experience of algorithm aversion after users see the algorithm errs. The primary purpose of this thesis is to give attention to the role of counterfactuals during the interaction with complex systems.

Chapter 1: Literature Review

1. The pervasive role of algorithms

More than twenty years ago, Mark Weiser envisioned the next future in which computation would be wholly embedded in the world. He coined the term "ubiquitous computing" to refer to the presence of technologies that do not require active attention to be used (Weiser, 1991).

Today with recent developments in information technology, many of Weiser's predictions have come true, with the advent of more and more powerful hardware and the pervasiveness of automated recommendations based on data. Examples of this ubiquity can be domains as shopping online (Amazon), listening to music (Pandora), or even more personal domains such as dating (Tinder).

Nowadays, with the advent of smartphones and big data, we are fully immersed in the use of programs and apps that are based on algorithms. Often we use the outputs of these apps without active thinking of them, and they are completely embedded in our everyday life. Due to their broader and broader adoption, it is fundamental to understand how users experience affects the decision-making process through these algorithm-based systems and the emotions that can be experienced after following a suggestion that leads to a non-optimal outcome. In the next sections, it will be presented an overview of the importance of algorithm-based application and the emotional aspects related to the decision process, in particular, when the outcome does not meet the initial expectations.

2. The importance of algorithm-based applications

Algorithms are well-defined scripts for sequences of mathematical calculations or procedural steps that lead to a solution to a particular problem (Cormen, Leiserson, Rivest, & Stein, 2002). They can complement, improve, and inform human judgment through advice and suggestions. Although companies have traditionally relied on human's gut feeling and expert intuitions to forecast future events, the advent of Big Data and algorithms-based systems lead to better predictions and smarter data-driven decisions (Brynjolfsson & McAfee, 2012; McAfee & Brynjolfsson, 2012; Sundsøy, Bjelland, Iqbal, Pentland, & De Montjoye, 2014).

Algorithms can surpass human accuracy for a few reasons. Aggregation of individual judgments can outperform a single user's forecast (Hastie & Kameda, 2005; Larrick & Soll, 2006; Youyou, Kosinski, & Stillwell, 2015). Even a simple average can perform better than a single user's judgment (Bachrach, Kosinski, & Gael, 2012; Turner, Steyvers, Merkle, Budescu, & Wallsten, 2014). Previous studies showed how the average between people's judgments reduces the error from individuals. This enhanced precision can be referred to as the *wisdom of crowds* (Galton, 1907; Mannes, Soll, & Larrick, 2014; Surowiecki, 2004), this process of aggregating opinions of people cancel the errors of individuals in order to boost decision accuracy.

However, algorithms are not usually based only on the simple average of human judgments, and they are trained with different and more complex inputs. Besides, more complex algorithms, like deep learning, overtake simple average in terms of accuracy, weighting more appropriately training cues. (Baron et al., 2014). This kind of complex application of algorithms results in what is called "black box" or unintelligible systems form, which is not possible to trace back the system's reasoning about the outcome. (Rudin, 2019). Dealing with complexity and complex unintelligible systems during a non-optimal decision process can provoke negative emotions, and this can affect the future in the system. The next sections will present two phenomena that can occur while following advice from an algorithm, algorithm aversion, and algorithm appreciation.

3. Algorithm aversion: distrust in algorithmic suggestions

Nevertheless, the higher algorithmic accuracy compared to humans, people usually show distrust in the advice made by these complex systems.

The dispute between algorithms versus human has a long story. Meehl's book "Clinical versus statistical prediction: A Theoretical Analysis and Review of the Evidence" (1954) is the first reference on the superiority of algorithms over human experts and the "irrational" skepticism in this superiority. As Meehl found in 20 forecasting tasks, even a linear model outperforms human judgment. Over decades other studies with the same anecdotal content spread in the research community (Dawes, 1979; Dawes et al., 1989). Then the claims over the distrust in algorithmic advice received attention from the decision-making research. Kahneman resumed Meehl's story on the hostility spread amongst clinical psychologists about the finding that even a simple algorithm can surpass and be more precise than human experts (2011, p. 227). Only in recent years, the attention has been paid on empirical evidence of the distrust in algorithmic suggestions, and the debate is still open. This irrational discounting of algorithmic advice has its roots in the "clinical versus actuarial" dispute, and only in the last few years, this phenomenon was found in forecasting domain (Önkal et al., 2009) and has been called "algorithm aversion"(Dietvorst et al., 2015; Prahl & Van Swol, 2017). This aversion could be explained by reasons stated by Dietvorst et al. (2015), and they include the perceived inability of algorithm to learn (Dawes, 1979; Dietvorst et al., 2015), the exaggerated role of human experience in forecasts (Dietvorst et al., 2015; Highhouse, 2008), the idea that algorithms are dehumanizing (Dietvorst et al., 2015; Grove & Meehl, 1996), the difficulty to accept that a machine can perform better than humans (Grove & Meehl, 1996), the users' desire for perfect forecasts (Dawes, 1979; Dietvorst et al., 2015; Highhouse, 2008; Madhavan & Wiegmann, 2007). This influential set of papers offers different hypotheses about the reason for algorithm aversion, while they lack empirical evidence.

The seminal empirical evidence can be found in the works of Dietvorst and colleagues (2015) and Dzindolet and colleagues (2002). With their works, the authors found that participants quickly lost their reliance on algorithmic advice when they see them err. To solve this issue, the research group of Dietvorst proposed a solution to overcome algorithm aversion, and through three studies, showed that people are more inclined to use algorithms if they can slightly modify them (Dietvorst, Simmons, & Massey, 2016). A recent paper demonstrated that the algorithm aversion is present in decisions under uncertainty, even if the algorithmic output is the best possible algorithm in the field (Dietvorst & Bharti, 2019). These experiments showed that participants relied more on themselves compared to the algorithm. These results have been found in the subjective domain as well, Yeomans and colleagues (2019) found that participants relied more on friends/other humans when they have to recommend jokes, even if the algorithm performs better.

4. Algorithm appreciation: reliance on algorithms

On the other hand, evidence from computer science showed that participants relied more on a suggestion given by an "expert system" compared to a suggestion given by a "human" solving logic problems (Dijkstra, Liebrand, & Timminga, 1998). Dijkstra (1999) found that even if the algorithm errs people trust in the actuarial advice compared to other humans. Dijkstra's results seem to be contradictory to Dietvorst's work. On this line, Banker & Khetani (2019) posited that consumers tend to totally depend on algorithmic recommendations even when they are aware of the inferiority of products or services proposed by the recommender. This phenomenon is what the authors called "algorithmic overdependence."

In this debate, it is worth to mention the work of Logg and co-workers (Logg, 2016, 2018; Logg et al., 2019). They found empirically that the context influences how people trust or not the suggestion given by an intelligent system. They found that people usually rely on an algorithm, in particular participants followed algorithmic advice more in objective decisions than in subjective decisions and when they have to cope with mathematical problems. Nevertheless, the authors' analysis did not take into account when the suggestion results in a wrong outcome.

Logg and colleagues (Logg, 2016, 2018; Logg et al., 2019) claimed that there is a general algorithm appreciation, contradicting the advocates and the literature of algorithm aversion (Dawes, 1979; Dawes et al., 1989; B. Kleinmuntz, 1990; D. N. Kleinmuntz & Schkade, 1993; Meehl, 1954, 1957) showing that this topic is not straightforward as it seemed as the literature suggests (e.g., Dietvorst et al., 2015), and she reconciled the incompatible results previously found. Logg and colleagues found that aversion only appears under certain conditions and reliance under others. Another point that needs attention is the role of expertise. Experts trusted more their gut feelings than the advice given by the algorithm. A general result is that people usually consider algorithmic advice from others when they face objective decisions, while they trust more human suggestions when they face subjective decisions.

These results appear to confute Dietvorst's findings (Dietvorst et al., 2015), however, while the work of Dietvorst and colleagues focused on the reliance after an algorithmic error, none of Logg and coworkers' studies found aversion before seeing the algorithm errs. Evidence for algorithm aversion can also be found in Prahl and Van Swol (2017), who tested in a repeated interaction task the reliance on algorithm before and after participants saw it err compared to a human. What they found is that participants rely on an algorithm in the same way as they did in the human condition, while there was a more significant discounting in the algorithmic advice after an err, compared to human advice, even when participants followed the human wrong suggestion.

In the presented studies, a high relevance had the domain of decision one is facing and the understandability of the machine's reasoning. Each of these works focused on the user's perception of an unintelligible algorithm-based system, or what is usually called a "black box." In the next section, the user's interaction with complex systems based on algorithms will be unfolded.

5. Reliance on complex and opaque algorithms

As seen in previous paragraphs, according to the literature, dealing with complex algorithms and the adverse outcome can lead to discounting precious algorithmic advice. Algorithm aversion and algorithm appreciation are fundamentally related to the idea of future collaboration with the systems to understand and prevent the abandonment of technology.

In this section, there is presented an overview of the interaction amongst users and complex or opaque systems based on algorithms. In different decision domains, people express reliance or discounting on the algorithmic advice in favor of human judgment, and the mistakes made by an algorithm result in more severe way compared to human mistakes. This different perception of the seriousness of the mistake can undermine the relationship and the cooperation between human-computer despite the fruitful advantages of this relationship.

Early on, the cooperation with artifacts was smooth due to the complete comprehension of the system's functioning because of its simplicity. The intention to cooperate has a relation with two main aspects; the complexity of the automation and trust in the algorithmic suggestion.

These two features are intertwined as the complexity leads to difficulties and different strategies of trust in the system.

Back in the history of algorithm-based artifacts, computers were just mere tools, with a clear and understandable path beyond their calculation (Bradshaw et al., 2009). According to Norman, simple tools allow us to understand precisely the model beyond the artifact (2004, p. 76). If we share the same mental model of the design, we can easily understand how the system works. On the other hand, with the advent of more complex systems, new theories have been developed.

As stated by Norman (2004), people interpret everything that they experience in human terms; this way of attribute motivations and emotions to other people, animals, and objects; this is called *Anthropomorphism.* As human beings, we have the natural tendency to attribute mental states, beliefs, intents, emotions, and so forth to every animate or inanimate object. We are biologically primed to social interactions, and the basis of this interaction is the ability to understand others' minds. (Rand & Nowak, 2013; Rand et al., 2014). Reeves and Nass (1996) demonstrated with many studies that people treat "computers as social actors." That is to say if an object interacts with us, showing specific characteristics like language, turn-taking, reactivity, we implicitly and naturally treat this object as a social actor. As Norman (2004) noted in his book "Emotional Design" this social interaction with artifacts allow us to get mad with a computer, thus we can experience emotions about the wrong outcome of a decision when following a suggestion.

Early research by Nass and colleagues suggested that people treat machines in a social manner (Nass & Moon, 2000; Nass, Steuer, & Tauber, 1994; Norman, 2004), that is to say, that people act with complexity, as computers and algorithms, as they do with humans. Another approach examining the degree to which computers are treated in social ways focuses on people's inclination to anthropomorphize non-human entities because of the perception of having a mind (Waytz, Gray, Epley, & Wegner, 2010). Two measurements constitute mind perception: experience, which is related to the experience of feelings and emotions, and agency, which is linked to responsibility of actions, self-control, and planning (Gray, Gray, & Wegner, 2007) According to this view if a person perceives a computer as an entity capable of agency as well as sensing and feeling experience, this perception predicts if the machine is treated like an actual person (de Melo & Gratch, 2015). On the other hand, recent studies showed that people cooperated less with a computer agent compared to a human agent (de Melo, Carnevale, & Gratch, 2014). This can be defined as a bias for the in-group, or people foster humans because they have a similarity due to agency and experience (Blascovich et al., 2002; Gray et al., 2007; Waytz et al., 2010). In other words, people can dehumanize machines because they do not have these mental abilities (Haslam, 2006). This perceived difference with machines can lead users to perceive algorithm-based systems as only as artifacts.

Early on, trust was an essential factor in HCI, and the advent of big data, more powerful calculation competence, and advanced graphical user interfaces such as natural language communication or anthropomorphic virtual agents make these artifacts more similar to teammates than mere tools. In this framework, new essential investigation raised on how people interact and trust with complex machines and how this reliance on them differs from the human-human trust.

Trust helps people to rely on automation and overcome the complexity faced in the increasing sophisticated interaction with more and more unintelligible machines (J. D. Lee & See, 2004).

Trust, a social psychological concept, seems particularly important for understanding human-automation partnerships. Trust can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability.

Trust in HCI is not a new field of study, and has extensive literature exploring how people treat and cooperate with computers.

As trust mediates human relationships, it can also mediate the interaction between humans and automation. Many studies showed that trust is an important dimension to define human-automation interaction (e.g. Lee & Moray, 1992; Zuboff, 1988).

Proper forms of trust forestall maladaptive human behaviors induced by distrust (e.g. Dietvorst et al., 2015) and overdependence (e.g., Banker & Khetani, 2019) (J. D. Lee & See, 2004). An important aspect is that when the trust is violated and the relationship is being broken. When the trust has been violated the trustor (i.e., the user of the system) can feel responsible for having relied on the system or blame the system. Nevertheless, this is not straightforward.

Over the years, human-computer interaction changed from using the machines as mere tools to view them as potential teammates in cooperative settings (Bradshaw et al., 2009; Nass, Fogg, & Moon, 1996). In human-human interaction, a typical and effective way to achieve this is blame. Blame attributions provide the opportunity to regulate behavior, identify problems, and help to fix them (Malle, Guglielmo, & Monroe, 2014). In order to achieve trusting human-computer cooperation, the HCI community requires cooperative frameworks addressing such limitations.

It has a paramount relevance to understanding the use and misuse of technology, as the automation and algorithm-based systems are becoming more and more complex.

In the next paragraph, an overview of the violations of trust will be presented, in order to understand the differences in the concept of blame, during a decision with an algorithm and another human.

6. Trust violations: blame in human-algorithm interaction

As seen in the previous paragraph, trust is a concept from social psychology defined by two agents, the trustor, and the trustee. Trust can be defined as the attitude that the trustee will help to achieve the trustor's goals in a situation defined as "uncertain and vulnerable" (Buchholz, Kulms, & Kopp, 2017; J. D. Lee & See, 2004).

According to Mayer and colleagues (1995), trust results from the evaluation of the trustee's abilities to fulfill the trustor's expectation. The violations of trust occur when the trustee does not satisfy the trustor's expectancies. These violations usually lead to three specific emotions: disappointment, anger, and regret for "having trusted in the first place" (Martinez & Zeelenberg, 2015, p. 119).

When trust is violated, people can experience blame. According to Malle, Guglielmo and Monroe (2014) blame is a form of public criticism that can only be directed towards a person, and strongly rely on social cognition and always requires a judgment on the reasons of blaming. Three different dimensions stem from this definition that distinguishes blame from wrongness judgments, anger, and event evaluation. Event evaluation is directed to an event; wrongness judgment has as its target a purposeful behavior, while anger can be experienced towards anything (Buchholz et al., 2017).

These three emotions do not require warrants; blame is a form of responsibility assignment that requires the processing of information and has its purpose in the regulation of social behavior.

The act of blaming can be seen as an invitation to communicate and allow to repair trust (Malle et al., 2014) whereas blame is a form of attribution of responsibility that has no intrinsic defined emotions in itself. The emotions that blame convoys depend on whether one blames her/himself or another person. Another important aspect is the intentionality of non-optimal behavior. When blame is addressed towards another person, one can experience anger if the behavior in question is perceived as intentional, or pity if the behavior was a result of unintentionality. On the other hand, if blame is addressed towards the self, one can experience guilt if there was intentionality or ashamed when the negative outcome was not on purpose.

The emotions that relate to the feeling of blame are different according to the sense of responsibility and have different behavioral consequences. Sometimes is not easy to distinguish these emotions that stem from the same trust violation or can be experienced at the same time. For example, the emotion of guilt has a lot in common with regret, they both can be experienced in the same time, and even if regret is a broader emotion than guilt, both are tied with interpersonal harm, even if guilt can be experienced when a negative outcome affect others and not only oneself (Zeelenberg & Breugelmans, 2008).

Martinez and Zeelenberg found in an experiment contrasting effects of these emotions (Martinez & Zeelenberg, 2015). Regret and anger are strongly related to losing trust while experiencing disappointment increases it.

Trust is not only affected by the experienced emotions of the trustor, but it is also affected by the repair strategies of the trustee. When the trustee feels responsible for the inability to help the trustor to achieve her/his goals, the trustee blames her/himself. Usually, feelings of regret and guilt that follow self-blame decrease trust (Kim, Dirks, Cooper, & Ferrin, 2006).

These feelings of the trustee can lead to an apology in order to repair trust. Different types of apologies affect different positive ways of trust violation (Kim et al., 2006).

Kim et al. (2006) found concrete evidence about the positive effects of apologies as a means to repair trust violations. Usually, successful trust repair is achieved after violation by apologizing with an external or an internal attribution; consequently, the trust-repairing agent needs to take the responsibility of the wrong outcome or attribute blame to an external source of influence.

These positive effects of apologies to recover trust can be found even in human and virtual agent interaction. De Visser et al. (2016) demonstrated that anthropomorphism enhances the effect of apology as a trust-repair strategy of a virtual computer agent in a cooperation task.

To sum up, the blame is experienced when a person is pointed as responsible for a norm-violation action or a negative event, and there is a reason to think that this person acted on purpose or had the capacity to prevent an adverse outcome (Malle et al., 2014). In addition, the attribution of the blame can have negative or positive effects on trust (Kaniarasu & Steinfeld, 2014; Martinez & Zeelenberg, 2015).

Apologies form the trustee are the basis to repair trust in case of negative consequences (Kim et al., 2006). Nowadays, the research has demonstrated that human interaction with the computer and virtual agents can be social and similar to human-human interaction (Lee & Nass, 2010). However, it is not clear how attributions of blame and trust decrease affect human-algorithm cooperation, in which there is a lack of anthropomorphic interface, and there are no trust violation repair strategies.

With this idea in mind, the author hypothesizes that it is more difficult to blame an artifact than the self, during a decision-making process with machines, compared to the same process with other humans.

7. Emotions related to wrong decisions: Regret and Disappointment

Regret and disappointment are two emotions that stem from counterfactual thoughts. Counterfactual thoughts can be defined as a mental simulation that occurs after a situation that happened, through the comparison between "what now is" and "what might have been." Counterfactuals emotions of regret and disappointment have on their basis the comparison between the factual and counterfactual result of a choice, in particular for regret is related with the outcome one could obtain with a different choice, or, as regards disappointment, if another state of the world happened (Zeelenberg, Van Dijk, Manstead, & Van Der Pligt, 1998). That is to say that these two emotions differ on the source of comparison from which they arise.

Regret results from the comparison of the outcomes of different actions given a particular state of the world, and it generates behavior-focused counterfactuals. Regret is bounded with a sense of self-responsibility in which one can change her/his behavior in a certain state of the world

Disappointment, instead, is provoked by external attribution of responsibility, by comparing different states of the world and subsequent outcomes given a choice. Disappointment generates situation-focused counterfactuals and implies a change in events beyond personal control (Zeelenberg, Van Dijk, Van Der Pligt, et al., 1998).

As seen in previous sections, algorithm aversion and appreciation are strongly related to the process of choice. People can decide to follow or not the suggestion presented by an algorithm, and if it is wrong can lead to suboptimal outcomes. Two emotions are quite often experienced during a suboptimal decision process and outcome: regret and disappointment.

Regret and disappointment are "negative, cognitively determined emotions that we may experience when a situation would have been better if: (a) we had done something different (in case of regret); or (b) the state of the world had been different (in case of disappointment)"(Giorgetta et al., 2012). In other words, both emotions can be defined as counterfactual (Zeelenberg, Van Dijk, Van Der Pligt, et al., 1998).

Counterfactual thoughts can be defined as a mental simulation that occurs after a situation that happened, through the comparison between "what is now" and "what might have been".

Due to the importance of the emotions of regret and disappointment, these two emotions received wide attention from economists as well as psychologists. Early theories have been studied by economists and investigated how the feeling of anticipated regret affects the decision process under uncertainty (Bell, 1982; Loomes & Sugden, 1982). From a psychological point of view, the focus was on how negative outcomes could intensify the experience of regret (Kahneman & Tversky, 1982). The attention was then dedicated to how disappointment influences decision making (Bell, 1985; Loomes & Sugden, 1986). The main difference between these two emotions is based on the feeling of responsibility that can be found at different levels.

Many theorists claimed that the role of self-blame and responsibility is one of the main aspects in the experience of regret (Frijda et al., 1989; Gilovich & Medvec, 1994; Ordóñez & Connolly, 2000; Sudgen, 1985; Zeelenberg, Van Dijk, & Manstead, 1998, 2000).

As stated before, counterfactual emotions arise through comparison between the alternative not experienced. Regret and disappointment are two different emotions related to the hedonic values of the outcomes.

These two emotions can be discriminated against on their content and behavioral consequences (Zeelenberg, Van Dijk, Manstead, et al., 1998). Regret can be experienced when one could not obtain the expected goals, while disappointment arises when there is a goal abandonment (Zeelenberg, van Dijk, Manstead, & Van Der Pligt, 2000). Zeelenberg and colleagues found that regret is related to a behavioral switch and a trust reduction, whereas disappointment increases trust (Martinez & Zeelenberg, 2015; Zeelenberg & Pieters, 2004). In other studies, Zeelenberg found that regret increases prosocial behaviors, while disappointment reduces prosocial behaviors (Martinez & Zeelenberg, 2015; Martinez, Zeelenberg, & Rijsman, 2011b; Martinez et al., 2011a; Van Kleef, De Dreu, & Manstead, 2006)

Given these differences in how these emotions affect behavior, it is relevant to know whether people feel disappointed or regretted and what are the relevant aspects of these affective reactions. In this work, the main interest is aimed towards understanding when these emotions arise after a suggestion results in an unexpected outcome, and the differences when it is given by a human or an intelligent computer.

8. Limitations of existing evidence

This section discusses the limitations to both of the current studies on regret and disappointment related to intelligent systems, and how this dissertation addresses these open questions.

Firstly, the limitations regarding the studies that involved regret were discussed in comparison with the repeated interactions with intelligent systems. Secondly, a discussion on the potential of this research about the importance of the understanding negative emotions that could arise after a wrong decision and the aspects that could involve HCI to handle and improve cooperation between different agents (human and computer).

8.1. Negative Outcomes

Prior works have been focused on investigating whether the users trust or not in an algorithm compared to a human after seeing it err (Dietvorst, 2016; Dietvorst et al., 2015; Dzindolet, Pierce, Peterson, Purcell, & Beck, 2002). This attention has led to what is called *algorithm aversion*, and the results have shown that people's trust relies more on humans than on algorithms suggestion. This phenomenon leads companies and organizations to have expensive improvements that could lead to disuse or misuse of the new intelligent technologies due to the user's perceptions of them (Parasuraman & Riley, 1997). As reported before, Hung and colleagues (2007) found that the measure of the avoidance of experienced regret could be useful to improve the success in the adoption and the usage of an intelligent system. Hence, the understanding of regret and disappointment could support the interface design to plan more appropriate applications that can be accepted in a more correct fashion.

8.2. The interest in maintaining a Human-In-The-Loop

In the near future, humans will face the continuous and increasing development of automation, systems based on AI and intelligent computers; these technologies will be more and more complex (Borst, 2016). Studies in this field agreed that to improve new approaches and framework, able to understand the cooperative relationship, and more information needs to be shared within the users and the system (e.g. Borst, 2016; Christoffersen & Woods, 2001; Norman, 1990). Nevertheless, there is no explicit agreement about what kind of information needs to be shared by the machine (e.g., information related to input, process, and output) without overloading the user's cognitive system. It is fundamental to take the user in the center of the design of automation and interfaces to build technologies that can be members of the human-intelligent system team. In order to develop cooperative relationships between human and computer agents, it is fundamental to understand how people use the information of algorithms, what individual characteristics influence the use of the system and its reliance, the environment in which these technologies will be used and the specific decision tasks that need to be accomplished.

Chapter 2: Do people feel more regret following a wrong suggestion from an algorithm or other people?

1. Motivation and aims

This study explores emotional reactions (Regret and Disappointment) to bad choice in a "purchase a flight ticket" scenario. This experiment aims to understand the underlying relations between the experience of counterfactuals and the source of suggestion (human vs. computer) taking into account the role of trust in the advice and the confounder variables of locus of control, numeracy, attitude towards technology and personality traits.

This first experiment aids the understanding of the experience of counterfactual emotions through analyzing participants' implicit trust, or the trust users showed in following the suggestion, in a task with a suggestion given by a human or an "intelligent machine." The primary purpose is building probabilistic exploratory models to explain how regret, disappointment, and attribution correlate with the presented source of suggestion.

Hence, this experiment explores how much people rely on the suggestion from an "intelligent computer" compared with the suggestion from human experts and the differences in the experienced emotions after they realized the "intelligent system" or the human experts provided a wrong outcome. In this work, the primary purpose is to explore the differences in regret and disappointment between a human suggestion and computer suggestion in a binary choice in which the user can rely on the given suggestion or not. The user has to make a decision based on the suggestion she/he received within a fictional flight ticket scenario, constructed to provide negative feedbacks generating the experience of counterfactual emotions.

According to Pieters and Zeelenberg (2007), regret and disappointment are two counterfactual emotions that are strongly related to decision-process with bad outcomes or disconfirmed expectancies. These emotions differ in what concerns their antecedents; regret is closely related to self-attribution (self-agency), while disappointment is more related to external attribution. One can experience regret while observing that the outcome of a decision would have been better if she/he had done something different, having direct responsibility of the outcome, on the other hand, one can experience disappointment while observing that the problematic situation she/he is facing is the result of unexpected events in the world, on which the person could not have direct control. Previous works have shown that different counterfactuals could lead to different behavioral consequences (Bougie et al., 2003). These consequences are particularly important during the use and cooperation with intelligent systems. The design has its basis in psychological literature presenting two different scenarios, albeit it has been modified to achieve the goals of the study through the lenses of algorithm aversion studies and advice taking framework.

2. Hypotheses

This experiment has its basis in the difference between regret and disappointment in terms of the perceived responsibility and the attribution of blame.

The main aim of this study is to compare the experience of counterfactual emotions and attribution of responsibility after having followed a poor suggestion provided by a human expert or by a trained algorithm. Secondly, the study aims to investigate the presence of a possible correlation between the emotions experienced by the participants and some individual traits present in the population. In particular, numeracy (the ability of doing simple mathematical calculation), the level of trust that the participant showed towards the source of suggestion, and her/his attribution of locus of control will be modeled seeking a correlation with the type and the intensity of emotion they have experienced in response to the human or artificial suggestion.

This manipulation and analysis aim to provide a more detailed description and explanation of the complex interaction between the individual emotional response to a bad outcome, and the source of the suggestion, being human or virtual. The final goal is to better understand how the emotional response of individuals changes when interacting with a computer instead of a human being. These effects give insight into how counterfactual emotions differ compared to the source of suggestion and disentangle the complexity of these emotions in relation to human-computer interaction. Two research questions are the core of this study. First, what are the main variables that affect the experience of regret, disappointment, and attribution? Second, is there a relationship between the measures of regret, disappointment, internal and external attribution, and trust? Eventually, discover what is the effectiveness of the confounding variables for counterfactual emotions and perceived responsibility in an algorithm scenario.

In this experiment, the hypotheses have been stated following Haslam (2006), according to the idea of dehumanization of the machines (see Chapter 1:.section Chapter 1:5). Hence, the main idea is that when following a suggestion given by a machine is more difficult to feel external responsibility from the outcome compared to the opportunity to blame someone else while following a wrong suggestion given by humans.

Despite the exploratory purpose of this study, five hypotheses guided this work:

H1: Participants feel more internal responsibility while following the suggestion of an intelligent machine compared to human advice;

H2: Participants feel more regret while following the suggestion of an intelligent machine compared to human advice;

H3: Participants feel more external attribution while following the suggestion of a human compared to participants who followed the machine suggestions

H4: Participants feel more disappointment while following the suggestion of a human compared to participants who followed the machine suggestions

H5: In the first task, according to previous literature (e.g., J. M. Logg et al., 2019) participants have more trust in the suggestion provided by the algorithm than in the suggestion provided by the human experts.

In this experiment, the hypotheses are related to understanding the algorithm aversion through the experience of regret, disappointment, the internal and external attribution of responsibility.

In particular, according to the ideas described in the previous section (Chapter 1:), the leading hypothesis is that regret, disappointment, and the attribution of the blame can be explained mainly by the source of the wrong advice.

This work also aims at finding experimentally the relationship between Objective Numeracy, locus of control, time of experience, and trust that can influence the participants' perceived regret, disappointment, internal and external attribution after a bad outcome. In the following sections, it will be presented the methods, followed by a detailed description of the experimental protocol. Then, results are discussed, and finally, this work concludes with a discussion on the main findings and the critical points encountered.

3. Equipment

The experimental task was built using Axure RP8[1], a software that allows to prototype websites. To use the prototype online, the preview of the website was then recoded in HTML programming language and uploaded on Trento University's servers.

The questionnaire was built through Google Forms to be easily spread and incorporated into the experimental online task.

4. Methods

This section describes the experimental design and the metrics recorded, then the description of the sample is presented.

