The Future of Work Environment: The Impact of Artificial Intelligence in the Employee's Levels of Psychological Safety

Tomás Bandeira

Dissertation written under the supervision of Filipa de Almeida

Dissertation submitted in partial fulfilment of requirements for the MSc in Management with Specialization in Strategy and Entrepreneurship, at the Universidade Católica Portuguesa, September 2021

Abstract

Title: The Future of Work Environment: The Impact of Artificial Intelligence in the Employee's Levels of Psychological Safety

Author: Tomás Bandeira

With the rise of automation, observed already in cars and factories, the next step becomes working side-by-side with us humans amongst our working environments. This research focuses on the potential future of workplaces where humans work alongside AI, and how impactful these changes can be on the employees' psychological safety (PS) at the company they work at. It reviews different aspects in terms of how people perceive both humans and AI in terms of social and intellectual traits and how impactful these perceptions can be to their PS levels. This study exposes participants to four different scenarios that consist in the addition of a new individual to their workplace. Each scenario differs from the other based upon two aspects: the type of entity that is added to the workplace (Human vs. AI); the position of the entity relative to the employee at the company (Colleague vs. Leader). The results suggest that although there is no direct significant difference on psychological safety between both types of entities, there is an indirect influence as different perceptions of these entities have an impact on PS. On the other hand, the position of the entity towards the self also has no significant impact, however, leadership has an impact on PS irrespectively of the type of entity. The findings are discussed and new research questions raised.

Keywords: Psychological Safety; Artificial Intelligence; AI Leadership; Perception; Work Environment.

Sumário

Título: O Futuro do Ambiente de Trabalho: O Impacto da Inteligência Artificial nos Níveis de Segurança Psicológica do Trabalhador

Autor: Tomás Bandeira

Com o surgimento da automação, já observada em carros e fábricas, é apenas uma questão de tempo até que comecemos a observar um nível mais sofisticado de automação a trabalhar ao mesmo nível que nós humanos trabalhamos nas empresas. Esta pesquisa foca-se num suposto ambiente de trabalho futurístico, onde os humanos trabalham ao lado da IA, e o impacto destas mudanças na segurança psicológica dos trabalhadores na empresa, onde estas alterações irão ser implementadas. Este estudo analisa diferentes aspetos em como nós, humanos, visualizamos os outros e a IA em termos de características sociais e intelectuais e como estas perceções podem ter impacto nos nossos níveis de segurança psicológica. Este estudo expõe os participantes a quatro cenários diferentes que consistem na adição de um novo indivíduo ao seu local de trabalho. Cada cenário difere do outro com base em dois aspetos: o tipo de entidade que é adicionada ao local de trabalho (Humano vs. IA); a posição da entidade em relação ao trabalhador na empresa (Colega vs. Líder). Os resultados sugerem que embora não haja diferença direta significativa no impacto entre os dois tipos de entidade, ainda há uma influência indireta, pois diferentes perceções destas duas entidades podem impactar o nível de segurança psicológica. Por outro lado, a posição da entidade face ao colaborador também não apresenta diferenças relativas entre si, no entanto, a liderança tem impacto nos níveis de segurança psicológica independentemente do tipo de indivíduo.

Palavras-Chave: Segurança psicológica; Inteligência artificial; Liderança de IA; Perceção; Ambiente de trabalho.

Acknowledgments

Finishing this dissertation is another completed step in my life and one I won't forget, where it marks the end of my student life and the beginning of my life as an employee in the job market. As part of my era as a student, I had the opportunity to meet a lot of different people, face a diversity of difficult obstacles and experience and create a lot of memories that will reside with me for the rest of my life. Not only that, during my experience as a college student, I was able to learn a lot with my college and was a great source of information that will be useful for the rest of my life both professional and personal terms, becoming a key era for the current person I currently define me as. Therefore, first and foremost, I thank all the people involved with my college, Católica-Lisbon School of Business & Economics, for everything they provided me with. Particularly, I would also like to thank:

Professor Filipa de Almeida for all the outstanding support and supervision she gave me, and always making sure I noticed the successful work I would be achieving and don't always think of what is yet to be done. Would also like to thank for the constant check-in with me and making sure everything was still progressing as it should.

My mom, for always being present for me and remembering where I needed to focus and not lose track of what had to be done and keeping me in check so I wouldn't deviate from my responsibilities despite the hard times we currently face due to the pandemic. Always motivating to keep doing what keeps me at my best and remembering that whatever I do, it should be something I praise and love to do.