1 https://www.axure.com/

4.1. Experimental design

The experiment had two cells (advisor: human experts vs. intelligent system) between-subjects repeated measure design that manipulated the source of advice participants received.

For each subject, the order of presentation of the tasks was counterbalanced to prevent order effects. To prevent the learning effect, for each task, two different yet analogous assignments were identified.

Participants had to choose whether to buy a flight ticket for work purposes; they could decide to follow the suggestion given, buying the ticket "Now," or they could wait to hope for a last-minute ticket. They did so twice, and the feedback of their choice was always negative. The two conditions differ for the source of the suggestion. In the case Human Expert, the suggestion was given by "some professional travel agents…"; while, in the Computer Agent condition, the suggestion was given by " a machine learning system with a probability of 80%" of correctness.

This study operationalized a decision making task in which the user has to buy a flight ticket and she/he can decide to follow the given suggestion or not. Hence, the reliance on the suggestion, and how the (always wrong) feedback influences counterfactual emotions were measured. The task was administered twice to see the differences between the first wrong suggestion and the second wrong suggestion (Zeelenberg & Pieters, 1999). The scenario was "buying a flight at the best price" task. This task simulated the suggestion that could be given by human experts and an algorithm based on historical data. After the user's choice, feedback about the wrong decision was given. That is to say, participants always had a bad outcome. This experiment follows two conditions mixed-design. Each participant had to make two choices (buy two flight tickets) and can follow or not the given suggestion, in any case, the feedback was negative to provoke counterfactual feelings. Due to the nature of the study, the participants were first debriefed with a different with different instructions about the actual aims of the experiment, while the real aims of the data collection were explained at the end of both trials.

4.2. Setting and Sample

The data collection took place through a non-lab setting in a web-based study.

The collected participants were people from Italy aging from 18 to 65 years old. The only inclusion criteria were that participants had to be native Italian speakers to understand the experimental instruction and scenarios. The final sample was collected amongst the students of the University of Trento. In total, 270 participants took part in the study. Due to technical issues with the laptop, an experimental session was interrupted, and the data were excluded from analysis; data from another participant has been excluded from the analysis because she was not a native Italian speaker. Ninety-two participants have been removed from the analysis because they did not complete the study for technical reasons at the early stage of running.

A final sample of 178 people was then analyzed. The sample had a mean age of 26.47 (SD=5.39) ranging from 18 to 63 years old. In this sample, 110 are women, and 68 are men. The average age of the female sample has a mean of 26.34 (SD=4.88), and the male sample has a mean equal to 27.79 (SD=6.05).

In the **control condition (human suggestion),** there were 77 participants. The sample had a mean age of 23.76 (SD=4.158) ranging from 19 to 35 years old. In this sample, 50 were women and 27 were men. The average age of the male sample had a mean equal to 28.52 (SD=8.52), and the female sample had a mean of 26.52 (SD=45.28).

In the **experimental condition (computer suggestion),** there were 101 participants. The sample had a mean age of 22.85 (SD=3.014) ranging from 19 to 31 years old. In the present sample, 60 were women, and 41 were men. The average age of the male sample had a mean equal to 27.32 (SD=3.68), and the female sample had a mean 26.20 (SD=4.57). Participants were recruited by word of mouth, and posting advertisements on dedicated pages on social networks.

4.3. Experimental Procedure

Each participant has been tested individually with an online form. In the beginning, the participants were informed about the study with a cover story to not influence the results of the study. If participants agreed, they accepted the initial informed consent. First, they were asked to fill the pre-test questionnaire, about numeracy and attitude towards technology (only in machine learning suggestion). Then, they were assigned to either the human or the machine(s) condition. The first task was then administered, with the possibility to accept or reject the presented suggestion. After the first scenario, the participants filled the first scale of counterfactuals. Finally, a second analogous task was administered with the presentation of the final scale about counterfactual emotions.

In the end, a debriefing with the actual purpose of the study is presented, and a second informed consent was accepted by the participant.

4.3.1. Suggestion manipulation

Each task consisted of deciding whether to buy a flight ticket now or wait a few days hoping that the price will go down.

Example of a task for the *human expert* condition:

"Today is March 1, 2017. Your task is to buy a ticket from Milano Malpensa to New York for March 15, 2017. The best price available today is 1050 euros. Do you want to buy it now or do you want to wait?

"<u>According to many professional travel agents</u>, the best time to buy a ticket is two weeks in advance of the flight, although in some cases it is possible for last-minute tickets; (purchased one or two days before departure) are even cheaper but it is a risky choice if the departure date is not flexible.

Example of a task for the *computer expert* condition:

"Today is March 1, 2017. Your task is to buy a ticket from Milano Malpensa to New York for March 15, 2017. The best price available today is 1050 euros. Do you want to buy it now or do you want to wait?

"According to the Machine Learning system, the best time to buy a ticket is two weeks in advance of the flight. Based on historical data, this forecast has a 80% probability of being corrected, although in some cases it is possible for last minute tickets; (purchased one or two days before departure) are even cheaper but it is a risky choice if the departure date is not flexible."

Both tasks were analogous and counterbalanced to prevent order effect. That is to say, if in the first task the participants had to decide whether to buy a flight ticket from Milano Malpensa to New York in the second task they read instruction related to buying a flight ticket from Milano Malpensa to Madrid with the same percentage of money loss in the negative feedback.

4.4. Materials

To collect the possible confounder variables a brief version of the Big5 personality questionnaire was administered (Gosling, Rentfrow, & Swann, 2003). In addition, measures of locus of control of behavior (Craig, Franklin, & Andrews, 1984), objective numeracy (Lipkus, Samsa, & Rimer, 2001), subjective numeracy (Fagerlin et al., 2007) were collected. Only in the computer condition, participants had to fill out a questionnaire measuring the attitude towards technology (Elias, Smith, & Barney, 2012). The counterfactual emotions of regret and disappointment have been measured adopting the Regret and Disappointment Scale by Marcatto and Ferrante (2008) In these results, it was not calculated a regret index and a disappointment index between external attribution and disappointment and internal attribution and regret because the collected data differ from Marcatto et al.'s hence the Cronbach's alpha is not strong enough (lower than 0.6) to use these measures as indexes. Marcatto & Ferrante used the RDS to assess two different scenarios, one for regret and one for disappointment, while the presented experimental design does not have this difference. According to these reasons, it was decided not to use regret and disappointment indexes as reported by Marcatto & Ferrante (2008). This questionnaire was slightly modified for this study, and the item on the affective reaction was removed. The reason to choose this questionnaire was to assess regret, disappointment, internal and external attribution with not many items to prevent participants' abandonment due to tiredness.

Objective numeracy and attitude towards technology scales were assessed because these dimensions can influence the reliance on algorithms as found for numeracy in Logg (2016) and attitude towards technology in Venkatesh & Davis (2000).

4.4.1. Dependent variables:

- **Implicit Trust** (implicit measure): how many times the participants followed the suggestions (0=never, 1=only the trust time, 2=only the second time, 3=both) (J. D. Lee & See, 2004);

- **Regret** (1 item): refers to the extent the participants wish to have made a different choice (Likert 7 points: 1: totally disagree; 7: totally agree) (Marcatto & Ferrante, 2008)
- **Disappointment** (1 item): refers to the extent participants felt that the result of their choice was beyond their control (Likert 7 points: 1: totally disagree; 7: totally agree) (Marcatto & Ferrante, 2008)
- **Internal attribution** (1 item): refers to the extent which participants felt responsible for the outcome (Likert 7 points: 1: totally disagree; 7: totally agree) (Marcatto & Ferrante, 2008)
- **External attribution** (1 item): refers to the extent participants felt that the events out of their control were the cause for the bad outcome (Likert 7 points: 1: totally disagree; 7: totally agree) (Marcatto & Ferrante, 2008).
- **Control item** (1 item): refers to the satisfaction of the wrong outcome (Likert 7 points: 1: totally disagree; 7: totally agree) (Marcatto & Ferrante, 2008).

4.4.2. Independent variables

- **Conditions**: human expert vs computer suggestion

4.4.3. Confounding variables

- **Demographic**: (age, gender)
- **General Objective Numeracy scale** (3 items): This measure refers to the general comprehension and use of simple percentages and basic calculations (coded as 0 to 3 according to the correctness of the answers) (Lipkus et al., 2001).
- **Attitude towards technology** (3 items): this measure was collected only in computer suggestion condition to collect measures on the perception of the technology and was adapted from Elias, Smith, & Barney (2012).
- **Locus of control of behavior** (17 items): scale to measure the locus of control of participants (Likert 6 points: 0: strongly disagree; 5: strongly agree) (Craig et al., 1984)
- **Extraversion** (2 items): refers to the extent to which participants are extraverted, enthusiastic (Likert 7 points: 1: strongly disagree; 7: agree strongly) (Gosling et al., 2003)
- **Agreeableness** (2 items): refers to the extent to which participants define themselves as sympathetic (Likert 7 points: 1: strongly disagree; 7: agree strongly) (Gosling et al., 2003)
- **Conscientiousness** (2 items): refers to the extent to which participants are dependable (Likert 7 points: 1: strongly disagree; 7: agree strongly) (Gosling et al., 2003)
- **Emotional Stability** (2 items): refers to the extent to which participants are calm and emotionally stable (Likert 7 points: 1: strongly disagree; 7: agree strongly) (Gosling et al., 2003)
- **Openness to Experiences** (2 items): refers to the extent to which people are open to new experiences (Likert 7 points: 1: strongly disagree; 7: agree strongly) (Gosling et al., 2003)

5. Results

The present section describes the findings obtained for the dimensions of regret, disappointment, internal attribution, and external attribution in conditions. All dimensions were collected both after the first, and second trials. In particular, the first section describes the descriptive statistics (means and standard deviations) of the dependent variables on the whole sample and for each condition.

Then the role of trust in the two experimental conditions is presented for any dependent variable. These analyses had the aim of understanding the role of trust in a wrong suggestion with the dimensions of regret, disappointment, internal and external attribution. Finally, linear mixed models are presented to evaluate the underlying dimensions that affect the counterfactual emotions and the attribution of responsibility for the wrong outcome.

5.1. Descriptive statistics

Table 1 to Table 6 report descriptive statistics of dependent variables in each condition.

Table 1 Means (and standard deviations) of the four dimensions in RDS scale scores and trust by gender, age group, locus of control, personality traits, objective numeracy score.

	Regret1	Disapp1	Internal Att. 1	External Att. 1	Regret2	Disapp2	Internal Att. 2	External Att. 2
Gender (p value)	0.001***	0.51	0.70	0.25	0.29	0.82	0.55	0.75
Female (n= 110)	4.39 (1.84)	3.72 (1.88)	3.827 (1.95)	4.49 (1.749)	3.645 (1.99)	3.55 (1.90)	3.71(2.0)	4.14(1.9)
Male (n= 68)	3.441 (1.96)	3.912(1.9)	3.72 (2.26)	4.2(1.75)	3.32 (1.94)	3.63 (2.01)	3.54(2.25)	4.22 (2.01)
Age (p value)	0.33	0.27	0.93	0.52	0.39	0.27	0.55	0.31
<20 (n= 7)	5.14 (0.69)	3.57 (1.81)	4.42 (2.07)	4.57 (1.61)	4.14 (1.43)	4.42 (1.99)	4.28 (2.36)	5 (1.419
20-30 (n= 124)	4.04 (1.92)	3.91 (1.85)	3.76 (2.04)	4.41 (1.72)	3.57 (1.93)	3.62 (1.90)	3.63 (1.99)	4.08 (1.90)
30-40 (n= 44)	3.72 (2.07)	3.40 (1.98)	3.75 (2.14)	4.18 (1.85)	3.27 (2.13)	3.29 (2.01)	3.52 (2.34)	4.16 (2.13)
40-50 (n= 2)	5.00 (1.41)	3.5 (0.7)	4.00 (4.42)	5 (1.41)	2.5 (2.12)	3 (1.4)	5.5 (2.12)	5.5 (2.12)
50-60(n= 0)								
>60(n= 1)	6.00 (na)	7 (na)	3.00 (na)	7 (na)	6 (na)	7 (na)	2 (na)	7 (na)
Personality Traits								
Extraversion (p value)	0.11	0.16	0.83	0.97	0.79	0.99	0.10	0,1
Low (n=79)	4.29 (1.88)	4.01 (1.83)	3.8 (2.03)	4.39 (1.69)	3.56 (1.98)	3.58 (1.92)	3.49 (1.99)	3.92 (1.85)
High (n=99)	3.81 (1.96)	3.61 (1.91)	3.75 (2.11)	4.37 (1.80)	3.48 (1.97)	3.58 (1.96)	3.77 (2.17)	4.37 (2.00)
Agreeableness (p value)	0.24	0.95	0.60	0.55	0.06	0.97	0.9	0,99
Low (n= 78)	3.83 (1.98)	3.8 (1.90)	3.7 (2.16)	4.29 (1.80)	3.2 (1.92)	3.59 (2.02)	3.32 (2.16)	4.15 (2.10)
High (n=100)	4.18 (1.89)	3.78 (1.87)	3.85 (2.01)	4.45 (1.71)	3.77 (1.98)	3.58 (1.88)	3.9 (2.01)	4.19 (1.83)
Conscientiousness (p value)	0.9	0.97	0.93	0.24	0.99	0.16	0.14	0,15
Low (n=63)	4.05 (1.96)	3.77 (1.71)	3.76 (1.84)	4.61 (1.59)	3.51 (2.08)	3.84 (1.90)	3.65 (2.06)	4.47 (1.83)
High (n=115)	4.02 (1.92)	3.8 (1.97)	3.8 (2.19)	4.25 (1.82)	3.53 (1.91)	3.44 (1.95)	3.64 (2.12)	4.00 (2.00)
Emotional Stability (p value)	0.38	0.44	0.64	0.54	0.43	0.90	0.8	0,81
Low (n=110)	3.93 (1.99)	3.7 (1.87)	3.85 (2.1)	4.31 (1.76)	3.42 (1.95)	3.60 (1.95)	3.65 (2.12)	4.19 (2.01)
High (n=68)	4.19 (1.83)	3.9 (1.89)	3.69 (2.05)	4.48 (1.74)	3.67 (2.00)	3.56 (1.92)	3.63 (2.06)	4.14 (1.85)
Openness (p value)	0.06	0.02*	0.20	0.89	0.09	0.06	0.68	0,69
Low (n= 89)	4.30 (1.88)	4.11 (1.78)	3.97 (1.95)	4.37 (1.76)	3.29 (1.90)	3.84 (1.89)	3.76 (1.99)	4.13 (1.81)
High (n=89)	3.75 (1.95)	3.47 (1.93)	3.59 (2.17)	4.39 (1.75)	3.75 (2.01)	3.32 (1.96)	3.52 (2.2)	4.21 (2.08)
Objective Numeracy Score (p value)	0,38	0,05	0,01*	0,59	0,35	0,04*	0,1	0,72
0 (n= 10)	3.10 (1.73)	2.4 (1.65)	3 (2.21)	4.5 (2.12)	3 (2.21)	2.3 (1.89)	2.9 (2.2)	3.7 (2.26)
1 (n= 36)	4 (1.75)	3.47 (1.84)	4.78 (1.9)	4.1 (1.43)	3.17 (2)	3.33 (2.01)	4.14 (2.19)	4.19 (2.01)
2 (n= 58)	3.966 (2.00)	3.93 (2.02)	3.65 (2.13)	4.45 (1.86)	3.57 (2.1)	3.5 (1.96)	3.93 (2.16)	4 (2.15)
3 (n= 74)	4.22 (1.99)	4.03 (1.76)	3.51 (1.96)	4.45 (1.78)	3.73 (1.86)	3.9 (1.84)	3.28 (1.92)	4.36 (1.7)

(*p*) = for each comparison an indication of the *p*-value. ns = non-significant.

Table 2 distribution of trust in the two conditions with percentages

Frequency and percentage (italic) of trust in human and computer conditions

	Trust in the suggestions				
	Never Trust (0)	Only in the first task (1)	Only in the second task (2)	Both task (3)	***Total***
Human	5 *(2.7%)*	8 *(4.4%)*	11 *(6.2%)*	53 *(30%)*	77 *(43.3%)*
Computer	7 *(3.9%)*	9 *(5%)*	16 *(9%)*	69 *(38.8%)*	101 *(56.7%)*
Total	12 *(6,6%)*	17 *(9.4%)*	27 *(15.2%)*	122 *(68.8%)*	178 *(100%)*

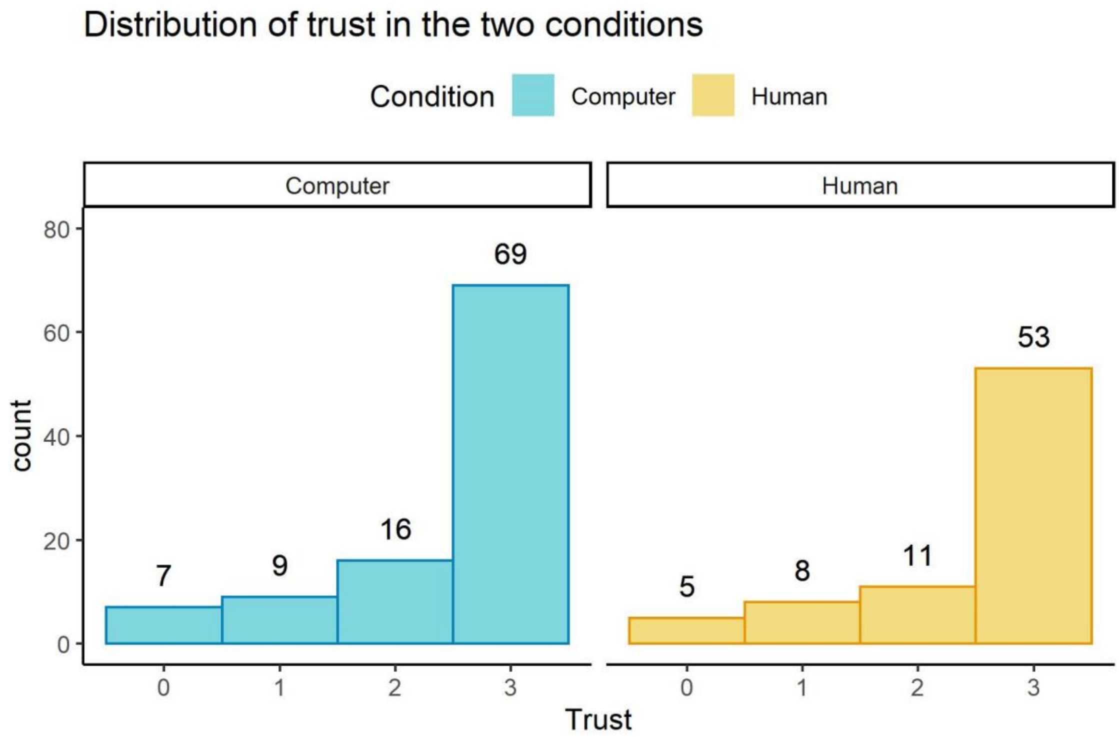

Figure 1 Distribution of trust in the two conditions

Results showed that there are no significant differences between the two conditions in the trust participants had in the suggestion.

Table 3 Means and standard deviation on regret between conditions

		M	SD	# Subj
Human	REGRET1	3.91	2.12	77
	REGRET2	3.45	2.1	
Computer	REGRET1	4.12	1.78	101
	REGRET2	3.57	1.86	

Table 4 Means and standard deviation on disappointment between conditions

		M	SD	# Subj
Human	DISAPP1	3.57	1.96	77
	DISAPP2	3.40	1.96	
Computer	DISAPP1	3.96	1.82	101
	DISAPP2	3.72	1.92	

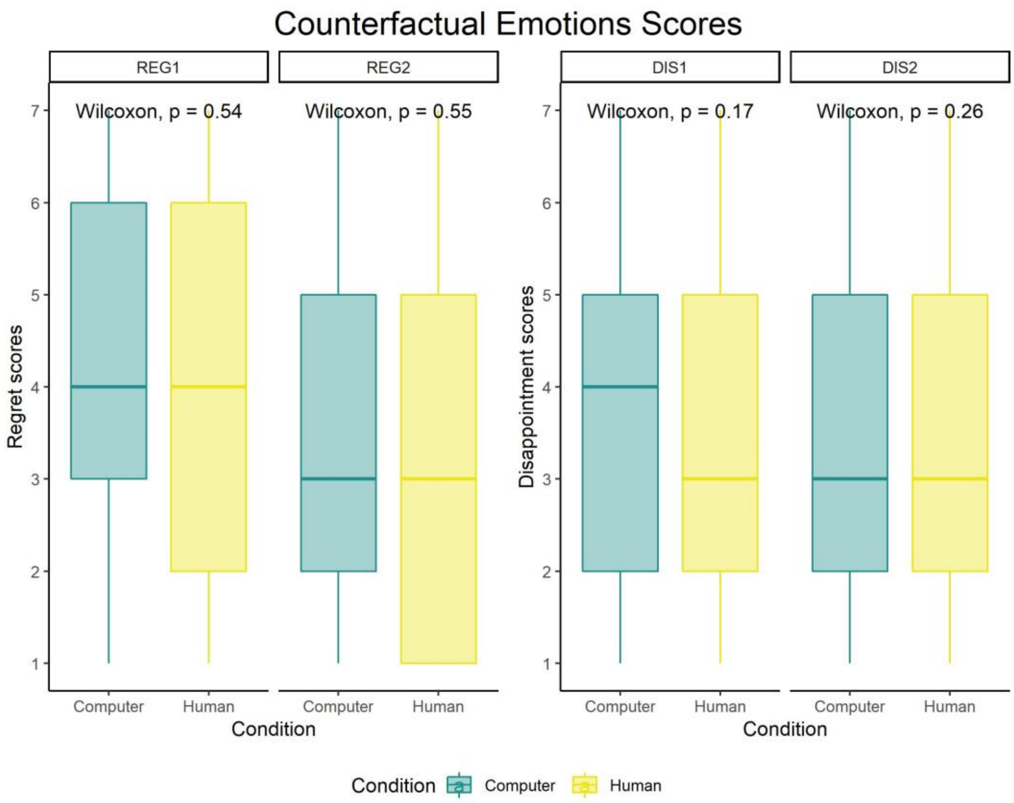

Figure 2 Boxplots of the counterfactual emotions (regret on the left) and the two conditions (human and computer). Lower and upper box boundaries 25th and 75th percentiles

Table 5 Means and standard deviation on internal attribution between conditions				
		M	SD	# Subj
Human	INTATT1	3.58	2.15	77
	INTATT2	3.72	2.23	
Computer	INTATT1	3.94	2.00	101
	INTATT2	3.58	1.98	

Table 6 Means and standard deviation on external attribution between conditions				
		M	SD	# Subj
Human	EXTATT1	4.39	1.87	77
	EXTATT2	4.18	2.03	
Computer	EXTATT1	4.37	1.67	101
	EXTATT2	4.16	1.89	

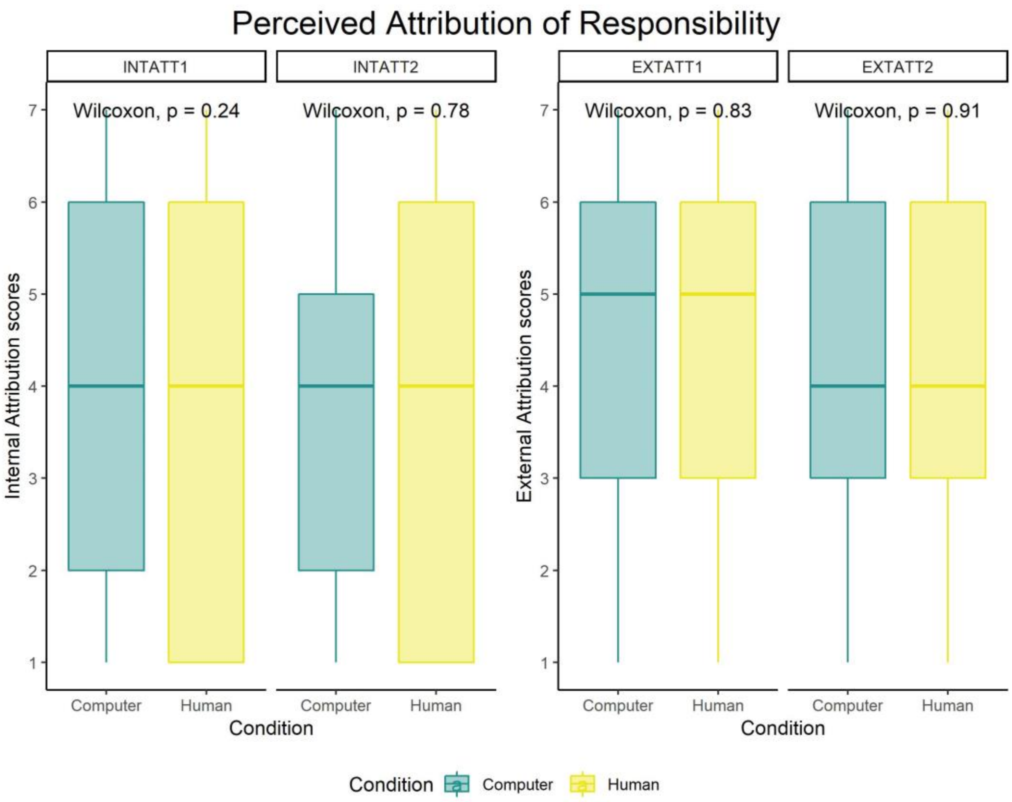

Figure 3 Boxplots of the attribution of responsibility (internal on the left) and the two conditions (human and computer). Lower and upper box boundaries 25th and 75th percentiles

5.2. The role of trust on the suggestion on the dependent variables

The plots presented below show how any DV is affected by the reliance on the suggestion. Trust is labeled as 0 for participants who never followed the advice; 1 for participants who followed the suggestion only in the first task; 2 for participants who followed the suggestion only in the second task; 3 for participants who followed the suggestion both times.

In these analyses, all the dependent variables have been taken separately as regards the first and the second choice.

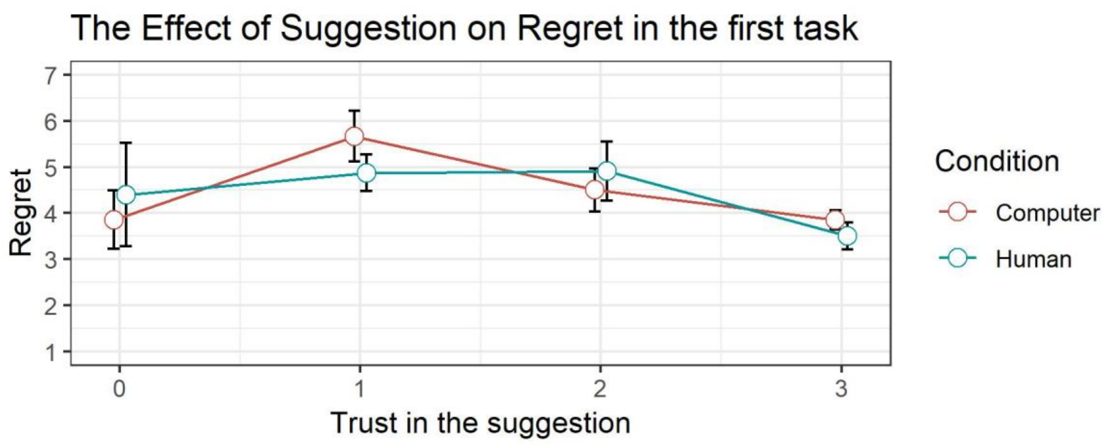

Figure 4 The effect of suggestion on regret in the first task

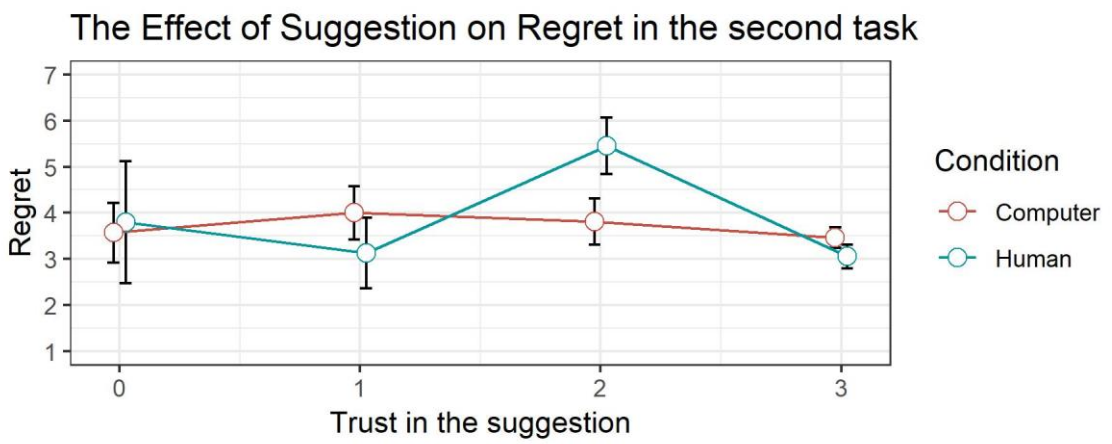

Figure 5 The effect of suggestion on regret in the second task

For regret in the first choice, results showed that there is no significant difference between the two conditions (Trust=0, p=0.68; Trust=1, p=0.26; Trust=2, p=0.61; Trust=3, p=0.33). It appears that the first decision provoked almost the same level of regret when participants followed a bad suggestion given by a human or a computer.