My aunt Teresa, for being there as a second mom, that despite loving and caring for me, always kept a hard-view on life for me to make me understand that life isn't as simple as some of us believe it to be and if we want to achieve something we must work for it.

My girlfriend Joana Vieira de Castro, for motivating me to keep working and always cancelling plans, so I would focus on my thesis instead of distracting myself while hanging out with her for a lot of time. But also, for making my days brighter during times I was feeling depressed.

And finally, my friend Tiago Marques for also keeping in-check with me and all the help he offered, not only with his thesis but in any other term, so that I would always I know if I needed something he would be there to help me and remembering to always work so that I could get on holidays as soon as possible.

Table of Contents

Table of Appendices

Glossary

&	And
AI	Artificial Intelligence
PS	Psychological Safety
ML	Machine Learning
df	Degrees of freedom
F	F distribution, fishers F ratio
H	Hypotheses
M	Sample mean
SD	Standard Deviation
N	Total number of cases
n	Number of cases per condition
p	p-value
Corr	Estimation of the Spearman's correlation coefficient
R-sq	Multiple Correlation Squared; measure of strength of association
MSE	Mean Standard Error
β	Regression Model's Coefficient with Final PS as outcome variable
θ	Regression Model's Coefficient with Social Traits as outcome variable
t	t distribution, Student's t-test

1. Introduction

This thesis explores how, in an eventual future, with no specific date in mind, where technology and more specifically, Artificial Intelligence (AI) has developed in a way that we will start coexisting more frequently with AI, smart robots, not only on a daily basis in every situation, but more specifically in our work environment. How will humans as employees feel when working alongside these machines?

For years, we have worked alongside many machines, being those computers, mobile phones, that help us achieve not only personal, but also company-related goals at a faster pace. A huge example are the factories that can produce canned foods at a completely different pace compared to previous times where everything was human-made and required a lot more of human-effort, due to low levels of labour productivity and slow productivity growth (Temin, 1997, p. 5). Nowadays, however, we are starting to see a pattern building up where technology is starting to become self-sufficient. A great example are the autonomous cars, that are already being tested and becoming a part of our daily lives (Wiseman, 2021).

When looking upon these, we can actually observe that the main purpose behind these machines is to specifically target tasks that we need to be done without applying much human-effort, to make them safer and to maximize the performance for each of these, and according to Mark Zuckerberg (Founder and CEO of Facebook), this will become more of a reality in the next five to ten years, where "AI will be able to deliver so many improvements in the quality of our life" (July 2017).

Despite these great improvements there are still a lot of people that believe AI could be a huge threat for humanity as a whole. This can perhaps be explained by the impact of the media, which depicts AI as a major threat to society that will rebel against humanity (as seen in movies such as 2001: A Space Odyssey or The Terminator), but also with some examples that we've seen occur, such as two autonomous self-learning computers that were able to create a completely new language unknown to us humans (Collins & Prigg, DailyMail, July 2017). However, leaders from large tech companies disagree with this perspective. For example, Mark Zuckerberg (July 2017) still does not understand people that share this opinion on this matter and Jeff Bezos (Founder and CEO Amazon) believes this evolution will become a renaissance or a golden age that will

allow us to solve more problems that were just mere fiction for the last several decades (May 2017).

One of the main issues regarding AI, which I will be targeting during the course of this research, is how will it affect the job market (Frey & Osborne, 2017). As mentioned above, machines have been increasing their levels of participation in the main activities within companies, such as production, manufacturing and agriculture. Regarding AI, Sundar Pichai (CEO of Google) believes that no matter what, "AI is going to progress. Technology has this nature, it's going to evolve", and in a matter of years, most jobs will be automated, and we will be substituted (MSNBC, January 2018). For example, and as Bill Gates (Founder of Microsoft) said once, "We used to all be farmers, now very few of us are", tending us to the fact that despite still being able to work on these particular jobs, "there's this other way to make those goods and services", allowing us to offer higher-performance services by using AI (Mint, November 2019).

People have shown a lot of concerns with the rise of all these technologies that are stealing "our jobs" and according to Jack Ma (Co-Founder of Alibaba Group), "if you're not creative enough, your job will be taken away by a lot of machines" (CNBC, September 2017). This tends to reassure most people that work in more innovative job-related positions. However, who can guarantee that even jobs that require creativity and innovation can forever stay free from all these evolving machines? After all, not many people even considered reaching these levels of technology, so why would it stop now? And if AI is indeed about to prevail at the workplace, how will workers react to it?

That's the main purpose of this thesis, to understand and observe how people feel about working alongside an innovative and creative AI that will help build decisions and work alongside us, as if they were just another additional colleague or worker at the company we are working at. Will there be any difference when facing this kind of scenario with AI, or will it have the same impact on people's PS as another human being working with us? In this thesis, I seek to understand if people really care about this possibility and if it would impact their overall PS in their current workplace or if it would only bring losses and generate conflicts among factions: humans vs AI. I will also look into how associated traits affect these levels of PS and what kind of traits people tend to associate or deem necessary for an AI (Social vs. Intellectual traits).

1.1. Problem Statement

In order to better understand and measure in a more efficient manner how human beings feel about working alongside AI, I decided to investigate the effect that the addition of an AI to a person's current work environment would have in terms of PS, both as a colleague and as a leader, to better evaluate all possibilities within a company and compare possible effects of different roles and relationships. Knowing that perceptions of trust and fairness in AI agents can be rather low (Lee, 2018), I decided to control for the possible effects that would come with the addition of a new entity by evaluating the addition of human beings instead of AIs. The purpose of this is to verify if people's reactions will differ when facing integration of humans versus AI in their workplace. Therefore, this research intends to answer the question: "How impactful can the addition of an AI be in the work environment?". This problem statement can be divided into X main questions:

- **Research Question 1:** Does the addition of a new AI team member negatively influence employee's PS comparatively with the addition of a new human team member?
- **Research Question 2:** Does the hierarchical level of the new AI team member matter, such that an AI leader negatively influences an employee's PS more than an AI colleague?
- **Research Question 3:** Do humans negatively influence PS in the same way as AI when in a leadership position versus a colleague position?
- **Research Question 4:** Are the ideal traits for a team member/leader different depending whether such team member is a human or an AI?

1.2. Relevance

This thesis will provide an original contribution to the existing literature on AI and PS in the working environment by combining both concepts in a unique way. As one of many potential outcomes in the future, understanding how people would feel working alongside AI will be important to better understand how one can potentiate this synergy, particularly as the popularity of this trend rises. Companies should start to look into this, as this technology keeps evolving and eventually will become a common part of society at a slow pace, with developments starting to be

seen already. There is also a lack of research on this matter, further justifying the need for this study, which I hope will be used as the foundation of future studies.

1.3. Structure

In order to achieve and accomplish these thesis goals, an experimental study was conducted. After this brief introduction, we will head on to the Literature Review chapter, where we will focus on the topics of AI and its Regulation, Explainable AI, and its effect on Algorithm Aversion, PS overall and related developed scales, finishing with Fear of changes in the Workplace, to better understand the concepts mentioned and facilitate the reading, which will be followed by a methodology chapter, which includes the research strategy and design, the participants, and a description of the whole process behind this study. Then, I will present the main results, which will be analysed and discussed in better detail, followed by the main conclusions from the study. In the end, limitations of the study and future research ideas will be discussed.

2. Literature Review

2.1. Artificial Intelligence

A few years ago, people would consider the world we are living in today feels like a Wonderland, similar to the feeling when someone reads about the one Lewis Carroll describes in his famous novels. Image recognition, smart speakers, and self-driving cars, used to be all part of a dream and imagination of people, and nowadays it is all possible due to the latest advances in AI.

According to Haenlein and Kaplan (2019), AI is defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (p. 5). Despite being established as an academic discipline in the 1950s, AI has had limited practical interest for over half a century coming out of obscurity only in the last few years. With the rise of Big Data and the improvements observed in computing power, it has started to become a part of the business environment and becoming one of the main topics amidst discussions in the larger tech companies (Haenlein & Kaplan, 2019).

Since the early ages of AI, there has been one concept of Artificial Intelligence, that experts have been predicting that AI will reach, which is the one concept I will focus upon throughout this research, referred to as Artificial General Intelligence (AGI) – "systems that show behaviour

indistinguishable from humans in all aspects and that have cognitive, emotional, and social intelligence" (Haenlein & Kaplan, 2019, p. 2). For now, it is still just an expectation from those experts, however, only time will tell if this becomes a reality, but for better or for worst, this study is here to test the human reaction to a possible future under this hypothesis.