The analysis of the second task showed that when participants relied only the second time on the suggestion (i.e., they trust the source of advice), users who received a wrong suggestion by human experts (M=5.45, sd=2.02) felt more regret compared to who trusted twice the advice of computer (M=3.81, sd=2.00, p=0.049).

There are no differences between the scores of regret in the two conditions when respondents never trusted the advice (p=0.88) or when they trusted twice (p=0.23).

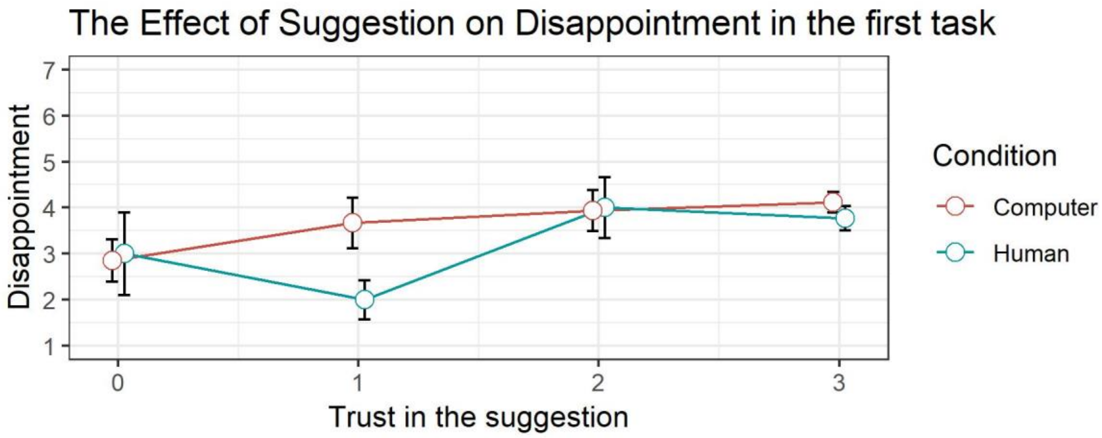

Figure 6 The effect of suggestion on disappointment in the first task

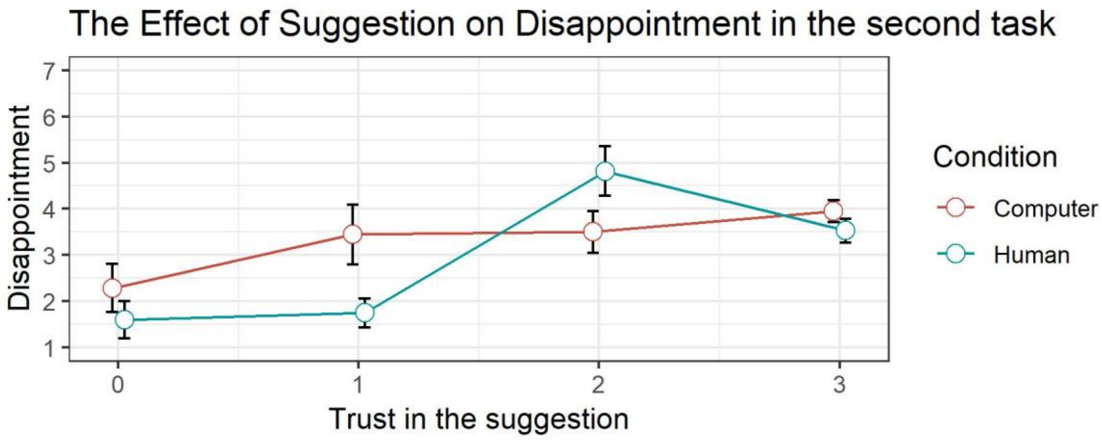

Figure 7 The effect of suggestion on disappointment in the second task

As regards disappointment in the first task, the main difference is shown by the trend in which participants felt more disappointment after a bad outcome with an intelligent system compared to human only in the first choice. An independent-samples t-test showed a non-significant difference between human (M=2; sd=1.20) and computer (M=3.67; sd=1.65) conditions, p=0.30.

In the second task, respondents showed a difference as regards how many times people chose the advice or not. When people followed the suggestion only in the first task, they felt more disappointment following a computer (M=3.44; sd=1.94) compared to those who followed human advice (M=1.75; sd=0.88, p=0.037). However, even if it is not significant, there is a trend showing participants who trusted the suggestion only in the second task they felt less disappointment while the suggestion came from a computer (M=3.5; sd=1.77) compared to a human (M=4.82; sd=1.78, p=0.07).

As previously found for regret, participants who never trusted the suggestion and participants who relied entirely upon the suggestion showed comparable scores in both conditions.

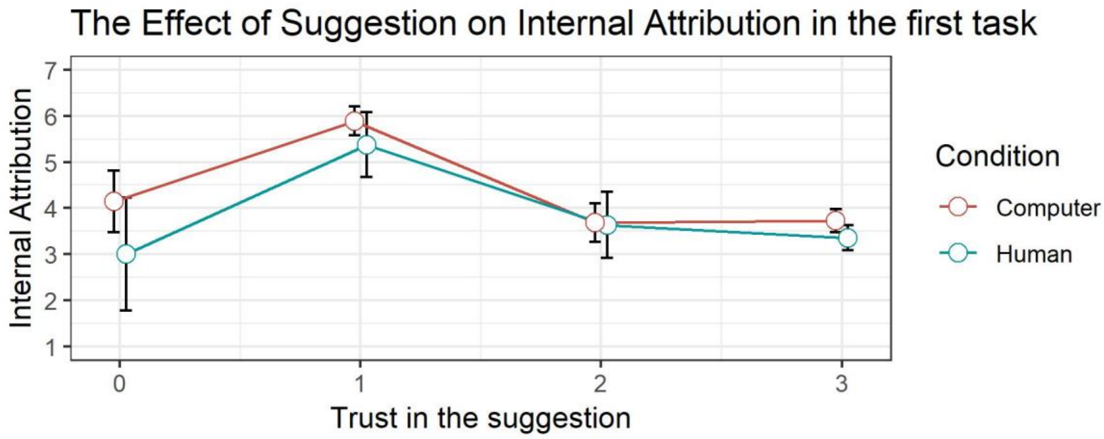

Figure 8 The effect of suggestion on internal attribution in the first task

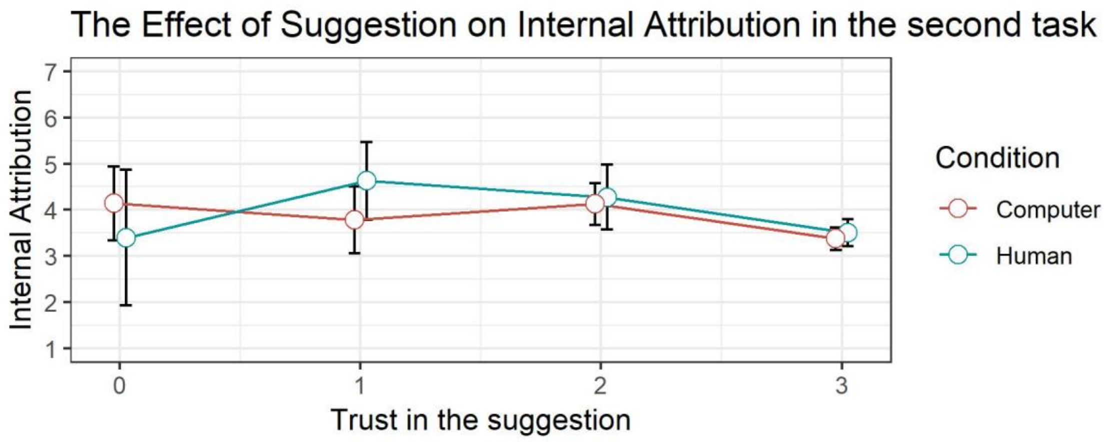

Figure 9 The effect of suggestion on internal attribution in the second task

The perceived internal attribution of blame after the first wrong outcome did not show any variation between the two sources of the suggestion, and it seems to follow the same trend concerning the trust in the suggestion.

Scores after the second bad outcome seemed to have the same trend more related to trust than to the source of suggestion.

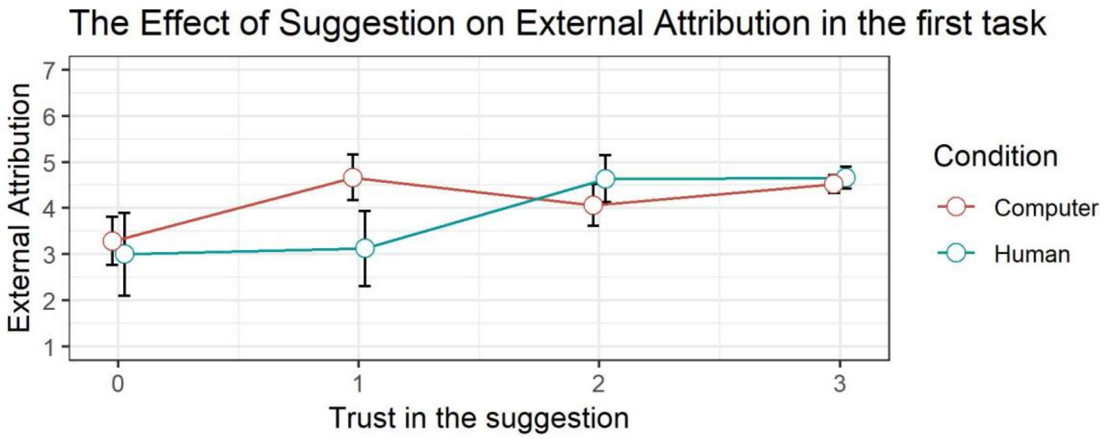

Figure 10 The effect of suggestion on external attribution in the first task

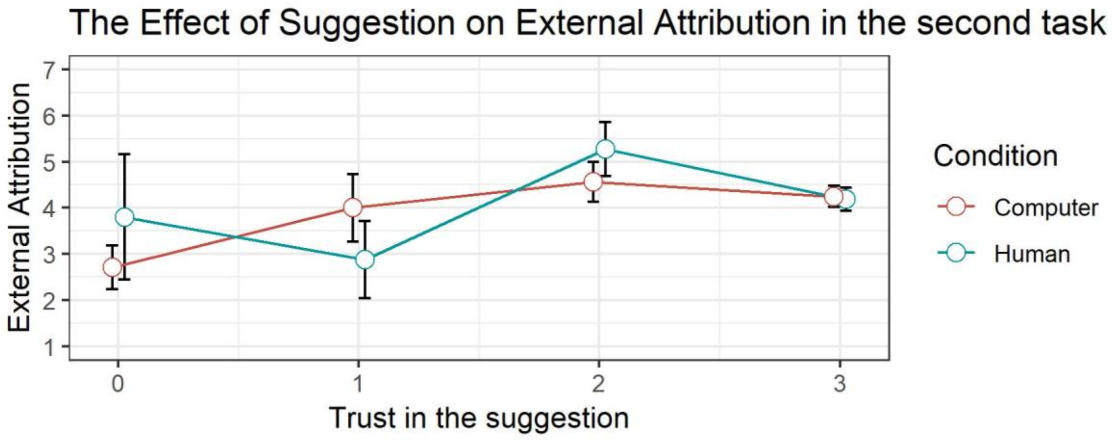

Figure 11 The effect of suggestion on external attribution in the second task

Concerning the scores of external attribution in the first task, the resulting trends seemed similar in both conditions. The most relevant difference is in the scores of external attribution after following the advice in the first task. People seemed to blame more the suggestion given by the computer compared to the suggestion given by a human.

After the second trial, respondents assigned more external attribution to the error of human experts if they trusted twice the suggestion, while if they trusted the suggestion only during the first task, they blamed more the "intelligent system."

In the next section, liner-mixed models have been performed to assess the role of confounding variables in the scores of regret, disappointment, internal and external attribution.

5.3. Hierarchical analysis of variance for Linear Mixed-Effects Models

For the explorative analysis, linear mixed-effects models[2] had been calculated to assess the relationship between regret values (in the first and second task) and the two conditions, adding the influence of the other measured confounding variables.

Initially, the exploration of the effects of the role of the suggestion given (human vs. computer) is presented comparing to regret scores in t1 and t2, and this analysis followed a bottom-up approach to get the best predictors involved in the regret scores.

As fixed effects, the first model had trust in the suggestions, objective numeracy score and locus of control (without interaction term) into the model. As random effects, intercepts for subjects random slopes for the effect of the condition is added. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-value was obtained by likelihood ratio tests of the full model (m5) with the effects in question against the model without the effects in question (m0).

The analysis specified four models of decreasing complexity concerning the parameterization of the subject-related variance/covariance matrix. The simple model is a dependent variable-condition model. The *lmer* specification for this model is:

```
m0=lmer(dependent var ~ Condition + (1|Subj))
```

The final model requests the addition of other parameters through an iterative process. This model estimates parameters for the four components trust, time (t1, t2), objective numeracy and locus of control.

The *lmer* specification for this model is:

```
m4=lmer(dependent val~Condition+Trust+Time+ObjectiveNumeracy+LOC+(1|Subj))
```

In detail, as regards regret, disappointment and external attribution, the model that better fits considers the time of experience (first and second task), objective numeracy score, locus of control.

---

2 These analyses have been developed with R Studio software (R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018)(R Core Team, 2018) (R Core Team, 2018)((Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2018) (Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)(Bates et al., 2015)

In this section, it will be presented the specific model used that best explains the considered dependent variable.

Any considered model comprehends the random effect of individual variability of intercept-values (1|subj).

The considered models are:

- m4: the model comprehends fixed effects of the condition, time of experience, numeracy and locus of control and random effect (~condition + trust in the suggestion + time of experience + objective numeracy + locus of control + (1|subj))
- m3: the model comprehends fixed effects for condition, time of experience and objective numeracy and random effect (~condition + trust in the suggestion + time of experience + objective numeracy + (1|subj))
- m2: the model considers fixed effects for condition, time of experience and random effect (~condition + trust in the suggestion + time of experience + (1|subj))
- m1: this model takes into account fixed effects for condition and trust in the suggestion with random effect (~condition + trust in the suggestion + (1|subj))
- m0: the simplest model comprehends only condition and random effect (~condition+ (1|subj))

The resulting LMEMs are compared beginning from the more complex and moving to the simpler ones. In case two models significantly fit the data, the simplest model was preferred.

5.3.1. Regret

The ANOVA model comparison of regret values in model4 displays a significant difference with other models, and m4 is, therefore, the best model exploring the data.

Regret values are not explained only for different conditions; the source of suggestion does not systematically vary regret value. As shown in Table 7, these analyses showed that the variables of condition, trust in the suggestion, objective numeracy, and locus of control significantly explain the participants' expressed regret compared to the model without these effects.

Model4 is chosen as the best fit to explain the date, since it is preferable over m3, m2, m1, and m0.

Table 7 Compared linear mixed models explanatory analysis models of regret

Models:

model0: RegretScores ~ Condition + (1 | Subjects)

model4: RegretScores ~ Condition + Trust+ Time+ Objective numeracy + Locus of control +(1 | Subjects)

	Df	AIC	BIC	LLR	Deviance	χ^2	Df	p-value
model0	4	1456.3	1471.8	-724.17	1448.3			
model4	10	1433.9	1472.7	-706.96	1413.9	34.423	6	>0.01 *

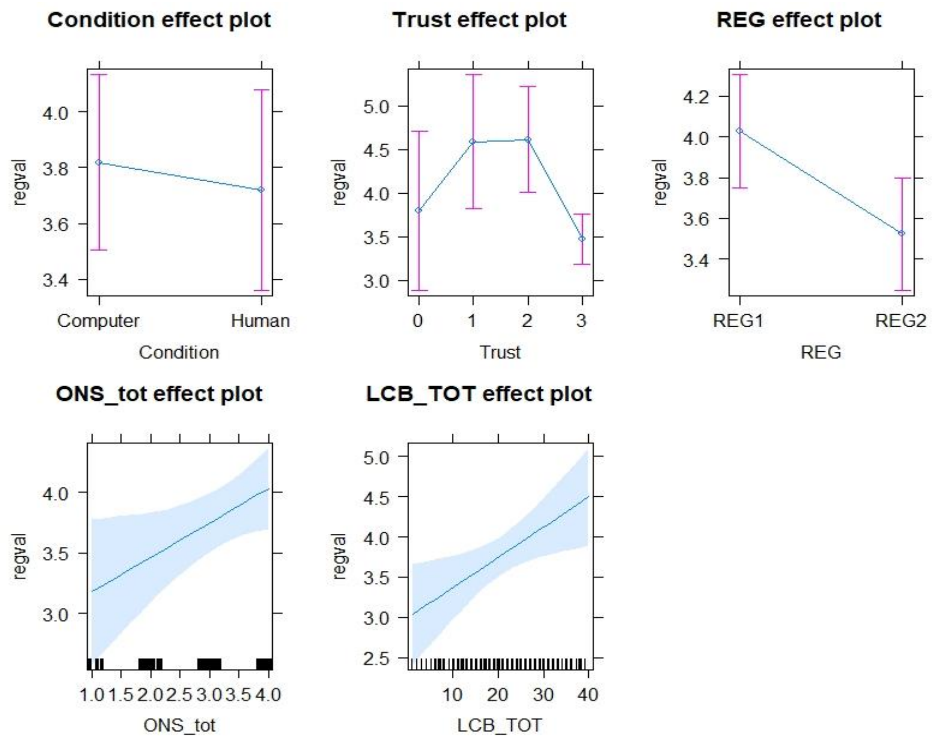

Figure 12 Linear mixed models results of regret scores

5.3.2. Disappointment

In each analysis, the best-fitting model was chosen by comparing models whose random-effects structure had a different degree of complexity. For each pair of models, the results of the likelihood ratio test were applied to evaluate whether the reduction of additional fixed-effects parameters provided a better fit of the model to the data.

As found for regret dimension, disappointment is not only explained by the different source of the suggestion. The best model that fits the data is model4, in which it has been taken into account how many times the participants trusted the advice, the time of the experience, the objective numeracy, and locus of control.

In Table 8 is presented the comparison between the simplest model and the model5 that better fit disappointment scores.

Table 8 Compared linear mixed models explanatory analysis models of disappointment

Models:

model0: DisappointmentScores ~ Condition + (1 | Subjects)

model4: DisappointmentScores ~ Condition + Trust+ Time+ Objective numeracy + Locus of control +(1 | Subjects)

	Df	AIC	BIC	LLR	Deviance	χ^2	Df	p-value
model0	4	1433.6	1449.1	-712.80	1425.6			
model4	10	1395.3	1434.1	-687.66	1375.3	50.293	6	>0.01 *

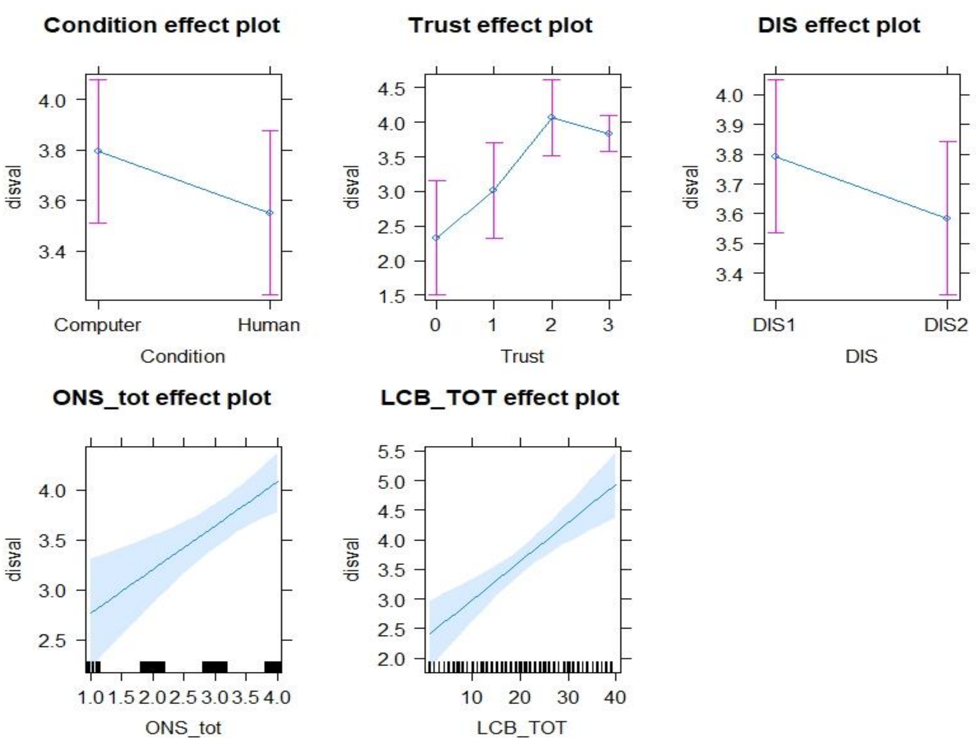

Figure 13 Linear mixed models results of disappointment scores

The formula of the model that best fit disappointment is the following:

```
m4= lmer(Disapp scores~Condition+Trust+Time+Numeracy+Locus of control+(1|Subj))
```

5.3.3. Internal attribution

Internal attribution results showed that the best-fitting model was model1, which incorporates trust in the suggestion and the random-effect for subjects. This model was found more explicative compared to more complex models evaluating the likelihood ratio.

In Table 9 is presented the comparison between the simplest model and the model1 that better fit internal attribution scores.

Table 9 Compared linear mixed models explanatory analysis models of internal attribution

Models:
model0: IntAttScores ~ Condition + (1 | Subjects)
model1: IntAttScores ~ Condition + Trust +(1 | Subjects)

	Df	AIC	BIC	LLR	Deviance	χ^2	Df	p-value
model0	4	1461.4	1476.9	-726.68	1453.4			
model1	49	1443.2	1633.0	-672.59	1345.2	108.18	45	>0.01 *

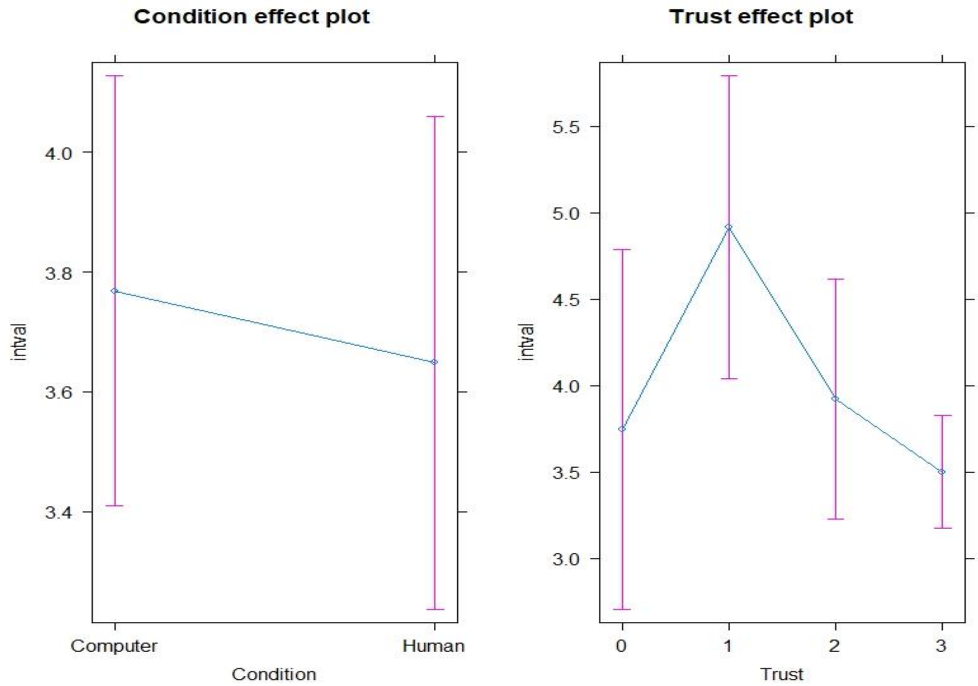

Figure 14 Linear mixed models results of internal attribution scores

The formula of the model that best fit internal attribution is the following:

```
m1=lmer(int attr. scores ~  Condition + Trust +(1|Subj))
```

5.3.4. External attribution

As found previously for regret and disappointment, the ANOVA model comparison showed that the best-fitting model that better explains external attribution was model4 that includes besides the source of suggestion, trust in the suggestion, time of experience, numeracy, and locus of control.

In Table 10 is presented the comparison between the simplest model and the model5 that better fit disappointment scores.

Table 10 Compared linear mixed models explanatory analysis models of external attribution

Models:

model0: ExtAttScores ~ Condition + (1 | Subjects)

model5: ExtAttScores ~ Condition + Trust+ Time+ Objective numeracy + Locus of control +(1 | Subjects)

	Df	AIC	BIC	LLR	Deviance	χ^2	Df	p-value
model0	4	1405.2	1420.8	-698.62	1397.2			
model5	10	1390.6	1429.4	-685.31	1370.6	26.624	6	>0.01 *

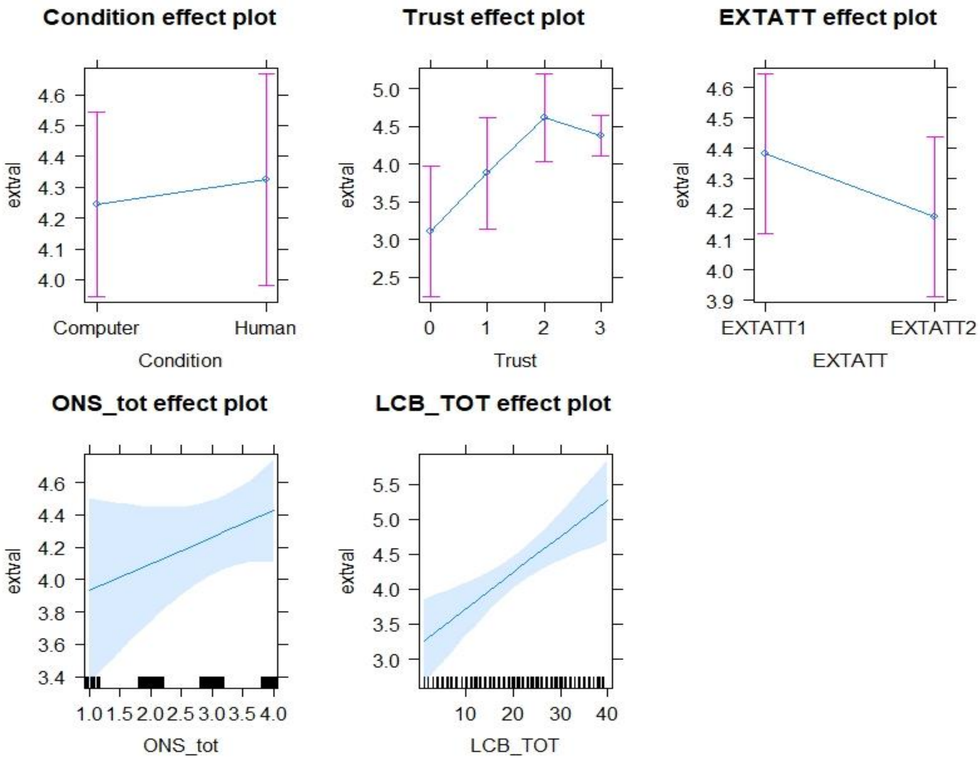

Figure 15 Linear mixed models results of external attribution scores

The formula of the model that best fit external attribution score is the following:

```
m5=lmer(Ext att scores~Condition+Trust+Time+Numeracy+Locus of control+ (1|Subj))
```

5.4. Summary of Linear Mixed-Effects Models analysis

Overall the LMEMs' analysis shows that the only source of suggestion was not enough to explain the DVs scores and other variables play a role in the best fitting model.

Regret and disappointment acted almost in the same way as regards the differences in the two conditions, in the time of experience and their relationship with objective numeracy and locus of control. Both of these counterfactual emotions are slightly higher in human condition compared to computer condition. They decreased over time between as response of the first negative feedback compared to the second. Higher numeracy correlates with higher scores in regret and disappointment; as for locus of control, which higher scores (higher externality of locus of control) correlate with higher scores in both regret and disappointment.

The main difference can be found in the role of trust in the suggestion; while regret and disappointment are comparable when the participants did not trust in the suggestion, these two emotions differ when the participants trust only the first time or only the second time. As for regret, participants who trusted only the first suggestion had a higher score in regret compared to disappointment, while for the second task, people who trusted only in the second suggestion showed similar levels of counterfactual emotions. An important distinction can be found for participants who trusted both the suggestion; they showed a lower level of regret and a higher level of disappointment

Pertaining to the attribution of responsibility, internal and external attributions differ from the underlying factors that influence these dimensions.

Internal attribution is mainly explained by the source of the suggestion and the reliance on the advice. External attribution is defined by the source of the suggestion, the trust in the advice, the time of experience, numeracy, and locus of control.