As previously stated, many assumptions have been made for a possible future alongside AGI, where some people fear them, while others embrace them. As an example, for the former case, Jack Ma (Co-Founder of Alibaba Group) said he was quite optimistic and doesn't think AI will become a threat, believing people are smart enough to learn that, while for the latter, Elon Musk, a person who has a lot of exposure to most cutting-edge AI, believes people should be really concerned about it, and thinks that "we have to be proactive in regulation instead of reactive because I think by the time, we are reactive in AI regulations, it's too late", using the extreme example of "people see[ing] robots going down the street killing people" and believing that without proper regulations and without proactive measures as mentioned, a World War III could be a major concern (New China TV, August 2019).

2.2. Fear of Changes in the Workplace

Fear is a basic emotion we have all felt at a moment or another. At its most basic level, fear is a combination of time and imagination, where we imagine something that has yet to occur that results in the emotional response we label as "fear" (Skinner 2004, p. 2).

One common fear to most developed countries is the threat people feel within their work environment resulting from interpersonal risks, and feeling threatened by possible changes within their workplace that result in high stress, such as the risk of losing their job, as some people tend to work in an environment, where this can happen to anyone at any time. However, this is not the concept that we will be looking into. As stated, the concept or type of change targeted will be regarding AI joining our workplace.

Being AI a possible outcome in the future that, although it's still currently an imagination of most people minds, some already believe it will be a realistic future, being a reality that currently promotes fear in eyes of some people and opportunity in the eyes of others, as observed already in leaders from large tech companies, but it's also believed to be shared among civilians and we will try to observe this in potential participants in our survey.

As Langer and de Certeau (1988) once pointed out, to know that the physical world will continue beyond our presence, it is an existential tragedy that foments considerable alarm on our part and although it is referred to physical rather than organizational "death", the general concern of imagining something without us is similarly felt in the possibility of a future where we are no longer needed. In case of a future with AI leading the world and substituting us in most routinely jobs (with self-driving cars becoming a reality and efficient assistants such as Google Duplex) and eventually more, people are starting to observe a future similar to this one due to involuntary job loss. With situations like this where people are situated in positions that lead to anxiety from lack of information about what is going to happen due to high levels of uncertainty.

Taking all of this into account, I will take no longer to explore what is this PS concept, how AI joining our workplace can have so much impact on this concept, compared to another human being and depending on the position it joins the company, and its importance to companies and individuals.

2.3. Psychological Safety

First explored by scholars in the 1960s and experiencing a renaissance that started in the 1990s and ongoing, PS describes people's perceptions of the consequences of taking interpersonal risks in a particular context, such as a workplace (Edmondson 1999), which this thesis will be focusing upon. According to Amy Edmondson, interpersonal risks can be seen at two different levels: at the lightest would be the risk that other people will not think well of me; and at a more extreme level would be the risk of rejection or humiliation in front of others. She even used a great analogy in an interview for Sarder TV (2015), where in evolutionary times, if we were rejected by the "tribe", we will quite literally die, this results in having us hard-wired to fear rejection and others' opinions about us. With this perspective in mind, we are fixated at an early age to always measure and build, both our reputation and image to mainly please what we believe others to think of us. And although this is quite efficient at keeping us safe from interpersonal risks, it can become quite problematic from the point of view of creating a learning organization, since it will be the main focus of people, managing what others think about them, rather than engaging with others with the intent of learning and/or teaching them how to overcome certain obstacles and adversities.

PS research for several years has tackled this issue, being a central theme that facilitates the willing contribution of ideas and action to a shared enterprise. It is now known that PS helps to explain why employees share information and knowledge (Collins & Smith 2006; Siemsen et al. 2009), speak up to make suggestions for organizational improvements (Detert & Burris 2007; Liang et al. 2012), and take initiative to develop new products and services (Baer & Frese 2002). In short, extensive research suggests that PS enables teams and organizations to learn (Bunderson & Boumgarden 2010; Carmeli 2007; Carmeli & Gittell 2009; Edmondson 1999; Tucker et al. 2007) and perform (Carmeli et al. 2012; Collins & Smith 2006; Schaubroeck et al. 2011).