Through a visual comparison between Figure 14 and Figure 15, one can notice the difference between the internal attribution trend and the external attribution. That is to say, internal attribution is higher in the computer condition compared to the human condition, while external attribution is higher in the human condition compared to the computer condition.

Even the role of trust in the system affects these variables differently. As regards internal attribution, participants who trusted only the first time in the suggestion perceived themselves more self-responsible for the bad outcome while participants who trusted only the second suggestion or both attributed the responsibility of the adverse outcome to events beyond their control.

6. Discussion and Limitations

The main goal of the current study was to investigate whether the dimensions of regret, disappointment, internal and external attribution behaved differently in the function of a suggestion provided by some human experts or a machine learning system. Further on, the role of the trust in the suggestion, objective numeracy, and different levels of locus of control was observed.

The results showed that the counterfactual emotions of regret and disappointment, as well as external attribution of responsibility, are affected by five main factors, that are: the source of the suggestion, the trust in the suggestion, the time of the experience, the ability to understand numbers measured by the objective numeracy, and the locus of control.

The dimension that is affected only by the source of the suggestion and the trust in it is the internal attribution of responsibility. Nevertheless, a model that takes into account the dimensions outlined above seems to be the most accurate to predict the values of counterfactual emotions during the interaction with an intelligent system.

These findings support the purpose of this research, which suggests that the source of suggestion has an impact on counterfactual emotions and attribution of responsibility. This study also makes a small contribution to establishing an evidence base for the confounding variables measured.

As regards the first hypothesis stated in section 2 the analysis showed that internal attribution is higher in the computer condition compared to the human condition. As a confirmation, the third hypothesis is confirmed by external attribution that is more related to the human suggestion compared to the computer advice.

The second hypothesis was about the higher level of regret while having a bad outcome after following a suggestion given by an "intelligent system." This second hypothesis is partially confirmed by trends in LMEMs, even if this difference is not significant per se and is affected by the dimensions of trust in the system, the time of experience, numeracy, and locus of control.

The third hypothesis was about the higher level of external attribution in the human suggestion condition. No significant differences were found between the two condition, even if LMER models showed a weak trend even if it needs to be confirmed by further studies.

The fourth hypothesis was about the higher level of disappointment in the human condition compared to the computer condition. This hypothesis was not confirmed, and trends in LMEMs showed contradictory results than expected.

The last hypothesis was about the higher trust in the computer suggestion in the first task compared to human suggestion, according to previous literature on reliance on the algorithm (Logg et al., 2019). Nevertheless, this hypothesis was not confirmed; however, the descriptive statistics for trust showed a slight trend in which participants trusted more algorithmic advice twice. Hence it seems people trust more algorithm advice, even if this trend needs more attention in the next studies.

This study had some limitations. First, the scores in the dependent variables appear to be quite low. There are two potential reasons for these results. One can be related to the nature of the scenario as referred by some participants at the end of the study. The task was about buying two flight tickets for work purposes, and some participants denoted that they did not care about the idea of saving money; hence, they preferred to buy the ticket as soon as possible to avoid the possibility of losing the flight. Another reason can be related to the use of the RDS scale. To prevent the abandonment of the study, in the design phase, it was decided to not collect the affective reaction item that investigates how sorry participants felt. This item was used by Marcatto & Ferrante (2008) to control the experience of counterfactuals subtracting it from the control item about how satisfied the participant was.

Another potential weakness can be related to the presence of a percentage of correctness in the computer scenario, which is not present in the human scenario. This decision was made to present to participants a plausible scenario, in which usually a suggestion given by a machine-learning system has a percentage of precision. This aspect could have affected the results in terms of trust and in terms of comparison between the two scenarios.

The next study overcomes these issues through the use of another purchase scenario validated in the previous study on regret and disappointment (Giorgetta, 2012) and taking into account the control measures used in the Regret and Disappointment Scale (RDS) (Marcatto & Ferrante, 2008).

Despite the limitations, these results create a first original input to study counterfactual emotions and their antecedents in the particular case in which the decision process ended with a wrong outcome with algorithm-based systems. This novel approach takes into account the emotional reactions of the users while facing wrong outcomes during the use of these complex systems. In particular, this first attempt to merge psychological literature on regret and disappointment during the interaction with intelligent systems. It is especially important to suggest to intelligent systems designers to take into account the users' emotional reaction to the potential incorrect suggestion an intelligent application can give. The prevention of the user experience of regret after a wrong outcome can significantly prevent the abandonment of the technology itself.

In the next chapter, study two investigated how participants experienced regret, disappointment, guilt and perceived responsibility in a typical purchase scenario from the psychological literature. The second experiment aimed at further investigating the specific role of counterfactuals in two scenarios without taking into account the implicit trust of the participants. The decision to avoid the measure of implicit trust was made to control and assess only the experience of counterfactuals without any variance related to participants' trust.

Chapter 3: Regret with Intelligent Systems predictions

As found by the previous study, counterfactual emotions are involved in the interaction with intelligent systems when the user faces a wrong outcome. This experiment was about the confirmation of the previous findings and is aimed at fixing the limitations encountered before.

The previous study found that the emotions of regret and disappointment, as well as the internal and external attributions, are affected by the source of the suggestion in the evaluation of a wrong outcome.

The design followed a previous experiment in psychological literature aimed at analyzing the attribution of responsibility, and the consequent counterfactual experience, in a purchase scenario.

In this experiment, it was chosen to overcome the issues found with the main dependent variables of the previous study changing the scale to assess the experienced counterfactual emotions. In this study, it was used the scale developed by Giorgetta (2012) maintaining the control measures used in Marcatto & Ferrante (2008) to check the actual experience of counterfactuals.

1. Motivation and aims

Several works are discussing that people abandon algorithm-based systems after seeing them err (e.g., Dietvorst et al., 2015; Logg, 2018). This phenomenon named "Algorithm Aversion" is due to the fact that users expect intelligent computer systems to behave perfectly, and they do not assume that different algorithms can have different accuracy rates, determining a percentage of success that almost always is far from 100%.

The central consequence of algorithm aversion is that people switch in a dichotomic way from entirely relying on the computer-based suggestion to go back completely relying on human advice, not taking advantage anymore of the calculating power that intelligent systems have and humans do not.

The hypothesis at the basis of this study is that a cause of this behavior can be found in the emotional response users have while interacting with an intelligent system, and how it determines the behavioral switch caused by regret.

Hence, the purpose of this research is to understand the relationship between the users' experienced counterfactual emotions after following a suggestion resulting in a wrong outcome that could describe the algorithm aversion.

This information could advise intelligent systems experts and computer scientists to design better and more useful decision support systems, enhancing or not the sense of responsibility to make a decision more nudging-oriented or more reasoning-oriented, preventing the abandonment and support the cooperation between human and these new AI-based systems.

This cooperation is essential in order to understand what kind of information the systems should provide to improve the decision-making process and prevent user's regret and the subsequent abandonment of intelligent systems.

2. Hypotheses

As described in the previous section, the expectations were that suggestion provided by an intelligent machine (in a technical and relatively complicated decision task) decreases the possibility for the user to blame the advisor and, therefore, should produce a lower rating on counterfactual emotions and responsibility.

Consequently, after a wrong suggestion by an intelligent machine, it was postulated the following hypotheses:

• Hypothesis on regret: the user experiences more regret after a poor suggestion provided by an intelligent machine compared to a poor suggestion provided by a human being.

• Hypothesis on disappointment: the user experiences less disappointment after a poor suggestion provided by an intelligent machine compared to a poor suggestion provided by a human being.

• Hypothesis on responsibility: the user feels less responsible for the choice and the bad outcome than after a wrong suggestion given by a human being.

• Hypothesis on guilt: the user experiences less guilt than after a wrong suggestion given by a human being.

3. Equipment

The experiment was built on Qualtrics XM[3]. Qualtrics XM is an online platform to build surveys that can be easily incorporated in Amazon Mechanical Turk services.

[3] www.qualtrics.com

4. Method

4.1. Experimental Design

To address this topic, the participants read two similar scenarios in a 2 cells between-subjects design controlled by the source of the suggestion (human advice vs computer advice).

To explore the effect of experienced counterfactuals, the scenario is presented twice in an analogous manner in each condition.

The dimensions investigated are Regret, Disappointment, Responsibility, and Guilt. Every dimension has two items one related to the experienced emotion on the choice, the other is related to the outcome (e.g., *Choice*: "I feel responsible for the choice that was made"; *Outcome*: "I feel responsible for having a phone that does not meet my needs").

The scenario is a typical purchase scenario (Giorgetta, 2012, pp. 89–129) in which the person needs to buy a new phone, and the clerk or an algorithm-based system suggests another alternative.

4.2. Sample

In total, 147 participants took part in this study. 47 participants were excluded from the analysis due to the repeating occurrence of an I.P. address. 15 participants were excluded because they failed the attention check; the other 15 participants failed the control items about the experience of counterfactual emotions.

A final sample of 70 people was then analyzed. The sample had a mean age of 37.5 (SD=11.7) ranging from 22 to 73 years old. In this sample, 25 are women, and 45 are men. The average age of the female sample has a mean of 37.8 (SD=11.66), and the male sample has a mean equal to 36.9 (SD=11.75).

In the **control condition (human suggestion),** there are 43 participants. The sample had a mean age of 37.72 (SD=11.9) ranging from 22 to 73 years old. In this sample, 12 are women and 31 are men. The average age of the male sample has a mean equal to 38.48 (SD=12.47), and the female sample has a mean of 35.75 (SD=10.56).

In the **experimental condition (computer suggestion),** there are 27 participants. The sample had a mean age of 36.44 (SD=11.39) ranging from 23 to 66 years old. In the present sample, 13 are women and 14 are men. The average age of the male sample has a mean equal to 33.42 (SD=9.49), and the female sample has a mean of 39.69 (SD=12.72).

They were randomly assigned to one of the two conditions, with computer suggestions or human suggestions.

4.3. Experimental Procedure

Participants were recruited through the Amazon Mechanical Turk service. After accepting the informed consent, the participants read the first scenario and then filled the first questionnaire. Then the participants read the second analogous scenario and responded to the second questionnaire.

The "human suggestion" scenario was as follows:

"*Imagine that you have to buy a new smartphone because yours has just stopped working. Even if you are not an expert, you go to the shop with the idea to buy XY10, because you think that is the model that best suits your preferences. Once in the shop* ***, you explain your preferences to the clerk who suggests*** *you buy smartphone WLx at the same price. Hence you decide to buy the suggested model WLx. Some time later, you realize that model XY10 would have been better for your needs, while the smartphone you have bought does not meet your expectations*."

The scenario in the "computer suggestion" condition described the same situation but change the source of the suggestion and was as follows:

"*Imagine that you have to buy a new smartphone because yours has just stopped working. Even if you are not an expert, you go to the shop with the idea of buying XY10, because you think that is the model that best suits your preferences. Once in the shop,* ***you enter your preferences in an algorithm-based website that suggests*** *you buy smartphone WLx at the same price. Hence you decide to buy the suggested model WLx. Some time later, you realize that model XY10 would have been better for your needs, while the smartphone you have bought does not meet your expectations*"

After having read this first scenario, participants were asked to fill the questionnaire on responsibility, regret, disappointment, guilt.

After having answered the items, participants were asked to read the second scenario. For the human suggestion condition was:

"*After a year, your smartphone stops working and you have to buy a new one. You go back to the shop with the idea to buy B30, because you think that is the model that best suits your preferences. Once in the shop* ***, you explain your preferences to the clerk who suggests*** *you buy smartphone T25 at the same price. Hence you decide to buy the suggested model T25. Some time later, you realize that model B30 would have been better for your needs, while the smartphone you have bought does not meet your expectations*."

While the second scenario in the "computer suggestion" condition was as follows:

"*After a year, your smartphone stops working and you have to buy a new one you go back to the shop with the idea to buy B30, because you think that is the model that best suits your preferences. Once in the shop,* ***you enter your preferences in an algorithm-based website that suggests you*** *buy smartphone T25 at the same price. Hence you decide to buy the suggested model T25. Some time later, you realize that model B30 would have been better for your needs, while the smartphone you have bought does not meet your expectations*."

4.4. Materials

After having read the first scenario, participants filled the item related to the dimensions of *responsibility, regret, disappointment,* and *guilt.* Due to the sampling method (Amazon MTurk), it seemed necessary to add one question as an attentional check, two questions about affective reaction and control item and an item checking the attention in the suggestion given.

For the answers, participants were provided with a 10-points Likert scale, ranging from *complete disagreement (1)* to *complete agreement (10).*

In this study, there were two items for each dimension. The former item was about the feeling of the dependent variable on the choice, while the latter item was about how participants felt for the bad outcome.

The items were then considered singularly for each task and averaged to have scores about the experienced emotions on choice and outcome.

The attention check question was assessed at the beginning of the experiment, to check whether participants read or not the instructions, to prevent distracted answers.

Participants who had higher values in the control item compared to affective reaction were excluded from the analysis as a check of the experience of counterfactual emotions.

4.4.1. Dependent variables:

- **Responsibility on Choice** (1 item): refers to the extent the participants feel responsible about the choice made (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)
- **Responsibility on Outcome** (1 item): refers to the extent the participants feel responsible for the negative outcome (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)(Marcatto & Ferrante, 2008)
- **Regret on Choice** (1 item): refers to the extent the participants feel regret on the choice made (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)
- **Regret on Outcome** (1 item): refers to the extent the participants feel regret on the negative outcome (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)
- **Disappointment on Choice** (1 item): refers to the extent the participants feel disappointed about the choice made (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)
- **Disappointment on Outcome** (1 item): refers to the extent the participants feel disappointed about the choice made (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)
- **Guilt on Choice** (1 item): refers to the extent the participants feel guilty about the choice made (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)

- **Guilt on Outcome** (1 item): refers to the extent the participants feel guilty for the unsatisfactory outcome (Likert 10 points: 1: totally disagree; 10: totally agree) (Giorgetta, 2012)

4.4.2. Independent variables

- **Conditions**: human expert vs computer suggestion

4.4.3. Control variables

- **Affective reaction (1 item):** refers to the feeling of being sorry about what happened to the participant, or the degree about how much the participant feel (Likert 10 points: 1: totally disagree; 10: totally agree) (Marcatto & Ferrante, 2008))
- **Control item (1 item):** refers to the feeling of being satisfied with what happened (Likert 10 points: 1: totally disagree; 10: totally agree) (Marcatto & Ferrante, 2008)
- **Age and Gender**

5. Results

5.1. Descriptive statistics

The analyses were based on simple comparisons between dependent variables in the two conditions. No multiple comparisons were analyzed. In this section, the descriptive statistics for the dependent variables are presented. In the beginning, non-parametric analysis was carried out on the total sample of participants. In the second section, Mann-Whitney tests were used to test significant differences between dependent variables of responsibility, guilt, regret, and disappointment in the two conditions (human vs computer). In the last section, a summary of the will enlighten the significant results.

Table 11 Means (and standard deviations) of the four dimensions in RDS scale scores and trust by gender, age group, locus of control, personality traits, objective numeracy score.

	First task				Second task			
On choice	Responsibility1	Regret1	Disappointment1	Guilt1	Responsibility2	Regret2	Disappointment2	Guilt2
Gender (p)	ns	ns	ns	ns	Ns	ns	ns	ns
Female (n= 110)	8.16	9.4	9.4	6.76	8.44	9.44	9.48	6.5
Male (n= 68)	8.42	9.2	9.16	6.02	8.73	9.07	9.40	7.32
Age (p)	0.04*	0.04*	ns	ns	ns	0.004***	ns	ns
18-30 (n= 16)	9.31	9.62	9.3	7.37	8.88	9.31	9.25	7.81
31-45 (n= 39)	7.85	9.00	9.07	5.94	8.28	9.10	9.35	6.59
46-60 (n= 9)	8.55	9.44	9.44	6.44	9.44	9.22	9.67	7.44
>60 (n= 6)	8.5	9.83	9.83	5.33	9.00	9.5	10.00	4.5
On Outcome	Responsibility1	Regret1	Disappointment1	Guilt1	Responsibility2	Regret2	Disappointment2	Guilt2
Gender (p)	ns	ns	ns	ns	ns	ns	ns	ns
Female (n= 110)	8.28	9.4	9.5	6.7	8.36	9.36	9.44	6.84
Male (n= 68)	8.44	9.09	9.09	6.2	8.86	8.87	9.27	6.5
Age (p)	ns	ns	ns	ns	ns	ns	ns	ns
18-30 (n= 16)	9.12	9.37	9.37	7.69	9.13	9.00	9.13	7.63
31-45 (n= 39)	8.02	8.87	9.02	5.94	8.28	8.97	9.25	6.38
46-60 (n= 9)	8.33	9.78	9.55	6.33	9.22	9.11	9.56	7.33
60-… (n= 6)	8.83	10	9.83	5.66	9.33	9.5	10.00	4.5

(*p*) = for each comparison an indication of the *p*-value. ns = non-significant.

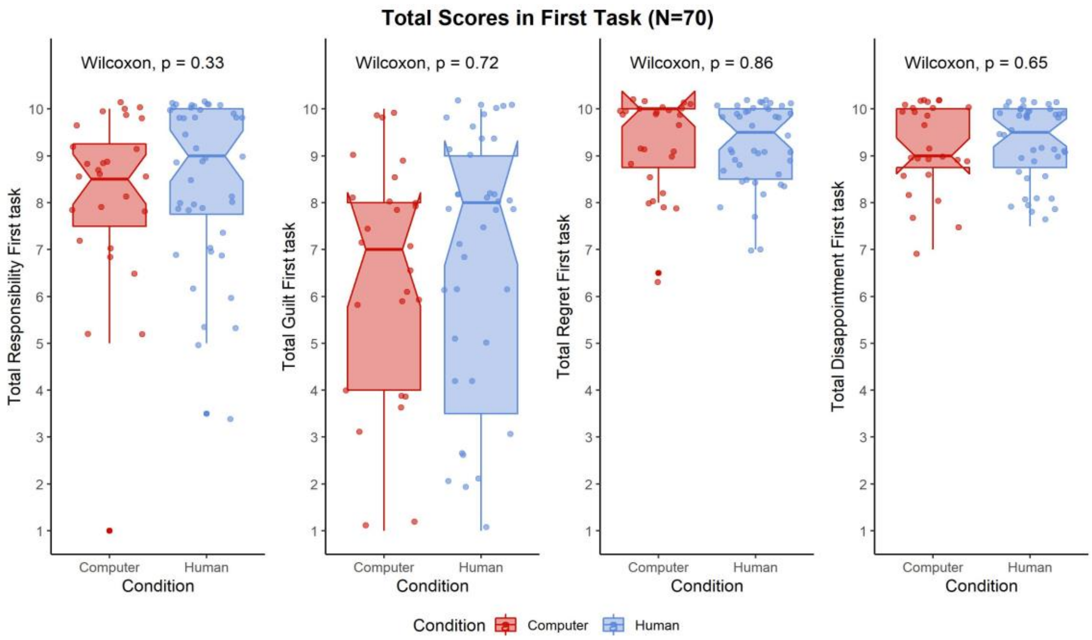

Figure 16 Total scores in the first task in human and computer condition by the dimensions of responsibility, guilt, regret, disappointment (N=70)

Table 12 Summary descriptive statistics on dependent variables in the first task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY TASK1	8.11	8.5	2.00	27
	GUILT TASK1	6.29	7	2.75	
	REGRET TASK1	9.21	10	0.96	
	DISAPP TASK1	9.17	9	0.93	
Human	RESPONSIBILITY TASK1	8.51	9	1.71	43
	GUILT TASK1	6.35	8	3.17	
	REGRET TASK1	9.23	9.5	0.88	
	DISAPP TASK1	9.29	9.5	0.82	

The medians of computer condition and the human condition were 8.5 and 9, respectively. To evaluate the difference in the responses of total responsibility in the first task, it was chosen a Mann-Whitney's U test. It was found a non-significant effect between the two groups (The mean ranks of computer suggestion and group human suggestion were 8.11 and 8.51, respectively; U = 501, p = 0.33).

Mann-Whitney's U test showed a non-significant difference for the dimension of total guilt in the first task. The scores of computer condition (Mdn=7) were not different compared to the human condition (Mdn=8), U= 551, p=0.72.

For total regret in the first task, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=9.5), U=595, p=0.86. Even for total disappointment scores in the first task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=9) condition and the human condition (Mdn=9.5), U=544.5, p=0.65.

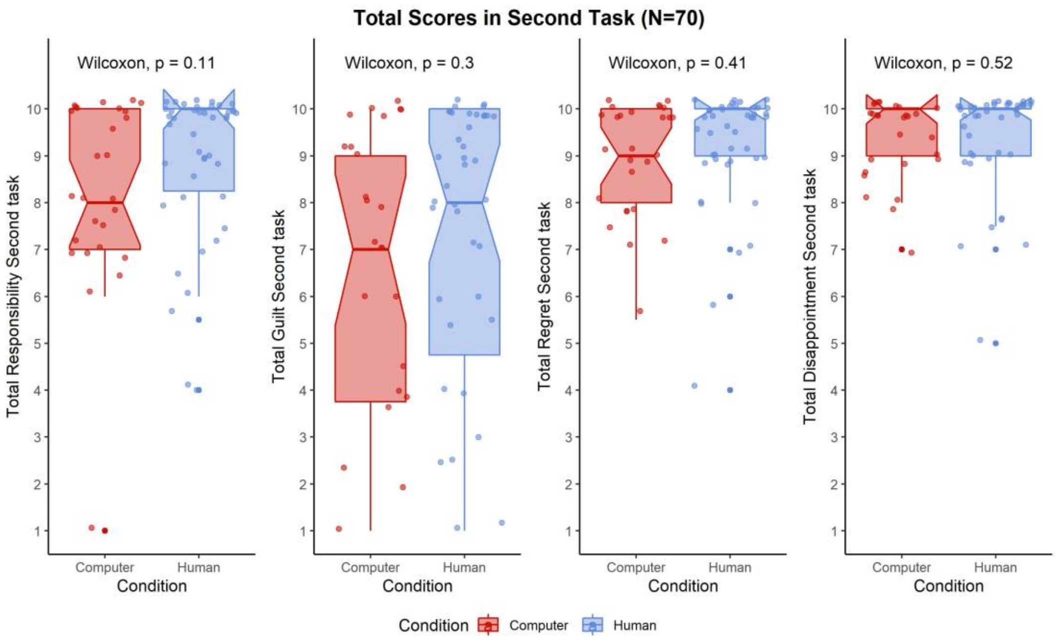

Figure 17 Total scores in the second task in human and computer condition by the dimensions of responsibility, guilt, regret, disappointment (N=70)

Table 13 Summary descriptive statistics on dependent variables in the second task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY TASK2	8.26	8	1.99	27
	GUILT TASK2	6.24	7	3.28	
	REGRET TASK2	8.98	9	1.23	
	DISAPP TASK2	9.37	10	0.86	
Human	RESPONSIBILITY TASK2	8.91	10	1.65	43
	GUILT TASK2	7.01	8	3.29	
	REGRET TASK2	9.21	10	1.25	
	DISAPP TASK2	9.38	10	1.12	

A Mann-Whitney's U test was chosen to evaluate the differences in the scores of total responsibility in the second task. The medians of computer condition and the human condition were 8 and 10, respectively. It was found a non-significant effect between the two groups, U = 455, p = 0.11.

Mann-Whitney's U test showed also a non-significant difference for the dimension of total guilt in the second task. The scores of computer condition (Mdn=7) were not different compared to the human condition (Mdn=8), U= 495, p=0.30.

For total regret in the second task, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=9) and human suggestion (Mdn=10), U=516, p=0.41. Even for total disappointment scores in the first task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=10) condition and the human condition (Mdn=10), U=533, p=0.52.

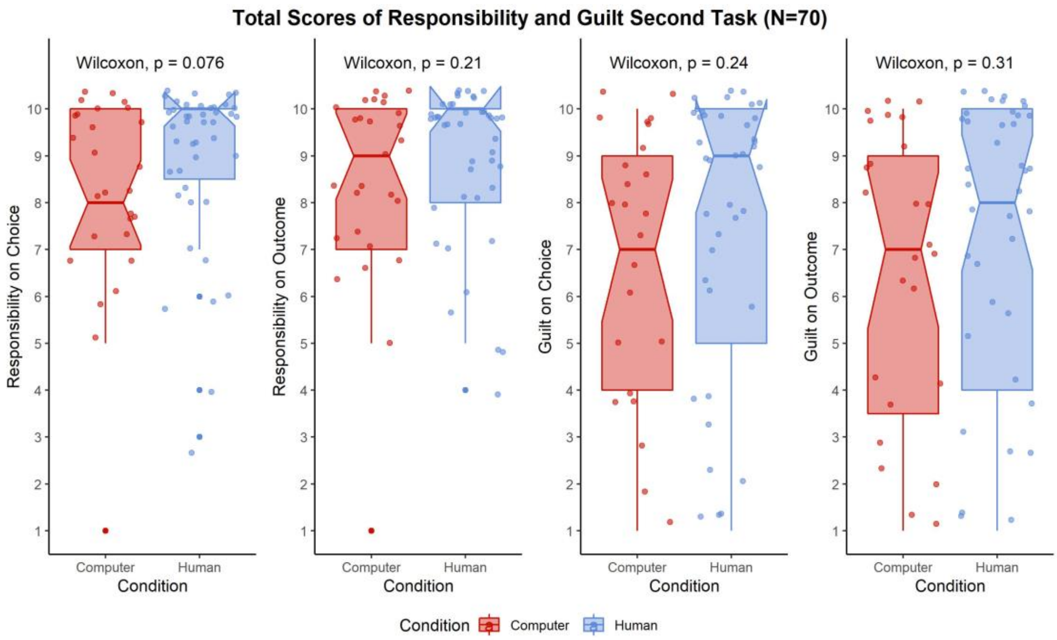

Figure 18 Scores in the first task in human and computer condition of the dimensions of regret and disappointment (N=70)

Table 14 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the first task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	REGRET ON CHOICE	9.26	10	1.02	27
	REGRET ON OUTCOME	9.22	10	0.97	
	DISAPPOINTMENT ON CHOICE	9.185	9	0.96	
	DISAPPOINTMENT ON OUTCOME	9.15	9	0.95	
Human	REGRET ON CHOICE	9.28	9	0.83	43
	REGRET ON OUTCOME	9.19	10	0.98	
	DISAPPOINTMENT ON CHOICE	9.28	10	0.83	
	DISAPPOINTMENT ON OUTCOME	9.30	10	0.89	

Mann-Whitney's U test showed a non-significant difference for the dimension of total regret on choice in the first task. The scores of computer condition (Mdn=10) were not significantly different compared to the human condition (Mdn=10), U= 602, p=0.78.

For regret on outcome in the first task, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=10), U=594.5, p=0.86.

Even for disappointment on choice in the first task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=9) condition and the human condition (Mdn=10), U=560, p=0.79. Non-significant results were found for the differences between the scores of disappointment on outcome in the first task between computer condition (Mdn=9) and the human condition (Mdn=10), as showed by Mann-Whitney's U test, U=525.5, p=0.47.

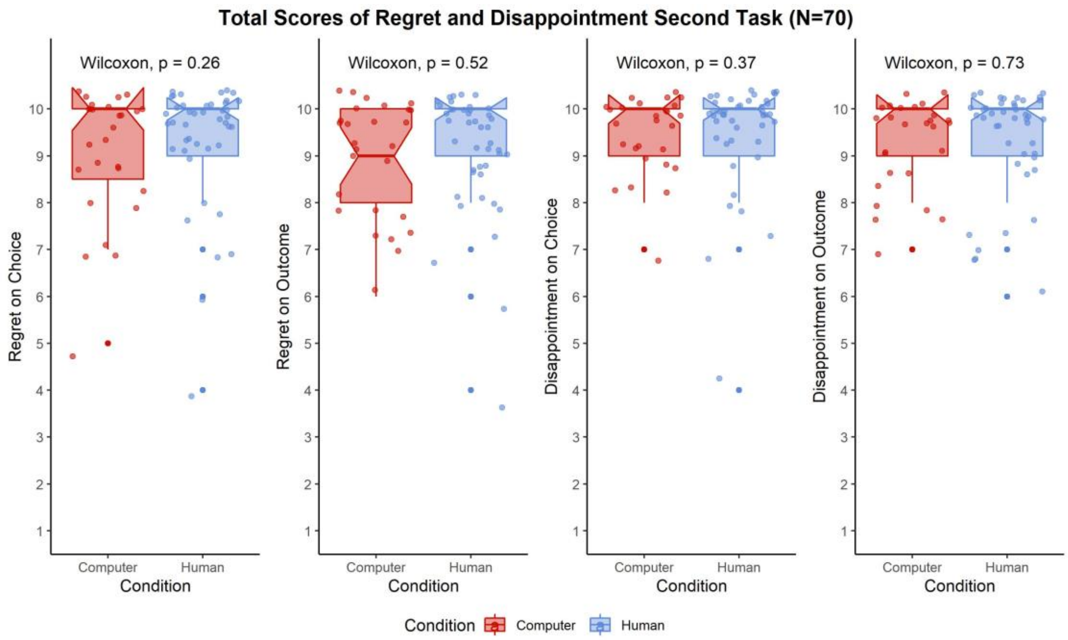

Figure 19 Scores in the second task in human and computer condition of the dimensions of regret and disappointment (N=70)

Table 15 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the second task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	REGRET ON CHOICE	9.04	10	1.31	27
	REGRET ON OUTCOME	8.93	9	1.27	
	DISAPPOINTMENT ON CHOICE	9.41	10	0.84	
	DISAPPOINTMENT ON OUTCOME	9.33	10	0.91	
Human	REGRET ON CHOICE	9.30	10	1.28	43
	REGRET ON OUTCOME	9.12	10	1.27	
	DISAPPOINTMENT ON CHOICE	9.44	10	1.18	
	DISAPPOINTMENT ON OUTCOME	9.33	10	1.12	

A non-parametric Mann-Whitney test indicated that regret on the choice on the second task was not different from computer suggestion (Mdn = 10) and human suggestion (Mdn = 10), U=498.5, p=0.26.

The median of experienced regret on outcome during the second task was not different between the experimental condition (Mdn = 9) and the control condition (Mdn = 10), a Mann-Whitney's U test showed non-significant results, U=530.5, p=0.52

As regards disappointment on choice on the first task, the Mann-Whitney test showed a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=10), U=518.5, p=0.37. Non-significant results were found for the differences between the scores of disappointment on outcome in the second task between computer condition (Mdn=10) and the human condition (Mdn=10), as showed by Mann-Whitney's U test, U=555, p=0.73.

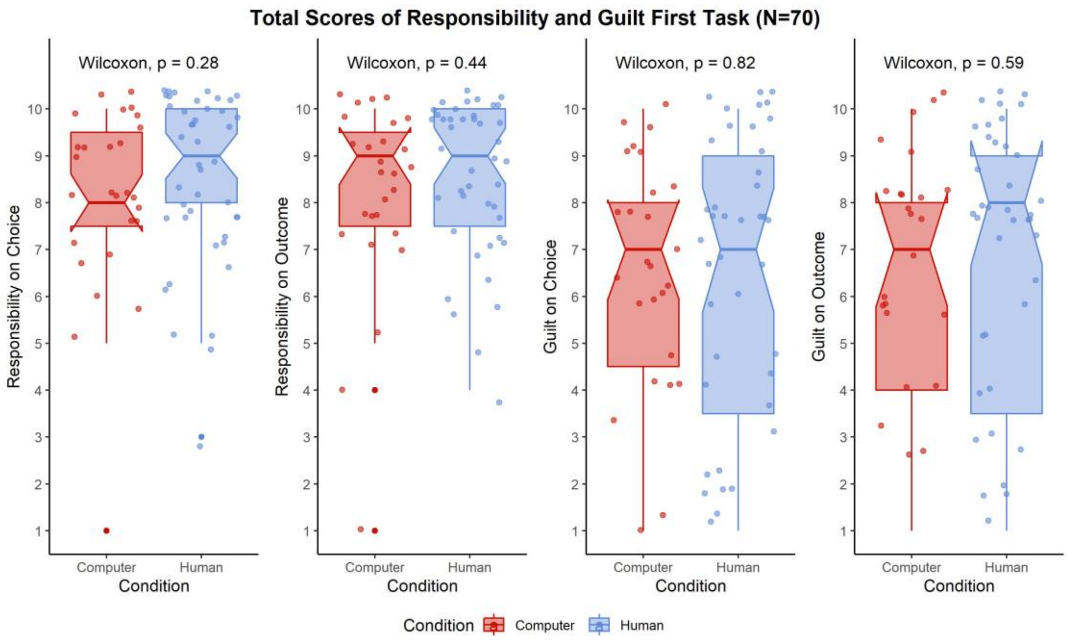

Figure 20 Scores in the first task in human and computer condition of the dimensions of responsibility and guilt (N=70)

Table 16 Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the first task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON CHOICE	8.07	8	1.98	27
	RESPONSIBILITY ON OUTCOME	8.15	9	2.09	
	GUILT ON CHOICE	6.33	7	2.72	
	GUILT ON OUTCOME	6.26	7	2.82	
Human	RESPONSIBILITY ON CHOICE	8.49	9	1.79	43
	RESPONSIBILITY ON OUTCOME	8.54	9	1.65	
	GUILT ON CHOICE	6.26	7	3.22	
	GUILT ON OUTCOME	6.44	8	3.16	

A Mann-Whitney's U test was chosen to evaluate the differences in the scores of total responsibility on choice in the first task. The medians of computer condition and the human condition were 8 and 9, respectively. It was found a non-significant effect between the two groups, U = 494.5, p = 0.28.

Mann-Whitney's U test also showed a non-significant difference for the dimension of responsibility on outcome in the first task. The scores of computer condition (Mdn=9) were not different compared to the human condition (Mdn=9), U= 517.5, p=0.44.

For guilt on choice in the first task, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=7) and human suggestion (Mdn=7), U=561, p=082.

Even for total guilt on outcome scores in the first task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=7) condition and the human condition (Mdn=8), U=536, p=0.59.

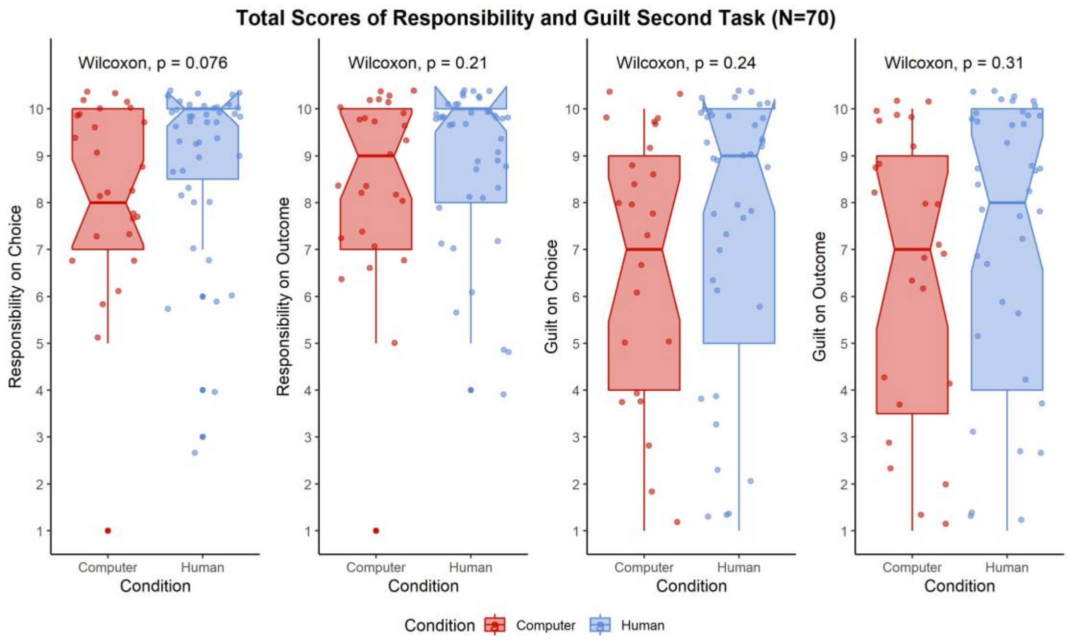

Figure 21 Scores in the second task in human and computer condition of the dimensions of responsibility and guilt (N=70)

Table 17 Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the second task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON CHOICE*	8.19	8	2.08	27
	RESPONSIBILITY ON OUTCOME	8.33	9	2.09	
	GUILT ON CHOICE	6.30	7	3.27	
	GUILT ON OUTCOME	6.19	7	3.34	
Human	RESPONSIBILITY ON CHOICE	8.91	10	1.72	43
	RESPONSIBILITY ON OUTCOME	8.91	10	1.66	
	GUILT ON CHOICE	7.12	9	3.34	
	GUILT ON OUTCOME	6.91	8	3.39	

A Mann-Whitney test indicated that responsibility on the choice on the second task was not different from computer suggestion (Mdn = 8) and human suggestion (Mdn = 10), U=442, p=0.07.

The median of responsibility on the outcome on the second task was not different between the experimental condition (Mdn = 9) and the control condition (Mdn = 10), a Mann-Whitney's U test showed non-significant results, U=484.5, p=0.21

As regards guilt on choice on the second task, the Mann-Whitney test showed a non-significant difference between computer suggestion (Mdn=7) and human suggestion (Mdn=9), U=483.5, p=0.24.

Non-significant results were found for the differences between the scores of disappointment on outcome in the second task between computer condition (Mdn=7) and the human condition (Mdn=8), as showed by Mann-Whitney's U test, U=498, p=0.31.

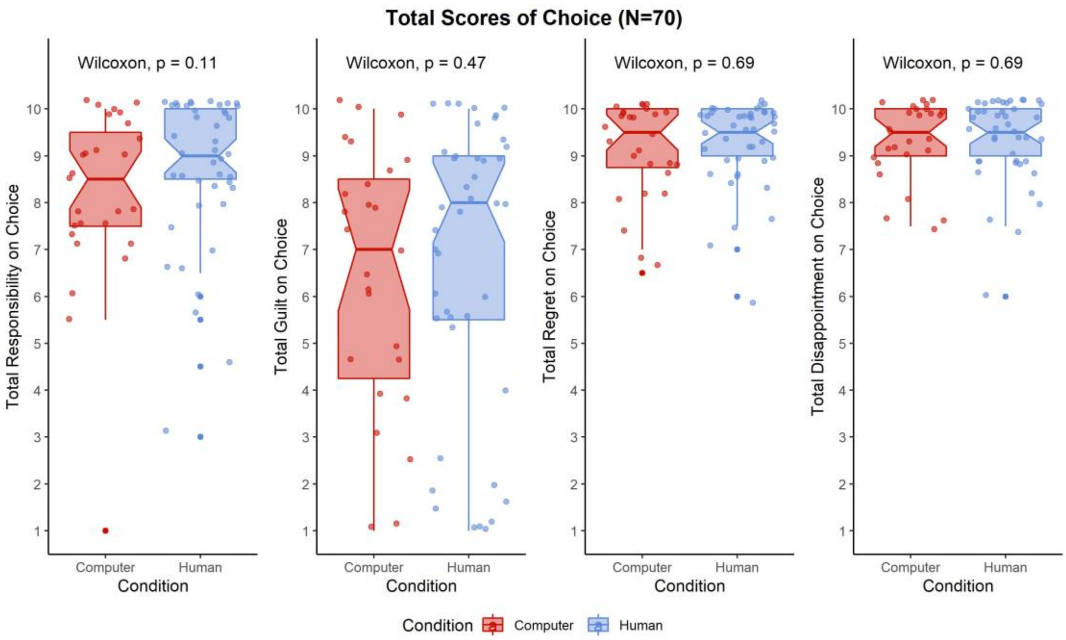

Figure 22 Scores in both tasks in human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on choice (N=70)

Table 18 Summary descriptive statistics of DVs on choice in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON CHOICE	8.13	8.5	1.92	27
	GUILT ON CHOICE	6.32	7	2.91	
	REGRET ON CHOICE	9.15	9.5	1.04	
	DISAPPOINTMENT ON CHOICE	9.3	9.5	0.85	
Human	RESPONSIBILITY ON CHOICE	8.7	9	1.65	43
	GUILT ON CHOICE	6.7	8	3.11	
	REGRET ON CHOICE	9.29	9.5	0.93	
	DISAPPOINTMENT ON CHOICE	9.36	9.5	0.88	

Mann-Whitney's U test evaluated the dimension of total responsibility of choice. The medians of computer condition and the human condition were 8.5 and 9, respectively. It was found a non-significant effect between the two groups; U = 448.5, p = 0.10.

Mann-Whitney's U test showed a non-significant difference for the dimension of total guilt on choice in both tasks. The scores of computer condition (Mdn=7) were not different compared to the human condition (Mdn=8), U= 520, p=047.

For total regret on choice, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=9.5) and human suggestion (Mdn=9.5), U=548.5, p=0.69. Even for total disappointment scores in the first task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=9.5) condition and the human condition (Mdn=9.5), U=549, p=0.69.

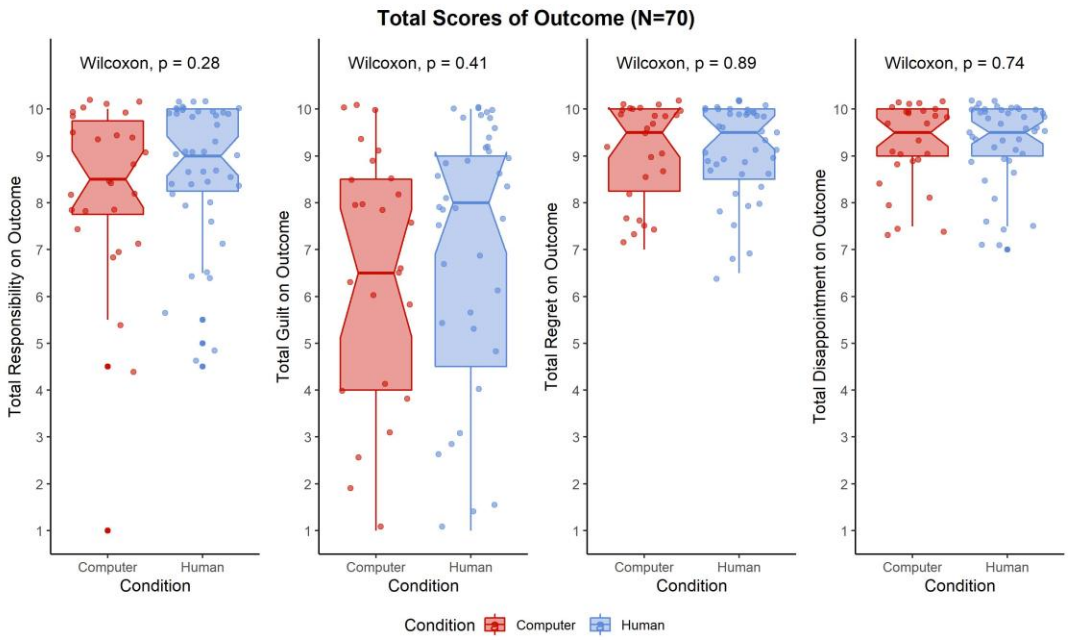

Figure 23 Scores in both tasks in human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on outcome (N=70)

Table 19 Summary descriptive statistics of DVs on outcome in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON OUTCOME	8.24	8.5	2.04	27
	GUILT ON OUTCOME	6.22	6.5	2.99	
	REGRET ON OUTCOME	9.01	9.5	1.06	
	DISAPPOINTMENT ON OUTCOME	9.24	9.5	0.88	
Human	RESPONSIBILITY ON OUTCOME	8.72	9	1.49	43
	GUILT ON OUTCOME	6.67	8	3.20	
	REGRET ON OUTCOME	9.15	9.5	0.97	
	DISAPPOINTMENT ON OUTCOME	9.31	9.5	0.89	

A Mann-Whitney's U test was chosen to evaluate the differences in the scores of total responsibility on the outcome. The medians of computer condition and the human condition were 8.5 and 9, in that order. It was found a non-significant effect between the two groups, U = 492, p = 0.28.

Mann-Whitney’s U test showed also a non-significant difference for the dimension of guilt on the outcome. The scores of computer condition (Mdn=6.5) were not different compared to the human condition (Mdn=8), U= 512, p=0.41.

Non-significant results were found for the differences between the scores of regret on the outcome between computer condition (Mdn=9.5) and the human condition (Mdn=9.5), as showed by Mann-Whitney’s U test, U=569, p=0.89. As regards guilt on the outcome, the Mann-Whitney test presented a non-significant difference between computer suggestion (Mdn=9.5) and human suggestion (Mdn=9.5), U=553.5, p=0.74.

5.2. Removed outliers

In this section, the data were handled excluding the outliers according to the subjects out the limits of the whiskers in the boxplots. Each dimension in each task was handled independently, and other non-parametric tests were carried out.

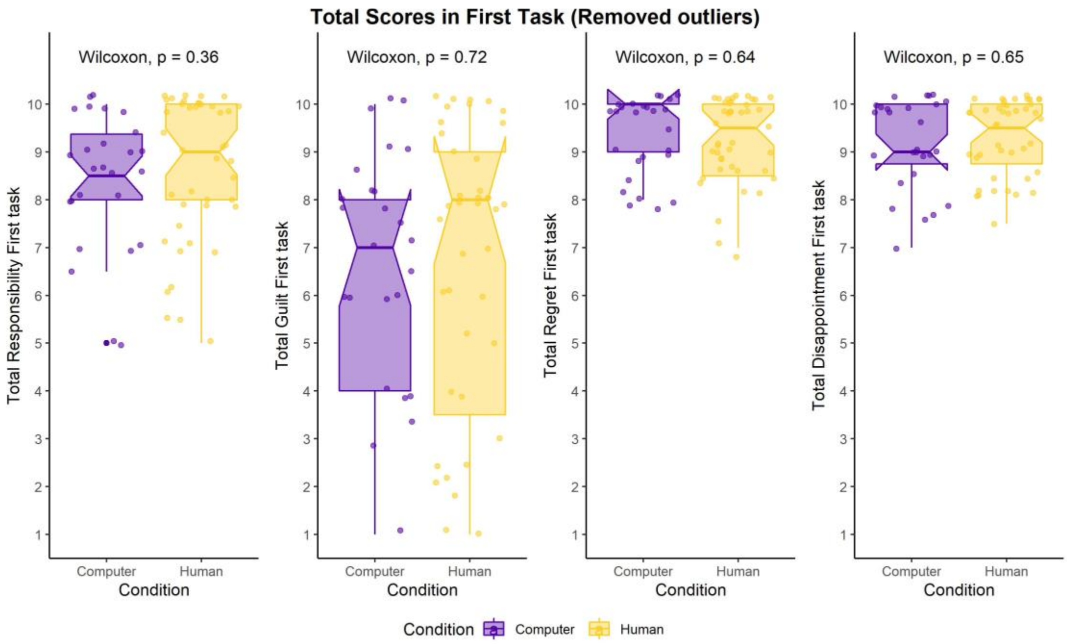

Figure 24 Scores in the first task human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on choice (N=68)

Table 20 Summary descriptive statistics on dependent variables in the first task in human and computer conditions

		M	Median	SD	# Subj
Computer	RESPONSIBILITY TASK1	8.39	8.5	1.43	26
Human	RESPONSIBILITY TASK1	8.63	9	1.53	42
Computer	GUILT TASK1	6.30	7	2.75	27
Human	GUILT TASK1	6.35	8	3.17	43
Computer	REGRET TASK1	9.35	10	0.81	26
Human	REGRET TASK1	9.23	9.5	0.88	43
Computer	DISAPP TASK1	9.17	9	0.93	27
Human	DISAPP TASK1	9.29	9.5	0.82	43

A Mann-Whitney test indicated that the scores of total perceived responsibility were not different in the human condition (Mdn= 9) than in Computer condition (Mdn=8.5), U= 475, p=0.369

As regards total guilt in the first task, the Mann-Whitney's U test showed a non-significant difference between computer condition (Mdn=8.5) and the human condition (Mdn=9), U=551, p=0.73.

Similarly, total regret in the first task showed a non-significant difference between the suggestion in the experimental condition (Mdn=10) and control condition (Mdn=9.5), Mann-Whitney's U test showed U=595, p=0.67. Even total disappointment in the first task showed a similarity in the medians in the computer condition (Mdn=9) and the human condition (Mdn=9.5); Mann-Whitney's U test was U=544.5, p=0.65.

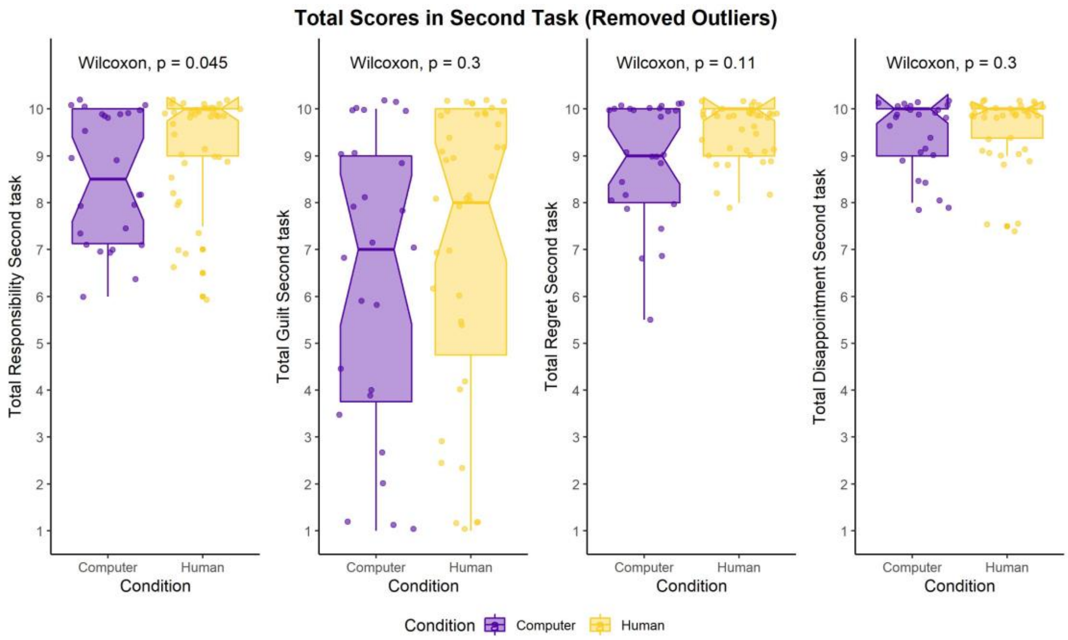

Figure 25 Scores in the second task human and computer condition of the dimensions of responsibility, guilt, regret and disappointment on choice

Table 21 Summary descriptive statistics on dependent variables in the second task in human and computer conditions

		M	Median	SD	# Subj
Computer	RESPONSIBILITY TASK2*	8.54	8.5	1.39	26
Human	RESPONSIBILITY TASK2*	9.24	10	1.13	40
Computer	GUILT TASK2	6.24	7	3.29	27
Human	GUILT TASK2	7.012	8	3.29	43
Computer	REGRET TASK2	8.98	9	1.23	27
Human	REGRET TASK2	9.54	10	1.24	39
Computer	DISAPP TASK2	9.46	10	0.73	26
Human	DISAPP TASK2	9.61	10	0.72	40

A Mann-Whitney test indicated that total responsibility in the second task was lower for computer suggestion (Mdn = 8.5) compared to human suggestion (Mdn = 9), U=377, p=0.04.

The median of guilt in the second task was not different between the computer condition (Mdn = 7) and the control condition (Mdn = 8), a Mann-Whitney's U test showed non-significant results, U=495, p=0.30.

As regards regret in the second task, the Mann-Whitney test showed a non-significant difference between computer suggestion (Mdn=9) and human suggestion (Mdn=10), U=413, p=0.11.

Even for disappointment in the second task, the Mann-Whitney test showed a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=10), U=453, p=0.30.

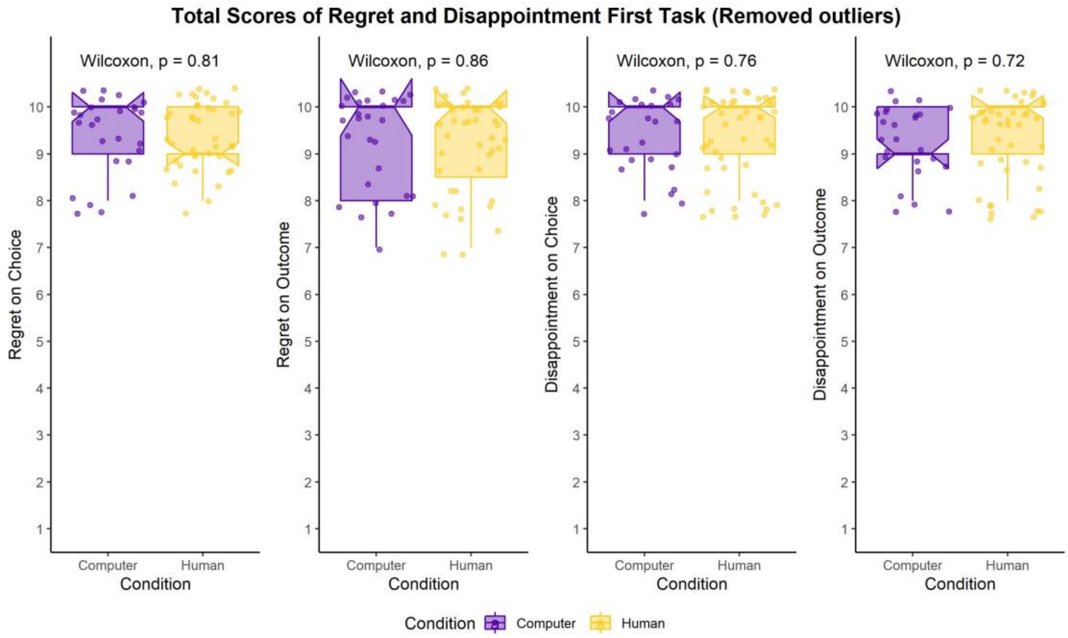

Figure 26 Scores in the first task human and computer condition of the dimensions of regret and disappointment

Table 22 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the first task in human and computer conditions

		M	Median	SD	# Subj
Computer	REGRET ON CHOICE	9.39	10	0.80	26
Human	REGRET ON CHOICE	9.39	9	0.67	41
Computer	REGRET ON OUTCOME	9.22	10	0.97	27
Human	REGRET ON OUTCOME	9.19	10	0.98	43
Computer	DISAPPOINTMENT ON CHOICE	9.36	10	0.76	25
Human	DISAPPOINTMENT ON CHOICE	9.28	10	0.83	43
Computer	DISAPPOINTMENT ON OUTCOME	9.32	9	0.74	25
Human	DISAPPOINTMENT ON OUTCOME	9.36	10	0.82	42

Non-significant results were found for the differences between the scores of regret on choice in the first task between computer condition (Mdn=10) and the human condition (Mdn=9), as showed by Mann-Whitney's U test, U=550, p=0.81.

As regards regret on outcome in the first task, the Mann-Whitney test presented a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=10), U=594.5, p=0.86.

For disappointment on choice in the first task, non-parametric Mann-Whitney's U test showed a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=10), U=560, p=0.76.

Even for the disappointment of choice in the first task, the two conditions did not differ comparing the medians of computer condition (Mdn=9) and the human condition (Mdn=10), Mann-Whitney's test U=499.5, p=0.72.

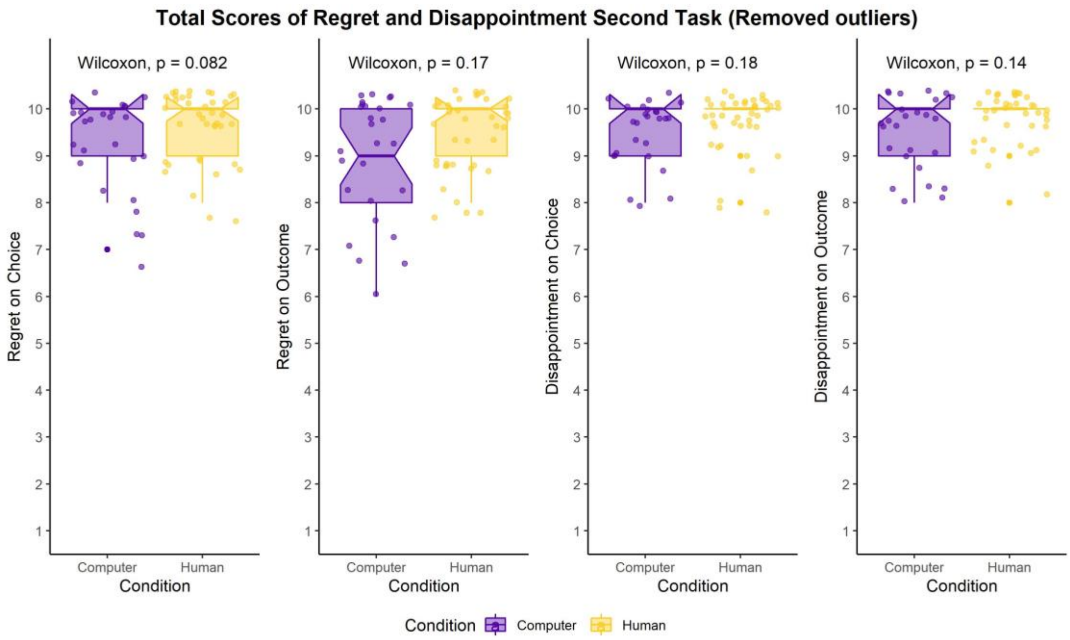

Figure 27 Scores in the first task human and computer condition of the dimensions of regret and disappointment

Table 23 Summary descriptive statistics of regret and disappointment (on choice and outcome) in the second task in human and computer conditions (N=67)

		M	Median	SD	# Subj
Computer	REGRET ON CHOICE	9.19	10	1.06	26
Human	REGRET ON CHOICE	9.64	10	0.63	39
Computer	REGRET ON OUTCOME	8.93	9	1.27	27
Human	REGRET ON OUTCOME	9.44	10	0.72	39
Computer	DISAPPOINTMENT ON CHOICE	9.5	10	0.71	26
Human	DISAPPOINTMENT ON CHOICE	9.7	10	0.61	40
Computer	DISAPPOINTMENT ON OUTCOME	9.42	10	0.81	26
Human	DISAPPOINTMENT ON OUTCOME	9.73	10	0.51	37

The dimension of regret on choice in the second task was evaluated by Mann-Whitney's U test. The medians of computer condition and the human condition were 10 and 10, respectively. It was found a non-significant effect between the two groups; U = 396.5, p = 0.08.