PS has become a significant phenomenon in part because of the enhanced importance of learning and innovation in today's organizations, being a key part for decision making overtime and how the company operates, as its environment and culture can influence these aspects extremely. It is fundamental to reduce interpersonal risk, which is associated with uncertainty and change (Schein & Bennis, 1965; McCarty, 1967), a common obstacle companies face every day. From a practical perspective, PS has become a timely topic with the rise of teamwork within companies and associate growth of knowledge, which leads to the expectation of employees sharing newly acquired ideas and information and collaborating to achieve common goals. Schein (1993) later argued that with PS individuals are free to focus on collective goals and problem prevention rather than on self-protection, a key component for an organization's success (Frazier et al., 2016).

William Kahn (1990) who rejuvenated research in PS, proposed that PS affects individual's willingness to "employ or express themselves physically, cognitively, and emotionally during role performances," rather than disengage or "withdraw and defend their personal selves" (p. 694). This demonstrates how PS actually affects employees' performance at their work, being inefficient for the company, as lower performance from employees will result in lower performance for the company overall. Further, Khan also argued that people tend to believe they will be given the benefit of the doubt when relationships within a given group are characterized by trust and respect, therefore, when feeling psychological safe in this aspect, believing they can make mistakes without being judged, they will be more open and less afraid of being subjected to interpersonal risk.

Research has shown that individuals work environments not always make individuals feel safe and free to speak up (Detert & Edmondson, 2011; Morrison & Milliken, 2003; Ryan & Oestreich,

1998). This can become a tough issue for organizations, as offering ideas and challenging the status quo is a vital force in helping companies grow from the inside. Therefore, PS in the work environment becomes a key aspect for higher performance, better decision-making and faster development in a company.

2.4. Hierarchy Levels and Relationships at a Company

Work relationships are a common concept amidst companies' environments, coming to the form the very foundation of organizations and the contemporary embodiment of how most work gets accomplished, and there has been increased scholarly interest in this area (Ferris et al., 2009). They are such an important part of organizations that companies start to promote their development through high-quality measures as they have significant implications for the achievement of not only individual but also organizational outcomes, therefore, being damaging to all entities related when these are composed by low-quality connections (Dutton & Heaphy, 2003).

Among these relationships we tend to have different types that depend on hierarchical levels between the people that set the relationship. Amongst these types of relationships, we have two major types within the company: Leader-Subordinate relationship, and peer relationship (at the same hierarchical level).

Similar to what Furnham (2002) did to extend the "imaginative research" regarding how different perspectives affect results, conducted by Cook and Emler (1999), I decided to also extend the original research made by Moreira (2020) by further exploring a new kind of perspective that she did not look upon when researching the levels of PS resulted from having AI at a leadership position. The explored contribution for this topic was having a side-by-side perspective (as colleagues) to working with AI to observer if there would be any kind of difference towards the participant's levels of PS when compared to the already explored bottom-top perspective of having AI as our team leader. This research also relates to the important and growing literature on 360º feedback where peer evaluations are important and discriminating (Furnham & Stringfield, 1994, 1998).

The results from the study (Furnham, 2002) also point out that there are high discrepancies between boss and the rest (colleague and subordinate) and I want to further explore this difference to observe if it can be also concluded in terms of working with AI in those

relationships. These discrepancies were focused not only in terms of perceptions, but also desired traits in each of these entities.

2.5. Perception of AI Traits

For a long time, research has been conducted and literature has been growing on the relationship between personality variables and behaviour at work (Roberts and Hogan, 2001). Not only that, but numerous studies have also shown how personality traits have a significant impact on predicting job satisfaction and motivation (Furnham, 2002). For instance, Furnham et al. (1999) found that extroverts were more sensitive to job motivators, while neurotics were more sensitive to hygiene factors at work (cit. in Furnham, 2002). This shows how certain personality traits might relate to the way people tend to evaluate potential bosses and peers and how we feel about these might end up having a potential influence on PS when reflecting on the interpersonal risks, already mentioned.

Mou and Xu (2017) developed research to understand and compare the initial interaction between two relationships: human-human and human-machine. In this study, they placed participants in two different scenarios: one where half the participants would speak to a specifically designed chatbot called Little Ice, which resembled a "17-year-old girl with a lively and outgoing, sometimes naughty personality" through WeChat, while the other half would speak to human friends. The results, although somewhat expected, were still surprising. Overall, "users were less open, agreeable, extroverted and conscientious when interacting with AI.". This being a potential reason behind us humans believing AI to be a less social being than a human and having higher difficulties to feel at ease and trust this technology.