Mann-Whitney's U test showed a non-significant difference for the dimension of regret on outcome in the second task. The scores of computer condition (Mdn=9) were not different compared to the human condition (Mdn=10), U= 429, p=0.17.

For disappointment on choice in the second task, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=10) and human suggestion (Mdn=10), U=438.5, p=0.18.

Even for disappointment on outcome in the second task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=10) condition and the human condition (Mdn=10), U=395.5, p=0.14.

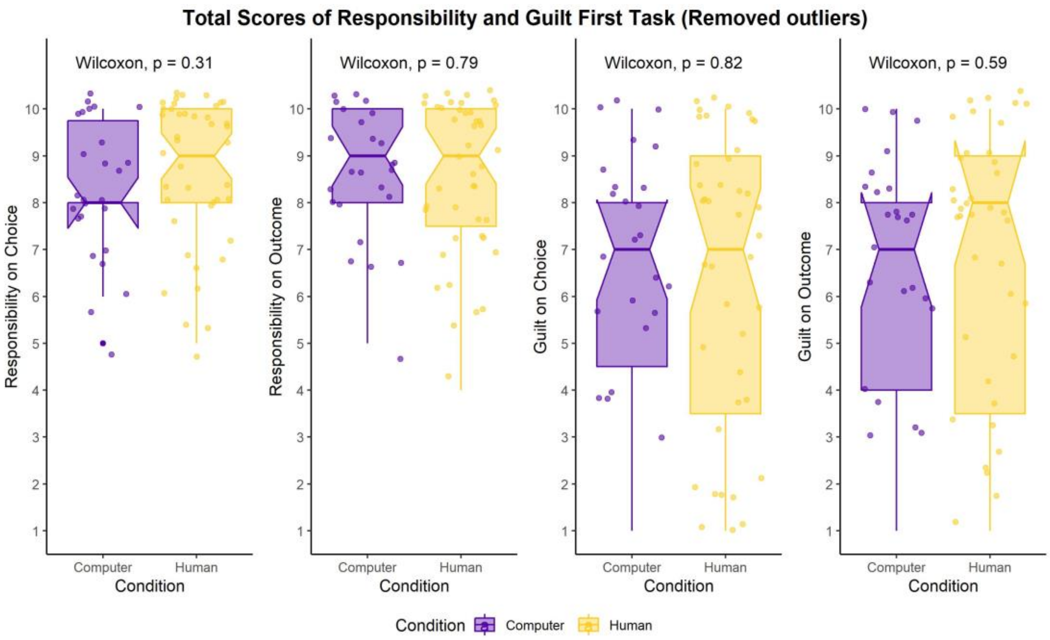

Figure 28 Scores in the first task human and computer condition of the dimensions of responsibility and guilt

Table 24 Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the first task in human and computer conditions

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON CHOICE	8.35	8	1.41	26
Human	RESPONSIBILITY ON CHOICE	8.62	9	1.59	42
Computer	RESPONSIBILITY ON OUTCOME	8.6	9	1.29	25
Human	RESPONSIBILITY ON OUTCOME	8.54	9	1.65	43
Computer	GUILT ON CHOICE	6.33	7	2.71	27
Human	GUILT ON CHOICE	6.26	7	3.22	43
Computer	GUILT ON OUTCOME	6.25	7	2.82	27
Human	GUILT ON OUTCOME	6.44	8	3.16	43

A Mann-Whitney's U test was chosen to evaluate the differences in the scores of responsibility on choice in the first task. The medians of computer condition and the human condition were 8 and 9, respectively. It was found a non-significant effect between the two groups, U = 468.5, p = 0.31.

Mann-Whitney's U test also showed a non-significant difference for the dimension of responsibility on outcome in the first task. The scores of computer condition (Mdn=9) were not different compared to the human condition (Mdn=9), U= 517 p=0.79.

For guilt on choice in the first task, Mann-Whitney's U test indicated a non-significant difference between computer suggestion (Mdn=7) and human suggestion (Mdn=7), U=561, p=082.

Even for total guilt on outcome scores in the first task, Mann-Whitney's U test showed a non-significant difference between the computer (Mdn=7) condition and the human condition (Mdn=8), U=536, p=0.59.

Figure 29 Scores in the second task human and computer condition of the dimensions of responsibility and guilt

Table 25: Summary descriptive statistics of responsibility and guilt (on choice and outcome) in the second task in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON CHOICE*	8.46	8.5	1.53	26
Human	RESPONSIBILITY ON CHOICE*	9.42	10	0.89	48
Computer	RESPONSIBILITY ON OUTCOME	8.62	9	1.52	26
Human	RESPONSIBILITY ON OUTCOME	9.02	10	1.49	42
Computer	GUILT ON CHOICE	6.30	7	3.27	27
Human	GUILT ON CHOICE	7.12	9	3.34	43
Computer	GUILT ON OUTCOME	6.19	7	3.34	27
Human	GUILT ON OUTCOME	6.91	8	3.39	43

A Mann-Whitney's U test was run to assess the differences in the scores of responsibility on choice in the second task. The medians of computer condition and the human condition were 8.5 and 10, in that order. It was found a significant effect between the two groups, U = 318, p = 0.009.

Mann-Whitney's U test also showed a significant difference in the dimension of responsibility on outcome in the second task. The scores of computer condition (Mdn=9) were not different compared to the human condition (Mdn=10), U= 485.5, p=0.23.

Non-significant results were found for the differences between the scores of guilt on choice in the second task between computer condition (Mdn=7) and the human condition (Mdn=9), as showed by Mann-Whitney's U test, U=483.5, p=0.24.

As regards guilt on the outcome, the Mann-Whitney test presented a non-significant difference between computer suggestion (Mdn=7) and human suggestion (Mdn=8), U=498, p=0.31.

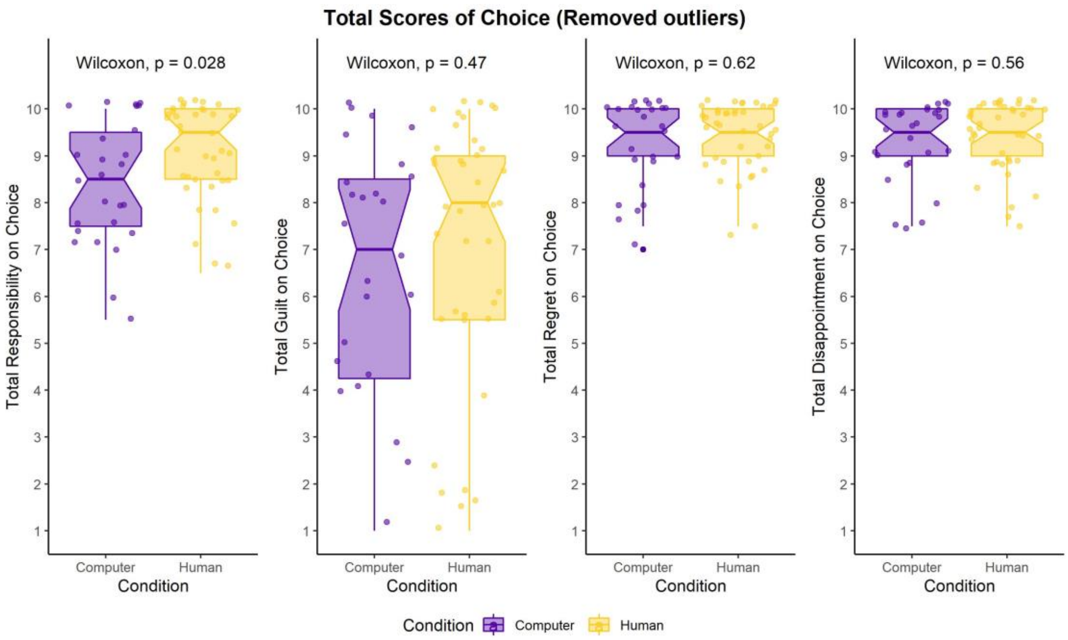

Figure 30 Total scores on choice in human and computer conditions of the dimensions of responsibility, guilt, regret, and disappointment

Table 26 Summary descriptive statistics of DVs on choice in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON CHOICE*	8.4	8.5	1.31	26
Human	RESPONSIBILITY ON CHOICE*	9.1	9.5	1.03	39
Computer	GUILT ON CHOICE	6.32	7	2.91	27
Human	GUILT ON CHOICE	6.69	8	3.11	43
Computer	REGRET ON CHOICE	9.25	9.5	0.91	26
Human	REGRET ON CHOICE	9.43	9.5	0.69	41
Computer	DISAPPOINTMENT ON CHOICE	9.3	9.5	0.84	27
Human	DISAPPOINTMENT ON CHOICE	9.44	9.5	0.72	42

A Mann-Whitney test pointed out that total responsibility was significantly different from computer suggestion (Mdn = 8.5) and human suggestion (Mdn = 9.5), U=346.5, p=0.03.

The median of guilt on choice was not different between the experimental condition (Mdn = 7) and the control condition (Mdn = 8), a Mann-Whitney's U test showed non-significant results, U=520, p=0.47.

As regards regret on choice, the Mann-Whitney test showed a non-significant difference between computer suggestion (Mdn=9.5) and human suggestion (Mdn=9.5), U=483.5, p=0.62.

Non-significant results were also found for the differences between the scores of disappointment on choice between computer condition (Mdn=9.5) and the human condition (Mdn=9.5), as showed by Mann-Whitney's U test, U=522, p=0.56.

Figure 31 Total scores on outcome in human and computer conditions of the dimensions of responsibility, guilt, regret and disappointment

Table 27 Summary descriptive statistics of DVs on outcome in human and computer conditions (N=70)

		M	Median	SD	# Subj
Computer	RESPONSIBILITY ON OUTCOME	8.68	9	1.25	25
Human	RESPONSIBILITY ON OUTCOME	9	9	1.1	40
Computer	GUILT ON OUTCOME	6.22	6.5	2.99	27
Human	GUILT ON OUTCOME	6.67	8	3.20	43
Computer	REGRET ON OUTCOME	9.07	9.5	1.06	27
Human	REGRET ON OUTCOME	9.15	9.5	0.97	43
Computer	DISAPPOINTMENT ON OUTCOME	9.24	9.5	0.88	27
Human	DISAPPOINTMENT ON OUTCOME	9.43	9.5	0.75	41

The dimension of regret on outcome was calculated by Mann-Whitney's U test. The medians of computer condition and the human condition were 9 and 9, respectively. It was found a non-significant effect between the two groups; U = 417, p = 0.25.

Mann-Whitney's U test showed a non-significant difference for the dimension of guilt on outcome. The scores of computer condition (Mdn=6.5) were not different compared to the human condition (Mdn=8), U= 512, p=0.41.

As regards regret on the outcome, the Mann-Whitney test presented a non-significant difference between computer suggestion (Mdn=9.5) and human suggestion (Mdn=9.5), U=569, p=0.89. Even for disappointment on outcome, the two conditions did not differ comparing the medians of computer condition (Mdn=9.5) and the human condition (Mdn=9.5), Mann-Whitney's test U=499.5, p=0.48.

5.3. Summary of results

In the first section, the analyses showed a lack of significance in the dependent variables of disappointment, regret, and guilt. While an exciting trend was found for the perceived responsibility on choice in the second task ($p=0.08$).

In the second section, the outliers were removed from the analyses according to the criteria of excluding participants who showed scores outside the whiskers of the boxplots (upper and lower quartiles).

These analyses showed how responsibility has higher scores after following the human suggestion. In particular total responsibility, or the sum of responsibility on choice and on the outcome, in the second task showed that participants felt more self-blame after made a mistake following twice a human advisor ($p=0.04$); the responsibility in the choice confirmed this result compared to the responsibility on the outcome. Participants felt more responsible for the wrong choice ($p=0.008$) when they trust again the same source of the suggestion than they felt responsible for the outcome ($p=0.21$). This effect showed a trend in the regret on choice in the second task ($p=0.08$), higher when people followed twice the bad advice of another human compared to following twice wrong advice by an algorithm-based system. This analysis, even if it is not statistically significant, showed a trend that needs more attention.

Eventually, the results of this study demonstrated that participants felt more responsibility on the wrong choice in both tasks ($p=0.028$) when the advisor is human than the same situation when a computer-based system gave the advice.

The results did not meet the initial expectations since any difference in regret, disappointment, and guilt between the human and computer-based task have not been found. The initial intuition was at least partially confirmed because the effect on responsibility was rather strong. Indeed, the effect on responsibility can be related to the antecedents of regret (Zeelenberg, van Dijk, et al., 2000).

Even if further studies need to be carried out, from a Human-Computer Interaction (HCI) point of view, the decrease of the sense of responsibility when the suggestion comes from the computer may induce risky choices and it should be counterbalanced by appropriate design solutions.

6. Discussion and limitations

This study presented preliminary research aimed at understanding the possible differences in the acceptance of an algorithmic suggestion with respect to a human being suggestion. Negative and counterfactual emotions were analyzed, presenting scenarios in which participants experienced two wrong outcomes due to bad suggestions.

The evidence from this study suggests that perceived responsibility on choice is lower when participants received a wrong suggestion by a computer rather than by a human being.

This evidence may suggest that decision making with computer advice may eventually induce risky choices due to the lower sense of responsibility.

This result needs to be confirmed by other studies, in particular, because of the lack of significance for the other negative and counterfactual emotions that are well-known related to a sense of responsibility. Primarily, it will be essential to vary the scenarios, to understand the context-specific effect of counterfactual emotions, and to control for possible confounding variables (such as personality traits or attitude toward technology).

Still, these results are appealing for a better understanding of the relation between users and intelligent machines in a decision-making process and for a better design of this type of systems. In particular, the contribution of this study to the design of intelligent systems is to make the designers conscious about how the context can influence counterfactual thoughts, and every decisional task can affect the emotions of regret and disappointment in different ways.

The study had some limitations. First, one may note that for all emotions and in both scenarios, the values tend to be quite high. These scores might be due to either a bias from recruiting the participants in a crowdsourcing service (as it might be apparent from the high number of participants that the author had to exclude from the analysis) or from the scenarios that may have appeared confused or unnatural. Second, the variance is quite large for all the emotions, and in both scenarios. Again, this might be due to the reasons above, or it might depend on some conditions that were not measured (for example, different perceptions of the true intentions of the clerk while the computer might have looked more neutral concerning hidden intentions). Alternatively, personality traits (for example, the Locus of Control (Ajzen, 2002) of participants) might be a confounding variable in this case. Another possibility would be that participants might have transferred the responsibility of the wrong decision outcome on the source of the suggestion, both perceived as experts (see for example the discussion in (Pieters & Zeelenberg, 2007)). Moreover, one of the limitations of the experiment based on scenarios is that they can be perceived as artificial by the participants, and the specific situation used can affect the results.

The next study will put the attention to the manipulation of algorithm aversion from a classic scenario from the literature (Logg et al., 2019; Moore & Klein, 2008; Sniezek & Buckley, 1995). It will focus on the exploration of different positive and negative emotions in the Judge-Advisor System paradigm between human and algorithm-based advisors.

Chapter 4: Counterfactual emotions and "algorithm appreciation."

1. Motivation and aims

The previous study showed the role of responsibility in the wrong outcome between a human and a computer suggestion in a purchase scenario. In particular, experiment 2 demonstrated that people feel more responsibility when they follow a suggestion from a human compared to a suggestion given by an intelligent system.

In the next study, to prevent the issues related to the scenario-based experiment and to maintain the multidisciplinary approach, the design was based on advice taking literature, in particular to algorithm aversion studies.

This study is adapted to the judge-advisor system (JAS) design experiment (Bonaccio & Dalal, 2006; Gino, 2008; Sniezek & Van Swol, 2001). The goal of this experiment was to replicate the results from Logg et al. (2019) and relate them to the experience of counterfactual emotions.

The fundamental importance of Logg's findings is that people appreciate more advice from an algorithm than from a human for the first time (2016; 2019), while there are no links between how bad feedback can influence the perceived responsibility in the outcome and the experienced subsequent counterfactuals.

The main idea of this study was to link the higher reliance on algorithms in the first instance with a greater surprise for seeing an algorithm errs, because of the expected perfection of the machine. According to the author's vision, this expectation can be linked to the experience of regret if the algorithm gives a wrong suggestion.

This study aims to replicate the previous findings in Logg (2016) that showed how people rely more on algorithmic suggestions and compare the previous results with counterfactual thinking and emotional experience of the participants. In this study, an error was added to the participants to give the salience to counterfactual thoughts. This design and the set up were inspired by (Logg et al., 2019), and as a variation, the method added wrong feedback after participants accomplished the task. Moreover, it adds three scales about the perception of the source of the advice, the experienced emotions, and the behavioral intention to use the systems again.

2. Hypotheses

This experiment has its roots in the Logg's findings of the higher reliance people expressed on algorithms during a visual estimation task (Logg et al., 2019).

This study aimed to investigate whether a higher reliance on the suggestion can affect significantly the experienced negative effects of counterfactuals emotions.

This experiment tests how much people rely on the advice provided by an algorithm or on the advice provided by other participants for an estimate with an objective standard.

In particular, the main purpose is to link the algorithm appreciation literature with algorithm aversion through the understanding of the underlying process of reliance. That is to say; participants follow algorithm suggestion because they think that the algorithm is perfect until they see it errs. In the author's idea, this can provoke a violation of trust and a higher feeling of self-blame and regret compared to following the suggestion provided by a human, even if it also leads to a wrong outcome.

This study has two research questions in its basis. First, do people with higher reliance in the algorithm suggestion, after a wrong outcome, show more negative emotions? Second, do people with higher reliance on the algorithm suggestion, after a wrong outcome, show less intention to use the same source of advice again?

Five hypotheses guided this work:

H1: Participants who rely more on the algorithmic suggestion feel more negative emotions about the choice.

H2: Participants who rely more on the algorithmic suggestion feel more negative emotion about the source of advice.

H3: Participants who rely more on the algorithmic suggestion show a lower intention to use the suggestion from the same source in the future again.

H4: Participants who trust more on the suggestion perceive a general lower ease of use of the suggestion nevertheless the source

H5: Participants who rely more on the suggestion perceive general lower usefulness of the suggestion nevertheless the source

3. Equipment

This experiment was built with Qualtrics XM platform. Qualtrics allows to dynamically present a suggestion with the same error reported to the participants' responses, according to the purposes of this experimental design. Besides that, Qualtrics software is suitable to be used on the Amazon Mechanical Turk platform, the crowdsourcing platform to collect participants.

4. Method

4.1. Experimental Design

The experiment had a 2-cell (advisor: participant vs. algorithm) between-subject design in which the source of the advice participants received was manipulated between an estimate given by an algorithm or given by a participant from a previous study.

The task was to estimate the weight of a person in a photograph, without any other information. Participants estimated the weight of the person twice and received advice before making the second choice. The primary dependent variable was how much the participants relied on the suggestion or, as it called in advice literature, Weight of Advice[4].

The instructions and the phrasing of the task are essentially the same as used by Logg (2019) in experiment1A. This paradigm was adopted to replicate the results achieved by the authors and investigate additional emotional implications. Specifically, three different scales were added. An emotional scale about the choice after receiving a wrong outcome(MacGeorge et al., 2013; Prahl & Van Swol, 2017), a scale on the perception of the source of the suggestion, and a scale on ease of use, usefulness, behavioral intention (Spagnolli, Guardigli, Orso, Varotto, & Gamberini, 2014).

4.2. Sample

For this experiment, 702 participants were collected through the Amazon Mechanical Turk platform. Every participant was from the US, and she/he was paid 0.25$. The aim was to collect 200 participants (100 in the human-suggestion condition and 100 in the computer-suggestion condition). Despite these arrangements, the final sample included 163 participants. The inclusion requirements were to be at least 18 years old, be fluent in English, understand weight in pounds and, read the instructions carefully. Nevertheless, 388 participants were removed by checking double (or more) occurrences in I.P. address, 21 participants did not finish the survey, 116 failed the attention check, 11 participants were excluded because of missing data or incomplete completion, 2 participants were removed because they failed the instructions of the first estimate and, 1 participant was excluded because she/he failed the instructions of second estimate (she/he gave an estimate out of the expected weight written in the instructions).

The final sample was composed of 163 participants (74 women M=42.68 SD=13.00; 89 men, M=34.72, SD=10.43) (83 computer suggestion, 80 human suggestion). For this and all the experiments

---

4 $SHIFT = \frac{FinalEstimate - InitialEstimate}{Suggestion - InitialEstimate}$. SHIFT= 0, totally discounting the advice; 0>SHIFT<1: rely on the advice

about estimates and forecasts, participants' precision was incentivized. In this experiment, participants were informed about the opportunity to enter a lottery of 10$ if the estimate was closer to the real weight.

To improve commitment and precision, participants earned a bonus according to their precision, according to the following bullets:

- $0.80 - perfectly forecast the actual weight
- $0.65 - within 2 pounds of the actual weight
- $0.50 - within 5 pounds of the actual weight
- $0.35 - within 7 pounds of the actual weight
- $0.20 - within 9 pounds of the actual weight
- $0.05 - within 12 pounds of the actual weight
- $0.00 - out 12 pounds of the actual weight

4.3. Experimental procedure

The procedure was adapted from Logg, Moore, Klein, and Gino (Logg et al., 2019). Participants viewed a photo of a person, made a first estimate about the weight, then received a suggestion (157 or 165 pounds) and made a final estimate before receiving the feedback about their (wrong) choice.

Before their second estimate, participants received different advice, according to their first estimate (157 pounds, if the first estimate was higher than actual weight; 165 pounds, if the first estimate was lower than actual weight). The suggestion was described as an estimate from either another participant or an algorithm. The advice had the same error (±4 pounds). The suggestion was actually quite accurate and only four pounds off from the person's actual weight (161 pounds).

This dynamic suggestion represents one of the main differences compared to Logg's design, in which the participants received the same static suggestion. This feedback was made to present a suggestion on a reasonable weight without going beyond plausible range.

Another important difference is the presentation of feedback. In Logg's design there was no feedback presented after the second judgment. In the present design, the wrong outcome has been presented.

4.3.1. Suggestion Manipulation

The presented suggestion followed exactly Logg's design. As stated before, the main differences between the present study and the study of Logg pertain to the dynamic suggestion presented to participants and the dynamic feedback about the error, in which the suggestion performed better compared to the participant. In the lines below, the wording of the two suggestions will be presented.

The task for the *human* condition:

"Here is some more information that may help you make your final estimate.

The average estimate of participants from a past study was: ### pounds."

The task for the *computer* condition:

"Here is some more information that may help you make your final estimate.

"An algorithm ran calculations based on estimates of participants from a past study.

The output that the algorithm computed as an estimate was: ### pounds."

4.4. Materials

To investigate the emotional experiences in the two conditions a questionnaire was build *ad hoc*. Such questionnaire is composed of three parts: the emotional scale on the suggestion, emotional scale on the source of advice, and a behavioral intention scale. The behavioral intention scale has its basis in Human-Computer Interaction literature, and here, the purpose was to assess the participants' willingness to trust the same source of the advice again despite the wrong suggestion. This scale was usually used in technology adoption, while here is used to understand how easy and useful was the advice, in addition to the user's behavioral intention to use the same source of advice again.

4.4.1. Dependent variables

- **Emotional scale on choice** (15 items): participants were asked to evaluate their emotions about their choice according to with the following features: happy, elated, sad, regret, guilt, anger, fear, disappointment, shame, self-blame, responsible, joy, relief, proud, competent (Likert 8 points: 1: strongly disagree; 8: strongly agree). This survey was adapted from previous advice literature (MacGeorge, Guntzviller, Hanasono, & Feng, 2013; Van Swol, MacGeorge, & Prahl,; Prahl & Van Swol, 2017), adding 8 emotions to have insight about a broader branch of emotions beside regret and disappointment. In particular, the purpose was to understand if other emotions affect interaction during the use of an algorithmic system.
- **Emotional scale on the source of suggestion** (15 items): it was asked to participants to evaluate their feeling about the source of the suggestion with the following features: trustworthy, predictable, dependable, reliable, clear, understandable, credible, inaccurate, unexpected, undependable, ambiguous, unintelligible, dishonest, authoritative, questionable (Likert 8 points: 1: strongly disagree; 8: strongly agree).
- **Ease of use** (2 items): the extent to which a user thinks the suggestion will not be tiring or difficult to use (Likert 6 points: 1: strongly disagree; 6: strongly agree) (Davis, Bagozzi, & Warshaw, 1989; Spagnolli et al., 2014).

- **Usefulness** (2 items): refers to the extent to which the user thinks that the use of a certain suggestion may improve her performance (Likert 6 points: 1: strongly disagree; 6: strongly agree) (Davis, 1985; Spagnolli et al., 2014).
- **Behavioral intention** (2 items): refers to the user's willingness to adopt and use a particular technology. In this case, how much a user would use again the suggestion given (Likert 6 points: 1: strongly disagree; 6: strongly agree) (Davis, 1985; Spagnolli et al., 2014).
- **Weight of Advice (WOA):** Weight of Advice (WOA) measures the degree to which participants move their estimate toward the advice from the first estimate to the second estimate (Sniezek & Buckley, 1995). This measure provides information about participant's reliance on the advice. Measures the degree to which participants move their estimates toward the advice from Time 1 to Time 2. The measure is continuous and provides more information about the participant's reliance on the advice than a binary choice measure can capture. In addition to the options of fully discounting or fully updating to the advice, participants can rely on the advice as little or as much as they would like. The measures collected how much participants relied on the advice by dividing the difference between the final and initial estimate by the difference between the advice and the initial estimate and confidence in each estimate. The higher the WOA, the greater the reliance on the information.
- **SHIFT**: is a variant of WOA, usually used in judge-advisor system (JAS) research, and it is used as a measure of trust (Bonaccio & Dalal, 2006; Önkal et al., 2009)
- **Weight of Own Estimate (WOE)**: WOE expresses a ratio and reveals the weight a participant discounted the advice (Yaniv & Kleinberger, 2000).
- **Confidence measures: a**fter each estimate participants indicated how confident they were about the estimate they gave. This measure was on a scale from 0= no chance to 100= absolutely confident.

4.4.2. Independent variables

- **Demographic** (age, gender)
- **Objective Numeracy Scale** (11 items): This measure point out the general comprehension and use of simple percentage, probability and mathematical concepts (coded as 0 to 11 according to the correctness of the answers) (Lipkus et al., 2001)

4.4.3. Attentional check

- **Reading the instruction – Education question** (1 item): this measure was added to prevent attention fail and random responses. It was asked to participants to fill education question with the information presentedin the description (Oppenheimer, Meyvis, & Davidenko, 2009).

5. Results

5.1. Descriptive Statistics

Below are presented the descriptive statistics of dependent variables of scales of the dimensions collected on the choice and the suggestion.

Table 28 Pearson's Product-Moment Correlation Table of variables of experienced emotions on outcome

	Happy	Elated	Sad	Regret	Guilt	Anger	Fear	Disappointment	Shame	Self-blame	Responsible	Joy	Relief	Proud
Happy	1													
Elated	0.74****	1												
Sad	-0.11	-0.01	1											
Regret	-0.25**	-0.18*	0.54****	1										
Guilt	0.04	0.17*	0.59****	0.44****	1									
Anger	0.01	0.07	0.59****	0.41****	0.64****	1								
Fear	0.22**	0.43****	0.44****	0.18*	0.58****	0.54****	1							
Disappointment	-0.35****	-0.33****	0.55****	0.69****	0.42****	0.39****	0.15	1						
Shame	0.03	0.19*	0.61****	0.50****	0.62****	0.59****	0.58****	0.42****	1					
Self-Blame	-0.21**	-0.14	0.62****	0.68****	0.54****	0.46****	0.27***	0.74****	0.58****	1				
Responsible	0.24**	0.14	0.1	0.11	0.05	0.03	0	0.11	0.06	0.19*	1			
Joy	0.79****	0.87****	0.01	-0.18*	0.16*	0.04	0.36****	-0.33****	0.13	-0.17*	0.15	1		
Relief	0.61****	0.64****	0.06	-0.20*	0.08	0.06	0.33****	-0.25**	0.13	-0.13	0.14	0.64****	1	
Proud	0.75****	0.75****	-0.09	-0.29***	-0.03	-0.04	0.18*	-0.41****	0	-0.29***	0.1	0.77****	0.70****	1
Competent	0.59****	0.56****	-0.13	-0.23**	-0.12	-0.05	0.1	-0.39****	-0.15	-0.30***	0.14	0.58****	0.49****	0.63****

$p < .0001$ '****'; $p < .001$ '***', $p < .01$ '**', $p < .05$ '*'

Table 29 Pearson’s Product-Moment Correlation Table of variables of experienced emotions on the source of suggestion

	Trustworthy	Predictable	Dependable	Reliable	Clear	Understandable	Credible	Inaccurate	Unexpected	Undependable	Ambiguous	Unintelligible	Dishonest	Authoritative
Trustworthy	1													
Predictable	0.53****	1												
Dependable	0.85****	0.52****	1											
Reliable	0.82****	0.52****	0.81****	1										
Clear	0.54****	0.46****	0.51****	0.51****	1									
Understandable	0.53****	0.32****	0.56****	0.53****	0.51****	1								
Credible	0.83****	0.50****	0.86****	0.88****	0.53****	0.60****	1							
Inaccurate	-0.62****	-0.26***	-0.59****	-0.64****	-0.25**	-0.37****	-0.63****	1						
Unexpected	0	-0.04	-0.01	0	-0.08	-0.01	0.01	0.16*	1					
Undependable	-0.63****	-0.23**	-0.58****	-0.58****	-0.29***	-0.33****	-0.62****	0.71****	0.28***	1				
Ambiguous	-0.08	0.01	-0.12	-0.17*	-0.08	-0.09	-0.1	0.21**	0.24**	0.31****	1			
Unintelligible	-0.25**	0.06	-0.22**	-0.21**	-0.27***	-0.19*	-0.26***	0.51****	0.30****	0.48****	0.32****	1		
Dishonest	-0.44****	-0.09	-0.43****	-0.45****	-0.30****	-0.29***	-0.51****	0.63****	0.30***	0.67****	0.30***	0.53****	1	
Authoritative	0.41****	0.46****	0.43****	0.46****	0.28***	0.26***	0.43****	-0.16*	0.20*	-0.14	0.12	0.14	-0.04	1
Questionable	-0.53****	-0.21**	-0.49****	-0.52****	-0.25**	-0.24**	-0.53****	0.70****	0.28***	0.70****	0.33****	0.55****	0.52****	-0.08

$p < .0001$ ‘****’; $p < .001$ ‘***’, $p < .01$ ‘**’, $p < .05$ ‘*’

Table 30 Pearson’s Product-Moment Correlation Table of variables of experienced emotions on the source of suggestion and participants' emotions on the wrong outcome

	Happy	Elated	Sad	Regret	Guilt	Anger	Fear	Disappointment	Shame	Self-Blame	Responsible	Joy	Relief	Proud
Trustworthy	0.24**	0.27***	0.01	-0.06	-0.07	-0.05	0.11	-0.03	0	0	0.03	0.28***	0.17*	0.18*
Predictable	0.24**	0.32****	-0.02	0.02	0.07	0.01	0.22**	-0.05	0.14	0.03	0.08	0.27***	0.28***	0.22**
Dependable	0.21**	0.22**	-0.05	-0.01	-0.07	-0.09	0.05	0.01	0	0.04	0.14	0.25**	0.11	0.16*
Reliable	0.14	0.24**	0.03	0.02	0.04	-0.05	0.20*	0.03	0.12	0.02	0.04	0.24**	0.1	0.08
Clear	0.06	0.05	-0.03	0.01	-0.14	-0.05	0.04	-0.02	-0.09	-0.03	0.12	0.02	0.07	0.01
Understandable	0.08	0.01	0.08	0.1	0.02	0.08	-0.09	0.09	0.05	0.14	0.19*	0.05	-0.05	-0.02
Credible	0.13	0.19*	-0.03	0.04	-0.07	-0.11	0.07	0	0.02	0.07	0.11	0.17*	0.09	0.1
Inaccurate	0.02	0	0.17*	0.11	0.25**	0.26***	0.16*	0.11	0.15	0.11	0.07	-0.04	0.08	0
Unexpected	-0.04	0.1	0.35****	0.23**	0.32****	0.31****	0.18*	0.24**	0.26***	0.25**	0.04	0.07	0.05	0.04
Undependable	-0.01	0.01	0.19*	0.21**	0.28***	0.32****	0.17*	0.17*	0.19*	0.16*	0.09	0	0.11	0.03
Ambiguous	0.18*	0.23**	0.16*	0.15	0.16*	0.21**	0.25**	0.12	0.19*	0.28***	0.18*	0.18*	0.21**	0.17*
Unintelligible	0.17*	0.35****	0.34****	0.16*	0.43****	0.40****	0.46****	0.1	0.42****	0.22**	0.04	0.26**	0.26***	0.18*
Dishonest	0.04	0.12	0.27***	0.19*	0.36****	0.40****	0.24**	0.17*	0.30****	0.16*	0	0.1	0.14	0.06
Authoritative	0.28***	0.39****	0.1	0.06	0.11	0.05	0.27***	-0.07	0.18*	-0.01	0.11	0.41****	0.26***	0.24**
Questionable	-0.03	0.02	0.25**	0.22**	0.31****	0.37****	0.15	0.17*	0.29***	0.21**	0.05	-0.03	0.04	0.03

$p < .0001$ ‘****’; $p < .001$ ‘***’, $p < .01$ ‘**’, $p < .05$ ‘*’

Table 31 Pearson's Product-Moment Correlation Table of Acceptance measures, confidence, reliance on suggestion and on self, objective numeracy scores and experienced emotions on outcome

	Happy	Elated	Sad	Regret	Guilt	Anger	Fear	Disappointment	Shame	Self-Blame	Responsible	Joy	Relief	Proud	Competent
BI	-0.01	-0.02	-0.1	-0.05	-0.13	-0.21**	-0.11	-0.02	-0.1	-0.02	-0.01	-0.01	-0.12	-0.03	-0.17*
PU	0.02	0.07	-0.06	-0.08	-0.13	-0.13	0.02	-0.01	-0.06	0	-0.02	0.08	-0.01	0.01	-0.07
EoU	-0.18*	-0.26***	-0.16*	-0.03	-0.23**	-0.14	-0.34****	0.1	-0.23**	-0.05	0.05	-0.23**	-0.20**	-0.19*	-0.14
Conf1	0.19*	0.23**	-0.02	-0.03	0.07	0.07	0.19*	0.05	0.05	0.06	0.05	0.20*	0.15	0.18*	0.17*
Conf2	0.18*	0.18*	-0.03	-0.01	0.06	0.03	0.08	0.13	0.02	0.12	0.07	0.17*	0.1	0.17*	0.14
WOA	-0.11	-0.05	0.08	0.12	0.15*	0.08	0.08	0.15	0.16*	0.17*	0.04	-0.05	-0.07	-0.16*	-0.15
WOE	0.05	0.14	-0.05	-0.15	-0.09	-0.1	0.11	-0.21**	-0.08	-0.21**	-0.1	0.13	0.16*	0.13	0.13
SHIFT	-0.07	-0.16*	0.03	0.14	0.08	0.1	-0.1	0.18*	0.08	0.20**	0.07	-0.15	-0.16*	-0.14	-0.14
ONS	-0.12	-0.20**	-0.06	-0.01	-0.12	-0.08	-0.23**	-0.1	-0.23**	-0.1	0.05	-0.15	-0.14	-0.11	-0.03

BI= Behavioral intention to use again the same suggestion; PU= perceived usefulness; EoU= Ease of Use; Conf1= confidence first task; Conf2= confidence second task; WOA=Weight of advice; WOE= Weight of own estimate; SHIFT=variant of WOA; ONS=Objective Numeracy Score. p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table 32 Pearson’s Product-Moment Correlation Table of Acceptance measures, confidence, reliance on suggestion and on self, objective numeracy scores and experienced emotions on the source of suggestion

	Trustworthy	Predictable	Dependable	Reliable	Clear	Understandable	Credible	Inaccurate	Unexpected	Undependable	Ambiguous	Unintelligible	Dishonest	Authoritative	Questionable
BI	0.58****	0.24**	0.57****	0.56****	0.31****	0.35****	0.60****	-0.67****	-0.02	-0.60****	-0.1	-0.42****	-0.52****	0.11	-0.56****
PU	0.70****	0.34****	0.70****	0.67****	0.40****	0.43****	0.71****	-0.70****	0	-0.62****	-0.08	-0.35****	-0.54****	0.25**	-0.52****
EoU	0.12	-0.06	0.13	0.1	0.26***	0.33****	0.19*	-0.23**	-0.24**	-0.19*	-0.26***	-0.36****	-0.31****	-0.14	-0.22**
Conf1	0.09	0.12	0.14	0.1	-0.02	0.05	0.13	0.1	0.14	0.03	0.15	0.23**	0.06	0.04	0.05
Conf2	0.05	0.12	0.11	0.06	-0.09	0	0.08	0.11	0.15	0.03	0.13	0.18*	0.05	0.02	0.03
WOA	-0.13	-0.05	-0.11	-0.08	-0.12	-0.1	-0.1	0.12	-0.01	0.1	0.01	0.11	0.20*	0.02	0.06
WOE	0.12	0.1	0.11	0.14	0.12	0.04	0.09	-0.14	0.08	-0.09	-0.02	-0.03	-0.13	0.05	-0.16*
SHIFT	-0.09	-0.09	-0.1	-0.14	-0.11	-0.04	-0.08	0.1	-0.07	0.06	0.02	0	0.14	-0.07	0.12
ONS	-0.02	-0.04	-0.08	-0.12	0.04	-0.03	-0.04	-0.08	-0.1	-0.1	-0.02	-0.26**	-0.16*	-0.11	-0.09

BI= Behavioral intention to use again the same suggestion; PU= perceived usefulness; EoU= Ease of Use; Conf1= confidence first task; Conf2= confidence second task; WOA=Weight of advice; WOE= Weight of own estimate; SHIFT=variant of WOA; ONS=Objective Numeracy Score. $p < .0001$ ‘****’; $p < .001$ ‘***’, $p < .01$ ‘**’, $p < .05$ ‘*’.

Table 33 Pearson's Product-Moment Correlation Table of Acceptance measures, confidence, reliance on suggestion and on self, objective numeracy scores

	BI	Usef	EoU	Conf1	Conf2	WOA	WOE	SHIFT
BI	1							
Usef	0.84****	1						
EoU	0.21**	0.23**	1					
Conf1	-0.06	-0.01	-0.19*	1				
Conf2	-0.1	-0.06	-0.19*	0.84****	1			
WOA	-0.12	-0.15	0.03	-0.05	0.05	1		
WOE	0.11	0.15	-0.15	0.11	0	-0.80****	1	
SHIFT	-0.09	-0.14	0.14	-0.13	-0.02	0.85****	-0.96****	1
ONS	0.03	-0.06	0.19*	-0.24**	-0.12	-0.06	0.03	-0.01

BI= Behavioral intention to use again the same suggestion; PU= perceived usefulness; EoU= Ease of Use; Conf1= confidence first task; Conf2= confidence second task; WOA=Weight of advice; WOE= Weight of own estimate; SHIFT=variant of WOA; ONS=Objective Numeracy Score. $p < .0001$ '****'; $p < .001$ '***', $p < .01$ '**', $p < .05$ '*'

5.2. Replication of Logg's results

Logg's experiment found that participants relied more on algorithmic suggestion compared to human suggestion. As found in Logg's, participants relied more on the same advice when it came from an algorithm (M= .55, SD=.30) than when they thought it came from other people (M=.42, SD=.32), F(1,161) = 6,78, p=.01. The same results have been found when controlling for gender, age and confidence after the first estimate, F(4, 158) = 2.38, p = .053. See Figure 32.

Compared to Logg's results, the current findings show no difference in terms of increased confidence at Time 2 in the computer condition (Mdn= 6, SD= 14,7) compared to the human condition (Mdn= 5, SD= 12.09), as shown in Wilcoxon Signed-ranks test Z= 3529, p=0.49.

Another difference found between the results of the current study compared to the original results of Logg's study regards numeracy; participants in the human condition did not show a correlation with the measure of reliance in the suggestion (WOA), as showed in Pearson's product-moment correlation, r=-0.17, p=0.11. A similar finding occurs in participants who received the algorithmic suggestion, r=0.01, p=0.91, although, in the original results, the authors found a correlation between numeracy and reliance on computer suggestion.

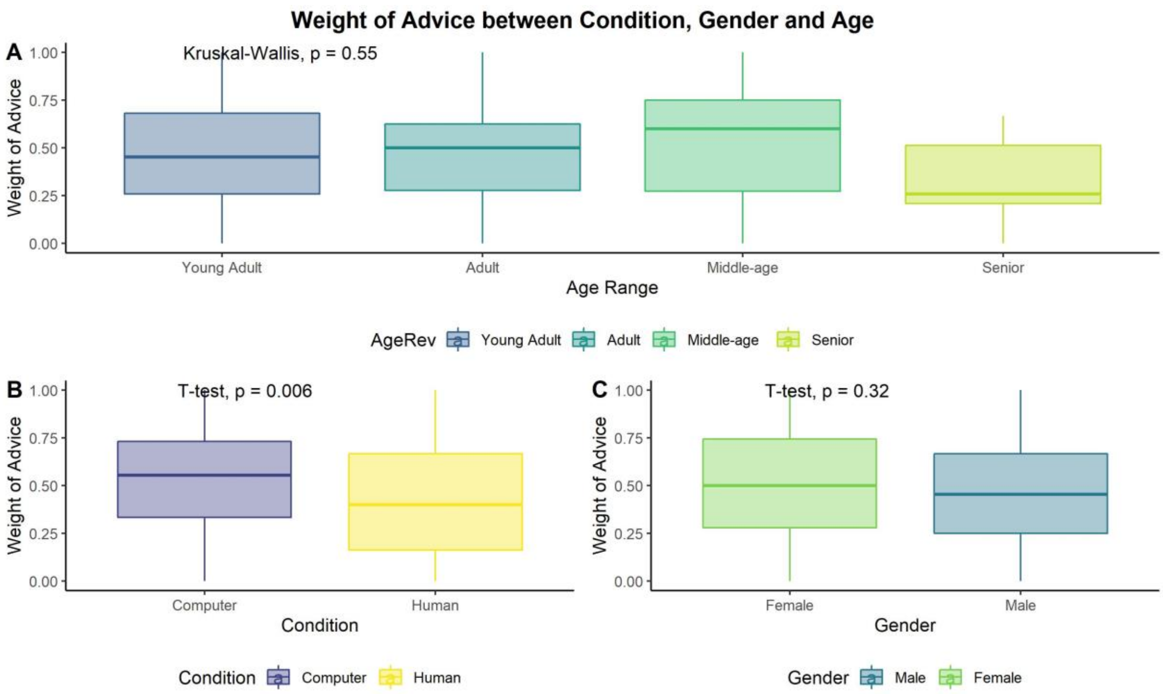

Figure 32 Boxplots of Logg's replicated results

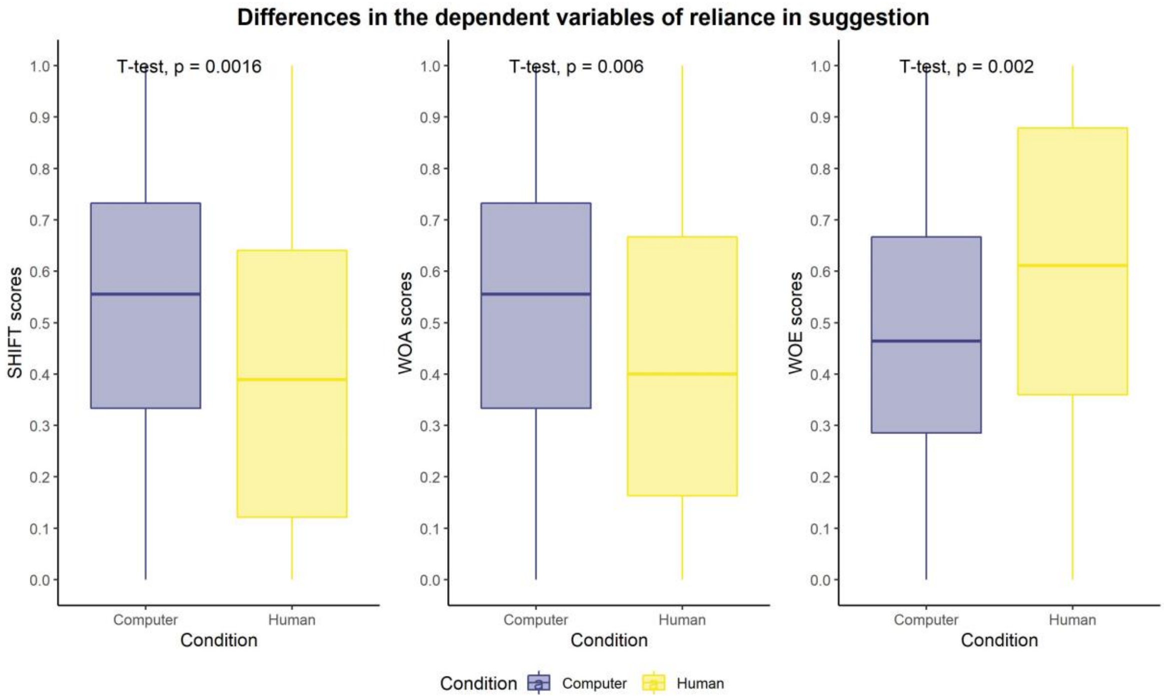

Figure 33 Reliance on the suggestion between the human condition and computer condition

Figure 33 presents the boxplots about the reliance on the suggestion according to three different indexes.

SHIFT was measured by an independent samples t-test to compare the differences between the computer and the human condition. Analyses found a significant difference between values in the computer condition versus the human condition.

Similar to SHIFT scores, scores in Weigh of Advice (WOA) showed a significant difference between the computer condition and the human condition in an independent sample t-test. This result follows Logg's finding.

These results are in line with the t-test in Weigh of Own Estimate (WOE) in which participants remained more on their initial hypothesis when the suggestion was given by another participant (human condition).

5.3. Results of emotions on choice

This paragraph presents the results about the first hypothesis about the relation between the reliance on suggestion and the participants' experienced negative emotions of regret, disappointment, guilt, and self-blame.

Results of Pearson correlation indicated that there was a not significant correlation between regret and reliance in the suggestion in human condition, $r(78) = .0.19$, $p = .097$, as in the computer condition, $r(81) = .0.03$, $p = .80$.

Pearson correlation showed a not significant association between disappointment and weight of advice in human condition, $r(78)=0.16$, $p=0.15$, as in computer condition, $r(81) = .0.16$, $p = .16$.

Pearson product-moment correlation indicated a significant positive association between the perceived guilt and the weight of advice in the human condition, $r(78)=0.26$, $p=.02$, while this association is not significant in computer condition, $r(81)=0.02$, $p=0.84$.

As regards self-blame, Pearson's r showed a positive association between the reliance on the suggestion from human and the feeling of self-blame, $r(78)=.27$, $p=0.02$, while this association is not significant for participants in computer condition, $r(81)=.03$, $p=0.80$.

5.4. Results of perception of suggestion

In this section, the association of the reliance on the suggestion with the perception of the source of the suggestion after a wrong outcome is presented. In particular, these results focused on the negative characteristics of how much participants though about the source of suggestion as inaccurate, unexpected, undependable, ambiguous, unintelligible, dishonest, and questionable.

Results of the Pearson's r indicated that there was a not significant correlation between the reliance and the perceived inaccuracy of the source of the suggestion in the human condition, $r(78) = .12$, $p = .26$, as in the computer condition, $r(81) = .13$, $p = .24$.

No significant correlations has been found between the WOA and the how much participants felt the suggestion unexpected, in both human, $r(78) = .05$, $p = .67$, and computer condition, $r(81) = -.12$, $p = .25$.

As concerns how much participants found the source of the suggestion undependable and the reliance on the suggestion, Pearson's r showed no association in the human condition, $r(78)=.1$, $p=.38$, while this association is not significant for participants in computer condition, $r(81)=.13$, $p=.23$.

The reliance on the suggestion and the perception of ambiguity about the presented advice has been tested through a Pearson's product-moment correlation and showed a not significant association in both human suggestion, $r(78)=.07$, $p=.53$, and computer suggestion, $r(81)=-.07$, $p=.53$.

A Pearson's r showed a not significant correlation between the reliance on the suggestion and how much participants' found the suggestion unintelligible in the human condition, $r(78)=.19$, $p=.09$, and in computer condition, $r(81)=-.03$, $p=.76$.

Results of the Pearson's r indicated that there was a not significant correlation between the reliance and the perceived dishonesty of the source of the suggestion in human condition, $r(78) = .09$, $p = .41$, while a significant positive correlation has been found in the computer condition, $r(81) = .26$, $p = .02$.

No significant correlations has been found between the WOA and the how much participants felt the suggestion questionable, in both human, $r(78) = .07$, $p = .55$, and computer condition, $r(81) = -.06$, $p = .61$.

5.5. Results on acceptance of the suggestion

The next paragraph will show the correlation between the three dimensions of acceptance, behavioral intention to use again the suggestion, perceived usefulness of the suggestion, and ease of use of the advice.

A Pearson's product-moment correlation found a not significant association between the reliance on the advice and the intention to use again the same wrong suggestion in both human $r(78) = -.02$, $p = .88$, and computer condition, $r(81) = -.20$, $p = .057$.

As regards the correlation between the reliance in the suggestion and the perceived usefulness of the advice, a Pearson's r showed a not significant association in the human suggestion $r(78) = -.05$, $p = .68$, and a significant negative association in the computer condition, $r(81) = -.25$, $p = .02$.

No significant correlations has been found between the WOA and the how much participants found the suggestion easy to use, in both the human, $r(78) = -.10$, $p = .36$, the and computer condition, $r(81) = .21$, $p = .06$.

6. Discussion and limitations

The experiment reported in the present chapter aimed at reproducing Logg's results on algorithm appreciation reported in experiment 1A (2019). The design followed Judge-Advisor System (JAS) design consistent with previous literature on advice research (Bonaccio & Dalal, 2006). Compared to the previous study, the suggestion given to participants was not static, but it was dynamic in a range to make the participants' estimate more reasonable. The reason was to make the participants closer to the actual weight (161 pounds) to improve the feeling of counterfactual emotions due to the "near miss" estimate. For example, if the first estimate was higher than the actual weight, the suggestion presented was 157 pounds; if the first estimate was lower than the actual weight, the suggestion presented was 165 pounds. However, this difference did not affect the replication with the previous main results. In line with Logg's findings, the presented results suggest that people relied more on algorithm compared to the advice that came from other people, even with the most straightforward description of the algorithm. As stated by Logg (2019), due to the suggestion given, closer to the real weight of the person, a greater reliance on the suggestion improved the judgment accuracy. The choice of replicate the previous study was to present an algorithm as a "black box" to allow participants the use of their personal interpretation of algorithmic advice.

To merge this reliance on algorithms with algorithm aversion, in the current experimental design, wrong feedback was presented after the second estimate to participants. After the feedback three scales measured the experienced emotions about the choice, how people felt about the source of suggestion, and a scale to collect data about acceptance of suggestion measuring the future behavioral intention to use again the same suggestion, the ease of use, and the perceived usefulness.

Results showed that participants experienced more guilt and self-blame when they relied more on human suggestion. This association was not present when they relied more on the algorithm suggestion. This result is consistent with what was found in the second experiment presented in Chapter 2: nevertheless, this result does not support the first hypothesis. As regards to the perception of the source of the suggestion, participants evaluated the source of suggestion as more dishonest when they relied more on the computer suggestion; this association was not found when people followed the human suggestion.

An interesting but not statistically significant trend that deserves more attention is the relation between the participants' reliance on suggestion and the expressed behavioral intention to use again the same advice. The more participants relied on the suggestion and less they are inclined to use again the suggestion from an algorithm. This trend was found only in the algorithm condition and supports the algorithm aversion literature. Similar results were found for the perceived usefulness, in which participants who relied more on algorithmic suggestion found the advice less useful. This association was not found for participants in the human suggestion.

To sum up, this experiment suggests that participants showed more self-blame and guilt the more they trusted human suggestion, while participants who relied more on the computer suggestion reported to find the algorithm more dishonest and showed a trend against the future use of the algorithm suggestion after a wrong outcome. Participants who showed higher reliance on algorithmic suggestion, after the negative feedback, evaluated the source of suggestion less useful compared to the people who relied upon another human. Nevertheless, these results can give suggestions to the design of intelligent automation about the role of the feedback. Insights from the comparison between the present findings and Logg's results showed that it seems that people are more willing to use algorithms when users do not see they making errors as the case when the algorithms are not visible to the final users. These findings suggest that algorithm appreciation and algorithm aversion are not two different phenomena, but they can happen after a wrong outcome, and the reliance on a suggestion has relations with the perceived negative emotions after a non-optimal outcome.

This study showed the importance of how people use and incorporate advice from the algorithmic source into their decisions. From an HCI point of view, it has a particular relevance to the result about the perceived usefulness after trusted the suggestion. This finding may shed light on the importance of the presentation of the feedback in terms of wording because besides the participants' idea about the source of the suggestion in this study the presentation of the advice was not manipulated. Participants who thought the advice came from an algorithm perceived it as less useful after bad feedback. As found in study 2, participants felt more self-blame after a wrong outcome following a human advice. This finding can be interesting for the designers of intelligent systems to put the attention on the level of anthropomorphism of the interface that can be tricky according to the field of the decisions in which the system is used. Next studies will be addressed to better understand the behavioral intention and its relationship with reliance on the suggestion.

Chapter 5: Final Discussion and Conclusion

1. General discussion

The present work has explored different aspects of the interaction with algorithms based systems focusing on the particular case in which the suggestion is not optimal in an attempt to unveil the users' experienced counterfactual emotions in order to understand the phenomenon called "algorithm aversion."

To the best of the author's knowledge, these are the first studies available on the relation between counterfactual experience and algorithm aversion. The emotions of regret and disappointment have never been studied from the lenses of HCI. These emotions are relevant and involved in the so-called "algorithm aversion" or the distrust people feel in the use of an intelligent system after see it errs.

The first experiment explored the dimensions of counterfactual emotions and the differences when the source of suggestion was an algorithm-based system or a group of human experts and the trust when participants needed to make a choice. The second theme of investigation, explored in the second study, involved a typical purchase scenario in which participants trust the same source of suggestion (i.e., algorithm or human) twice with the results of a wrong outcome both time. Finally, an evaluation of counterfactuals during a forecasting task was compared between an algorithm suggestion and a human suggestion.

These three studies were chosen to address different aspects of both counterfactual emotions and aversion to algorithms. The first experiment aimed at uncovering the users' choice and preference between an algorithm and a human suggestion. The second study aimed at the understanding of counterfactual emotions in relation to algorithm suggestion from a classical psychological perspective. The third experiment aimed at understanding from an algorithm aversion study from advice taking literature, the role of counterfactuals and other negative emotions.

In the first experiment, during a fictional "flight ticket purchase" scenario, participants were asked to make two decisions on whether to buy the flight ticket ("now" or "wait two weeks") in two different and analogous choices. The aim was to explore the underlying dimensions related to trust violation, counterfactual emotions, and the perceived attribution of responsibility for a wrong outcome. For this purpose, four linear-mixed effects models were created to explain the dimensions of regret, disappointment, internal and external attribution of responsibility. Despite the exploratory purpose, it was observed that regret, disappointment and external attribution of responsibility are affected by the dimensions of the source of the suggestion, how many times the person trusted the suggestion, when the person trusted the suggestion, the participant's general ability to understand simple mathematical concepts, and locus of control. Internal attribution of responsibility is affected by the source of the suggestion and how many times the person trusted the suggestion.

It was observed that participants had experienced similar emotions in both conditions. Even if it is not significant, there is a trend that showed higher scores in regret and internal attribution in computer condition. This result is confirmed for the higher external attribution scores in the human condition.

This experiment has some limits that could have influenced the results. One of the main limitations of this study was the story in the scenario. As reported by some participants at the end of the experimental session, the purpose was to buy a flight ticket for a work purpose in two analogous tasks. This purpose could lead to the thought that buying a more expensive ticket can be more consistent with the aim of the flight. Other participants' feedbacks were about the perception about spending more money on the same flight tickets, but they explained it as the preference to spend more to go to the work appointment is better to save money and lose the job appointment.

These limitations could explain the scores that are almost similar and averaged at the middle-lower part of the scale. Despite the limits found, experiment1 is the first attempt to consider the users' counterfactual emotions during decisions with algorithms compared to decisions with a human advisor.

The second experiment compared two purchase scenarios following a different suggestion in a repeated task. The first condition consisted of clerk advice, who suggested a non-optimal choice to the participant, while the experimental condition presented a non-optimal suggestion from a web-based system.

The results showed that people feel more responsible for a bad choice after having followed a suggestion given by a human compared to an algorithm-based system. This result has been found even aggregating the responsibility of choice and outcome and removing the outliers. That is to say that participants experienced more self-blame after a mistake following twice a human advisor. An interesting trend was found as regard regret on choice in the second task, after the first mistake, even if it is not significant. Furthermore, these results do not meet the initial expectations, and no significant differences were found in the dimensions of disappointment, regret, and guilt. Indeed, the effect of responsibility can be linked to the antecedents of regret, and these results need more attention.

The third experiment compared the reliance and discounting of advice between an algorithm and another human. The participants were asked to guess the weight of a person from a photograph, before and after a suggestion. This experiment partially replied Logg et al. (Logg et al., 2019) first experiment. As in Logg's design, the main measure was the Weight of Advice (WOA) or how much participants trusted the suggestion while the main differences between Logg's experiment are the presence of feedback (always wrong) after the second estimate and the collected measures about the emotions on the unsatisfactory outcome and the perception of the source of suggestion.

The aim was to link the higher reliance on algorithms people have compared to humans in this kind of task (Logg et al., 2019) and the higher emotions related to the violation of trust and expectations. Results found that participants relied more on the algorithmic suggestion compared to human suggestion. In particular, participants who discounted the algorithmic advice felt more guilt compared to participants who discounted human suggestion. This result can be important in the understanding of the underlying users' emotions while interacting with an algorithm-based system.

One of the most important limitations regards the important dimensions that can be related to the experience of relying on or discounting the advice and users' internal dispositions. In a work in progress study dimensions as the locus of control, personality traits, and attitude towards technology will be measured to check the individual differences.

Finally, a potential shortfall of this research that needs more consideration is that people can view the algorithm as too complex, or as a "black box," and this complexity could justify similar reactions between the two conditions. Despite these limitations, this work could be a springboard to discuss this relationship between how people deal with the algorithmic suggestion, in the particular case of a wrong outcome, and the emotional relevance for the future use and trust of algorithm-based systems.

This research is the first attempt to merge literature from algorithm aversion, decision-making, and counterfactual emotions. The present results showed that the aversion to algorithms is not a straightforward subject, as discussed in previous research. This difficulty was found in the replication phase of previous studies that sometimes have not been achieved. One of the main reason for these results can be found in difficulty to provoke actual regret and disappointment during web studies. The next works will take into account the opportunity to take advantage of a lab setting to prevent these issues.

The approaches used in this dissertation were different, and these studies encourage a broader field of research from algorithm aversion, decision-making with automation, and emotions with intelligent machines.

As stated in the literature review, regret and disappointment are not easy to distinguish because they both stem from bad decisions and disconfirmed expectancies, and often these emotions overlap (Martinez & Zeelenberg, 2015; Roese & Olson, 1995). This shared origin is one of the reasons why psychological research usually has studied counterfactual emotions through different scenarios for regret and disappointment. Less attention has been given to the negative outcome without manipulating the perceived responsibility. In this dissertation, different methods from different fields were used to address counterfactual emotions involved with computer suggestions. Despite the limitations of the results, these experiments represent the first attempt filling the gap of studying counterfactual emotions after a wrong outcome in uncertainty decision-making. Considering the findings altogether, it is possible to draw a set of general recommendations that can be useful to consider when studying the interaction with "intelligent systems" in the particular case the suggestion provided by the system is not correct.

1. **Importance of the context of the decision**, in different decisions, different perceptions of the magnitude of the error can be perceived, and this can lead to a higher or lower sense of internal responsibility and potential technology abandonment

2. **The role of counterfactuals in the interaction with intelligent machines**, studies about how people react to a wrong outcome are fundamental to design better machines
3. **Go further anthropomorphism**, previous research showed much interest in the study of users' emotions that could affect trust in virtual agents, while less attention has been given to those systems based on data that do not have anthropomorphic interfaces. As the technology evolved from mere tools to complex systems

Maintaining the human in the loop consist of thinking about users' emotions and behavioral consequences to advance the cooperation with systems based on data to improve wiser decisions.

To say it with Norman's (2010) words "we need new standards" to disentangle the important role of emotions during the use of complex systems based on algorithms.

2. Summary of results

This thesis aimed to extend current knowledge of algorithm aversion and the people's negative emotions that can be involved in the decision-making process with an "intelligent system." To the author's knowledge, this is the first attempt to merge algorithm aversion, regret and disappointment, and the relationship with decisions with algorithms.

The results in the previous sections suggest that people experience a different degree of responsibility more than explicit counterfactual emotions while they deal with a computer suggestion.

Experiment1 showed four different explanatory models that explain regret, disappointment, internal and external attribution. These models described how experienced regret, disappointment, internal and external attribution are triggered and how the source of suggestion influences the experience of these emotions. Other relevant factors are trust in the suggestion, locus of control, and the ability to apply mathematical concepts. It is essential to notice that people who followed the suggestion only the first time have a higher score in internal attribution, while participants who followed the suggestion only the second time experienced more disappointment and external attribution, regret is higher when people followed the suggestion or the first time or the second time.

Experiment 2 tested twice a purchase scenario in which participants did not have the opportunity to choose the outcome after following a suggestion given by a human or a computer.

The results showed that participants felt more self-blame when they made a mistake following a suggestion given by a human than when they followed a suggestion given by an algorithm (Experiment 2). Although these unexpected results, the explanation could rely on the perceived expertise of the advisor from the participants.

Experiment 3 replicated Logg's (Logg et al., 2019) design and results on reliance on the algorithm suggestion compared to the human suggestion. This experiment has its roots in judge-advisor system (JAS) literature (Gino & Moore, 2007; Sniezek & Buckley, 1995) and weight of advice (WOA). The task was a guess on the weight of a person showed in photo, in which the participants had to give an initial estimate, received a suggestion, and give the second estimate.

As stated before, findings corroborate Logg's results (Logg et al., 2019). In particular, people trust more algorithms, compared to human advice, when they have the opportunity to follow a suggestion. Additional findings between human and computer suggestions showed unsatisfactory evidence about the involvement of regret and disappointment. Nevertheless, there is a trend as regards the participants' reliance on the human suggestion correlates with a higher feeling of guilt and self-blame compared to participants who trusted algorithmic advice. The algorithmic suggestion was defined as more dishonest compared to human suggestion when participants moved more from their first estimate closer to the advice, and they evaluated less useful the suggestion from the system.

3. Implications and contributions

This dissertation is the first step towards understanding what emotions people experience while taking decisions with algorithm-based systems, mainly if the suggestion results in a wrong outcome. To improve the use of "intelligent systems," or systems based on big data, one cannot only focus on the accuracy of the suggestion delivered to the decision-maker but also on how the experience of negative emotions after a wrong outcome.

The study of the emotions in the particular case of wrong outcomes can help the designer of these new intelligent systems to take into account what could happen in terms of perceived responsibility of the wrong outcome and trust. Even if more research is needed to develop robust theories and guidelines for HCI, this dissertation found three guidelines for designers that need more attention in the design process of intelligent systems. The first study showed that counterfactual emotions are involved during a non-optimal interaction with intelligent systems compared to humans, and it is important to take into account how users deal from an emotional point of view a wrong outcome given by these systems. The first experiment showed that the source of the suggestion, whether it is from human or from machine learning based system, impact the experience of regret and disappointment. Hence during the design phase on these systems, particular attention needs to be given to the presentation of the feedback during the occurrence of a wrong outcome. The second experiment gave a contribution in the understanding how the experience of counterfactuals is not straightforward and is strongly related to the context of the application, that is to say, that each decision task needs particular focus on how people experience and cope with negative emotions after seeing an algorithm makes an err. The second experiment showed the importance in the design phase to analyze these users' counterfactuals feelings in the decision context in which the system is used, during the occurrence of a wrong outcome. The third study has its roots in judgment and advice taking literature, and the insight for HCI found is that people blamed more an algorithm when they provided with a negative feedback in a guessing task; that is to say that the way a negative outcome is presented to the final users can affect people future behavior and need a particular attention to prevent the abandonment of these technologies. From an HCI perspective, the third experiment gave insights about anthropomorphism and anthropomorphic interfaces. Even if it is a promising field, the results showed how people felt more self-blame when they though the wrong advice came from another human, that is to say, that is not always the best choice to present advice through anthropomorphic interfaces but it needs more attention to the context in which the decisions occur.

The growth of “big data”(Davenport & Harris, 2017) encourages the spreading of the algorithms in the everyday life of each of us. Nevertheless, many users remain reluctant to trust automation. This dissertation showed that this resistance could be related to users’ emotions and the sense of responsibility after a wrong outcome through following imperfect advice from intelligent systems. Design can have a central role in preventing the abandonment of these systems, which can be problematic in reaching better decisions approaches. Possible development in HCI with intelligent systems can be found in Feng & MacGeorge (2010), they highlighted that the perception of advisor, so the consequent emotional reactions, goes through message features and performance feedback. These aspects suggest better practice for researchers and designers of the interaction with intelligent automation.

4. Future directions

There are many exciting future directions for this research. Future works will address and improve research in advice-taking with machines through the emotional

Task dependent study

As mentioned above, even if regret has been identified as “a primary negative emotion” (Inman & Zeelenberg, 2002), it is difficult to collect data without a specific scenario to discern it from disappointment. Further experimental investigations need to be done to study these two emotions with an actual task with real users of systems based on algorithms.

The main fields in which these technologies are used are medical, financial, and legal (Gunning, 2017; Yeomans et al., 2019). The opportunity to test actual emotions and the attributed extent of responsibility with expert users can change the results presented for lay users. Different jobs have different levels of responsibility, and the magnitude of the consequences for a wrong outcome could differ significantly in the experience of counterfactual emotions.

Complexity, trust, and counterfactuals

People perceive machine learning-based systems as a "black box," either because complex, intelligent systems do not give access to understand the procedures beyond the output or the algorithms base their calculation in an opaque inaccessible even for expert programmers. Future experiments will consider the accessibility of an algorithm and whether laypeople can understand it. It can be interesting to add qualitative measures as interviews about the description of the experienced counterfactual emotions to have a more comprehensive description of the experience.

Another fascinating approach lies in Judge Advisor System (JAS) (Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001) and recent studies are giving attention to building a paradigm for algorithm aversion and constantly replicate the effect of the error and the subsequent lack of trust in non-optimal algorithmic advice. An example can be found in Prahl and Van Swol (2017), who provided further shreds of evidence for algorithm aversion in a repeated task. A possibility can be the study of counterfactuals after see an algorithm err with this design.

5. Conclusion

Intelligent machines based on algorithms and big data are more and more widespread and influence decisions in the everyday life of everyone. The majority of these applications based on data are everywhere around us. Furthermore, portable devices are now ubiquitous, enabling us to receive algorithm-based recommendations almost everywhere and for many domains of our everyday decisions.

Organizations are investing in gathering information through mining "big data" and inform human decision-makers through more and more advanced systems. Nevertheless, without understanding how people deal with suggestions from automation that leads to wrong outcomes, these systems and investments may remain unused because of algorithm aversion.

With the advent of "big data," it is fundamental to understand how algorithmic advice is perceived by users. Understand how people deal with wrong outcomes after following a wrong suggestion can save time and money for companies and reduce the frustration users can feel. It is an essential aspect of Human-Computer Interaction to learn how to present suggestions accurately according to the level of responsibility and the magnitude of the potential failure.

The usefulness of "intelligent systems" advice is undeniable. At the same time, the relevance of the context of the decision a user is facing and the perception of the source of the advice can lead to a non-optimal use of these technologies. This research raised questions about dealing with the wrong outcomes after following the wrong advice. In particular, organizations need to worry about how people cope with the wrong outcomes to prevent the abandonment of systems based on algorithms. Uncovering how emotions that are strongly related with the decision process, that is to say, regret and disappointment, and the user's perceived sense of responsibility can improve the design of these applications and will help to repair the violation of trust, without losing the advantages of these technologies and prevent abandonment.

As a final remark, it is desirable that future research on decisions with algorithms will take into account a more comprehensive perspective on users' experienced emotions and attribution of responsibility to prevent distrust in intelligent machines. This comprehensive perspective can enhance understanding of users' feelings and help designers of these systems to built better human-computer interfaces. Interaction with these new systems based on "big data" is not straightforward and a multidisciplinary approach will be necessary to gain insights from different domains. This work represents the first attempt and a step forward in this multidisciplinary direction.

Bibliography

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, *32*(4), 665–683. https://doi.org/10.1111/j.1559-1816.2002.tb00236.x

Bachrach, Y., Kosinski, M., & Gael, J. Van. (2012). Crowd IQ - Aggregating Opinions to Boost Performance. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, 535–542. Retrieved from http://www.aamas-conference.org/Proceedings/aamas2012/papers/5C_1.pdf

Banker, S., & Khetani, S. (2019). Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-Being. *Journal of Public Policy & Marketing*, *38*(4), 074391561985805. https://doi.org/10.1177/0743915619858057

Baron, J., Mellers, B., Tetlock, P. E., Stone, E., Ungar, L. H., Baron, J., … Stone, E. (2014). Two Reasons to Make Aggregated Probability Forecasts More Extreme. *Decision Analysis*, *11*(July 2015), 133–145.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bell, D. E. (1982). Regret in Decision Making under Uncertainty. *Operations Research*, *30*(5), 961–981. https://doi.org/10.1287/opre.30.5.961

Bell, D. E. (1985). Disappointment in Decision Making under Uncertainty. *Operations Research*, *33*(1), 1–27. https://doi.org/https://doi.org/10.1287/opre.33.1.1

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. https://doi.org/10.1016/j.obhdp.2006.07.001

Borst, C. (2016). Shared Mental Models in Human - Machine Systems. *IFAC-PapersOnLine*, 195–200. Retrieved from https://www.researchgate.net/profile/Clark_Borst/publication/308901718_Shared_Mental_Models_in_Human-Machine_Systems/links/58511a4c08aecb6bd8d21d54/Shared-Mental-Models-in-Human-Machine-Systems.pdf

Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry Customers don't Come Back, They Get Back: The Experience and Behavioral Implications of Anger and Dissatisfaction in Services. *Journal of the Academy of Marketing Science*, *31*(4), 377–393. https://doi.org/10.1177/0092070303254412

Bradshaw, J. M., Feltovich, P., Johnson, M., Breedy, M., Bunch, L., Eskridge, T., … Van Diggelen, J. (2009). From tools to teammates: Joint activity in human-agent-robot teams. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5619 LNCS, pp. 935–944). https://doi.org/10.1007/978-3-642-02806-9_107

Brynjolfsson, E., & McAfee, A. (2012). Race Against The Machine: How The Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and The Economy. *The Lancet. Diabetes & Endocrinology*, *2*(January), 8. https://doi.org/10.1016/S2213-8587(14)70016-6

Buchholz, V., Kulms, P., & Kopp, S. (2017). It's (Not) Your Fault! Blame and Trust Repair in Human-Agent Cooperation. Retrieved from https://www.researchgate.net/profile/Stefan_Kopp/publication/319102949_It%27s_Not_Your_Fault_Blame_and_Trust_Repair_in_Human-Agent_Cooperation/links/59915f37aca2721d9b74d3a9/Its-Not-Your-Fault-Blame-and-Trust-Repair-in-Human-Agent-Cooperation.pdf

Byrne, R. M. J. (2016). Counterfactual Thought. *Annual Review of Psychology*, *67*(1), 135–157. https://doi.org/10.1146/annurev-psych-122414-033249

Christoffersen, K., & Woods, D. D. (2001). How to Make Automated Systems Team Players, 1–13. https://doi.org/10.1016/S1479-3601(02)02003-9

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2002). *Introduction to Algorithm*. (McGraw-Hill, Ed.). New York: The MIT press.

Craig, A. R., Franklin, J. A., & Andrews, G. (1984). A scale to measure locus of control of behaviour. *British Journal of Medical Psychology*, *57*(2), 173–180. https://doi.org/10.1111/j.2044-8341.1984.tb01597.x

Crandall, J. W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J. F., … Rahwan, I. (2018). Cooperating with machines. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-017-02597-8

Davenport, T. H., & Harris, J. G. (2017). Competing on Analytics, Updated, with a New Introduction: The New Science of Winning. *September 19*.

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Massachusetts Institute of Technology. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14224511

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573

de Melo, C. M., Carnevale, P. J., & Gratch, J. (2014). Social categorization and cooperation between humans and computers. In *Proceedings of the 36th Annual Cognitive Science Conference* (pp. 2109–2114). Retrieved from https://cloudfront.escholarship.org/dist/prd/content/qt5k62p6c3/qt5k62p6c3.pdf

de Melo, C. M., & Gratch, J. (2015). Beyond believability: Quantifying the differences between real and virtual humans. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9238, pp. 109–118). https://doi.org/10.1007/978-3-319-21996-7_11

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*(3), 331–349. https://doi.org/10.1037/xap0000092

Dietvorst, B. (2016). *Algorithm aversion*. Retrieved from http://repository.upenn.edu/edissertations

Dietvorst, B., & Bharti, S. (2019). People Reject Even the Best Possible Algorithm in Uncertain Decision Domains. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3424158

Dietvorst, B., Simmons, J., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B., Simmons, J., & Massey, C. (2016). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour and Information Technology*, *18*(6), 399–411. https://doi.org/10.1080/014492999118832

Dijkstra, J. J., Liebrand, W. B. G., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour and Information Technology*, *17*(3), 155–163. https://doi.org/10.1080/014492998119526

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors*, *44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Dzindolet, M. T., Pierce, L. G., Peterson, S. A., Purcell, L., & Beck, H. P. (2002). The Influence of Feedback on Automation Use, Misuse, and Disuse. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(3), 551–555. https://doi.org/10.1177/154193120204600370

Elias, S. M., Smith, W. L., & Barney, C. E. (2012). Age as a moderator of attitude towards technology in the workplace: work motivation and overall job satisfaction. *Behaviour & Information Technology*, *31*(5), 453–467.

https://doi.org/10.1080/0144929X.2010.513419

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Feng, B., & MacGeorge, E. L. (2010). The influences of message and source factors on advice outcomes. *Communication Research*. https://doi.org/10.1177/0093650210368258

Frijda, N. H., Kuipers, P., & Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *2*(57), 212.

Galton, F. (1907). Vox Populi. *Nature*, *75*(1949), 450–451. https://doi.org/10.1038/075450a0

Gilovich, T., & Medvec, V. H. (1994). The Temporal Pattern to the Experience of Regret. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.67.3.357

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes*, *107*(2), 234–245. https://doi.org/10.1016/j.obhdp.2008.03.001

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21–35. https://doi.org/10.1002/bdm.539

Giorgetta, C. (2012). *The Emotional Side Of Decision-making: Regret And Disappointment*. LAMBERT - Academic Publishing.

Giorgetta, C., Grecucci, A., Bonini, N., Coricelli, G., Demarchi, G., Braun, C., & Sanfey, A. G. (2013). Waves of regret: A meg study of emotion and decision-making. *Neuropsychologia*, *51*(1), 38–51. https://doi.org/10.1016/j.neuropsychologia.2012.10.015

Giorgetta, C., Zeelenberg, M., Ferlazzo, F., & D'Olimpio, F. (2012). Cultural variation in the role of responsibility in regret and disappointment: The Italian case. *Journal of Economic Psychology*, *33*(4), 726–737. https://doi.org/10.1016/j.joep.2012.02.003

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of Mind Perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The Clinical-Statistical Controversy. *Psychology, Public Policy, and Law*, *2*(2), 293–323. https://doi.org/10.1037/1076-8971.2.2.293

Gunning, D. (2017). *Explainable Artificial Intelligence (XAI)*. Retrieved from http://listverse.com/

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Hastie, R., & Kameda, T. (2005). The Robust Beauty of Majority Rules in Group Decisions. *Psychological Review*, *112*(2), 494–508. https://doi.org/10.1037/0033-295X.112.2.494

Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology*, *1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Hung, S. Y., Ku, Y. C., Liang, T. P., & Lee, C. J. (2007). Regret avoidance as a measure of DSS success: An exploratory study. *Decision Support Systems*, *42*(4), 2093–2106. https://doi.org/10.1016/j.dss.2006.05.006

Inman, J. J., & Zeelenberg, M. (2002). Regret in Repeat Purchase versus Switching Decisions: The Attenuating Role of Decision Justifiability. *Journal of Consumer Research*, *29*(1), 116–128. https://doi.org/10.1086/339925

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kaniarasu, P., & Steinfeld, A. M. (2014). Effects of blame on trust in human robot interaction. *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions*, 850–855. https://doi.org/10.1109/ROMAN.2014.6926359

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less : The implications of internal vs . external attributions for the repair of trust after a competence- vs . integrity-based trust violation ଝ, *99*, 49–65. https://doi.org/10.1016/j.obhdp.2005.07.002

Kleinberg, J., Ludwig, J., Cohen, M., Crohn, A., Cusick, G. R., Dierks, T., … John, J. (2017). Human Decisions and Machine Predictions. *NBER Working Paper*. https://doi.org/10.3386/w23180

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*(3), 296–310. https://doi.org/10.1037/0033-2909.107.3.296

Kleinmuntz, D. N., & Schkade, D. A. (1993). Information displays and decision processes. *Psychological Science*, *4*(4), 221–227. https://doi.org/10.1111/j.1467-9280.1993.tb00265.x

Larrick, R. P., & Soll, J. B. (2006). Intuitions About Combining Opinions: Misappreciation of the Averaging Principle. *Management Science*, *52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

Lee, J.-E. R., & Nass, C. (2010). Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In D. Latusek & A. Gerbasi (Eds.), *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives* (pp. 1–15). IGI Global.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General Performance on a Numeracy Scale among Highly Educated Samples. *Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177/0272989X0102100105

Logg, J. (2016). *When do people rely on algorithms? ProQuest Dissertations and Theses*. UC Berkeley. Retrieved from https://escholarship.org/uc/item/5jw727t9

Logg, J. (2018). Theory of Machine: When Do People Rely on Algorithms? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2941774

Logg, J., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Loomes, G., & Sugden, R. (1982). An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, *92*(368), 805–824. https://doi.org/10.2307/2232669

Loomes, G., & Sugden, R. (1986). Disappointment and Dynamic Consistence in Choice under Uncertainty. *Review of Economic Studies*, *53*(2), 271–282. https://doi.org/https://doi.org/10.2307/2297651

MacGeorge, E. L., Guntzviller, L. M., Hanasono, L. K., & Feng, B. (2013). Testing Advice Response Theory in Interactions With Friends. *Communication Research*, *43*(2), 211–231. https://doi.org/10.1177/0093650213510938

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry*, *25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social*

Psychology, *107*(2), 276–299. https://doi.org/10.1037/a0036677

Marcatto, F., & Ferrante, D. (2008). The Regret and Disappointment Scale : An instrument for assessing regret and disappointment in decision making. *Judgment and Decision Making*, *3*(1), 87–99. Retrieved from https://pdfs.semanticscholar.org/87da/b1f618d7cf820bd9f493a6b93ef3f31dbc38.pdf

Martinez, L. F., & Zeelenberg, M. (2015). Trust me (or not): Regret and disappointment in experimental economic games. *Decision*, *2*(2), 118–126. https://doi.org/10.1037/dec0000025

Martinez, L. F., Zeelenberg, M., & Rijsman, J. B. (2011a). Behavioural consequences of regret and disappointment in social bargaining games. *Cognition and Emotion*, *25*(2), 351–359. https://doi.org/10.1080/02699931.2010.485889

Martinez, L. F., Zeelenberg, M., & Rijsman, J. B. (2011b). Regret, disappointment and the endowment effect. *Journal of Economic Psychology*, *32*(6), 962–968. https://doi.org/10.1016/j.joep.2011.08.006

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McAfee, A., & Brynjolfsson, E. (2012). Big data: The Management Revolution. *Harvard Business Review*, 1–9. https://doi.org/00475394

McNemar, Q., & Meehl, P. E. (1955). Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. *The American Journal of Psychology*, *68*(3), 510. https://doi.org/10.2307/1418552

Meehl, P. E. (1954). *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. *The American Journal of Psychology*. Minneapolis, MN: University of Minnesota Press. https://doi.org/10.2307/1418552

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology*, *4*(4), 268–273. https://doi.org/10.1037/h0047554

Moore, D. A., & Klein, W. M. P. (2008). Use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior and Human Decision Processes*, *107*(1), 60–74. https://doi.org/10.1016/j.obhdp.2008.02.005

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. Retrieved from http://www.few.vu.nl/~wissen/downloads/seminar/2000_Nass.pdf

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Conference companion on Human factors in computing systems - CHI '94* (p. 204). https://doi.org/10.1145/259963.260288

Norman, D. A. (1990). The "Problem" with Automation: Inappropriate Feedback and Interaction, not "Over-Automation." *Philosophical Transactions of the Royal Society B: Biological Sciences*, *327*(1241), 585–593. https://doi.org/10.1098/rstb.1990.0101

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. New York: Basic Civitas Books. https://doi.org/10.1145/985600.966013

Norman, D. A. (2010). The way I see it: Natural user interfaces are not natural. *Interactions*, *17*(3), 6. https://doi.org/10.1145/1744161.1744163

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390–409. https://doi.org/10.1002/bdm.637

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Ordóñez, L. D., & Connolly, T. (2000). Regret and Responsibility: A Reply to Zeelenberg et al. (1998). *Organizational Behavior and Human Decision Processes*, *81*(1), 132–142. https://doi.org/10.1006/obhd.1999.2834

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Pieters, R., & Zeelenberg, M. (2007). A Theory of Regret Regulation 1.0. *Journal of Consumer Psychology*, *17*(1), 3–18. https://doi.org/10.1207/s15327663jcp1701_6

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.06.003

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4677

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.

Roese, N. J., & Olson, J. M. (1995). Counterfactual thinking: A critical overview. *What Might Have Been : The Social Psychology of Counterfactual Thinking.*, 1–55.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sniezek, J. A., & Buckley, T. (1995). Cueing and Cognitive Conflict in Judge-Advisor Decision Making. *Organizational Behavior and Human Decision Processes*, *62*(2), 159–174. https://doi.org/10.1006/obhd.1995.1040

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes*, *84*(2), 288–307. https://doi.org/10.1006/obhd.2000.2926

Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., & Gamberini, L. (2014). Measuring User Acceptance of Wearable Symbiotic Devices: Validation Study Across Application Scenarios. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8820, pp. 87–98). https://doi.org/10.1007/978-3-319-13500-7_7

Sudgen, R. (1985). Regret, recrimination and rationality. *Theory and Decision*.

Sundsøy, P., Bjelland, J., Iqbal, A. M., Pentland, A., & De Montjoye, Y. A. (2014). Big data-driven marketing: How machine learning outperforms marketers' gut-feeling. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8393 LNCS, pp. 367–374). https://doi.org/10.1007/978-3-319-05579-4_45

Surowiecki, J. (2004). The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. *Choice Reviews Online*, *42*(03), 42-1645-42–1645. https://doi.org/10.5860/CHOICE.42-1645

Turner, B. M., Steyvers, M., Merkle, E. C., Budescu, D. V., & Wallsten, T. S. (2014). Forecast aggregation via recalibration. *Machine Learning*, *95*(3), 261–289. https://doi.org/10.1007/s10994-013-5401-4

Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2006). Supplication and appeasement in conflict and negotiation: The interpersonal effects of disappointment, worry, guilt, and regret. *Journal of Personality and Social Psychology*, *91*(1), 124–142. https://doi.org/10.1037/0022-3514.91.1.124

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010, August). Causes and consequences of mind perception. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2010.05.006

Weiser, M. (1991). The Computer for the 21st Century. *Scientific American*, *265*(3), 94–104. https://doi.org/10.1038/scientificamerican0991-94

Yaniv, I., & Kleinberger, E. (2000). Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2118

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

Zeelenberg, M., & Breugelmans, S. M. (2008). The Role of Interpersonal Harm in Distinguishing Regret From Guilt. https://doi.org/10.1037/a0012894

Zeelenberg, M., & Pieters, R. (1999). Comparing service delivery to what might have Been: Behavioral responses to regret and disappointment. *Journal of Service Research*, *2*(1), 86–97. https://doi.org/10.1177/109467059921007

Zeelenberg, M., & Pieters, R. (2004). Beyond valence in customer dissatisfaction: A review and new findings on behavioral responses to regret and disappointment in failed services. *Journal of Business Research*. https://doi.org/10.1016/S0148-2963(02)00278-3

Zeelenberg, M., Van Dijk, W. W., & Manstead, A. S. R. (1998). *Reconsidering the Relation between Regret and Responsibility*. *ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES* (Vol. 74). Retrieved from https://s3.amazonaws.com/academia.edu.documents/45775065/Reconsidering_the_Relation_between_Regre20160519-24500-l695og.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1554744610&Signature=IlqveM8CSWbZNT1tD0s3vKrAWtU%253D&response-content-disposition=inline%25

Zeelenberg, M., Van Dijk, W. W., & Manstead, A. S. R. (2000). Regret and Responsibility Resolved? Evaluating Ordóñez and Connolly's (2000) Conclusions. *Organizational Behavior and Human Decision Processes*, *81*(1), 143–154. https://doi.org/10.1006/obhd.1999.2865

Zeelenberg, M., van Dijk, W. W., Manstead, A. S. R., & Van Der Pligt, J. (2000). On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. *Cognition & Emotion*, *14*(4), 521–541. https://doi.org/10.1080/026999300402781

Zeelenberg, M., Van Dijk, W. W., Manstead, A. S. R., & Van Der Pligt, J. (1998). The Experience of Regret and Disappointment. *Cognition and Emotion*, *12*(2), 221–230. https://doi.org/10.1080/026999398379727

Zeelenberg, M., Van Dijk, W. W., Van Der Pligt, J., Manstead, A. S. R., Van Empelen, P., & Reinderman, D. (1998). Emotional Reactions to the Outcomes of Decisions: The Role of Counterfactual Thought in the Experience of Regret and Disappointment. *Organizational Behavior and Human Decision Processes*, *75*(2), 117–141. https://doi.org/10.1006/obhd.1998.2784

Zuboff, S. (1988). *In the age of the smart machine*. NY: Basic Civitas Books.