Moreover, personality traits impact trust (Sharan & Romano, 2020). Another key aspect that built trust towards machines and AI is these machines' ability to express social behaviours (e.g., turn-taking, emotional expressions) and anthropomorphism (Sanders et al., 2011).

To extend on this aspect and further explore personality traits in this study, I decided to assess the personality traits associated with AI agents, following the seminal paper by Rosenberg et al. (1968), where they offer a list of personality traits and whether these represent social or intellectual traits.

2.6. Literature Review Summary and Hypotheses

AI can end up creating a major impact on how safe people feel in their workplace as a possible outcome for the future. How people will feel about when sharing their opinion and beliefs at the company and how they will feel in terms of taking interpersonal risks when facing an AI in their workplace instead of another human being is a relevant and timely question. Based on the revised literature, I hypothesize that:

H1) *AI negatively influences employee's PS compared to a Human.*

The influence of AI leadership in the overall effect on an employee's PS has already been investigated (Moreira, 2020). However, there are two ways (out of many) that are possible for an AI to be related to an employee within a company, both as a Leader, but also as a Colleague. Knowing that people perceive company associates differently depending on their relationship type in the company's hierarchy, as mentioned in the "Hierarchy Levels and Relationships at a Company" section, I hypothesize that:

H2) *Having AI as a Leader has a higher negative influence on PS than having AI as a colleague.*

Feeling more in touch with this change and being recognized as a common event in every single workplace, compared to AI, the addition or replacement in any kind of position by a human being instead of AI is perceived as something more natural despite still resulting in possible fears in the workplace, as seen in the section "*Fear of Changes in the Workplace*". Then, I hypothesize that:

H3) *AI has a bigger impact on an employee's PS than a human being when in a leadership position compared to a colleague position.*

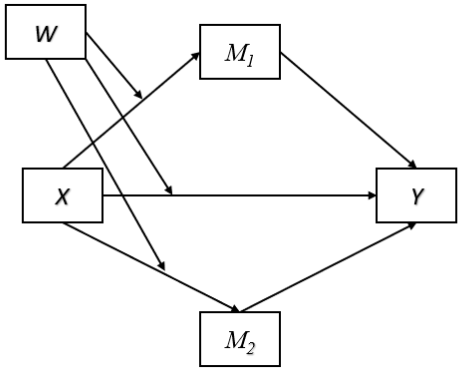
Furthermore, how we, human beings, tend to perceive others and how we characterize them in terms of personality traits can have a direct impact on how safe we feel towards the other entity, looking into deeper detail on the intellectual and social ones, as explored by Rosenberg et al. (1968), therefore, based upon this idea, further explored in the "Perception Levels of AI Traits", I hypothesize that:

H4a) *More Social traits associated with the entity, positively impacts PS.*

H4b) *More Intellectual traits associated with the entity, positively impacts PS.*

Along with these research questions and corresponding hypotheses, I will explore a little bit deeper the perception levels of traits in AI in terms on how necessary certain types of traits are compared to how we associate with these entities. Check section "*Exploratory Analysis*" for more information.

2.7. Conceptual Model



The conceptual model can be seen in the picture above and it is important to mention that while looking through all conceptual models built for SPSS and SAS by Andrew F. Hayes and The Guilford Press, I wase able to find a specific model that would take into account all the desired effects that I wanted to measure for my thesis. Therefore, I will be using this conceptual model (model 8), that will allow us to have more than one mediator (M_1 and M_2). In the Methodology section, each variable will be explained. M_1

3. Methodology

3.1. Research Strategy and Design

In this research I aim to test the overall effect and difference in an employee's level of PS after a certain change in their current work environment (or the last they were a part of) has occurred. Simply put, this change is represented by the addition of an entity in the participant's workplace at a given position. There are a total of four different scenarios (entity: human vs. AI) x (position: leader vs. colleague), each of them separately representing this change, where we evaluate the participant's level of PS before and after. All participants were randomly assigned to each of these scenarios, to increase the study validity (Malhotra et al., 2017) and an experimental study was made as it is the most common way to test for causality in hypothetical situations (Malhotra, Nunan, & Birks, 2017). To avoid the possibility of transference of knowledge and influence (bias) from one scenario to another, a between-subjects design was conducted, allowing for possible comparison between participants from different scenarios. This study was designed in the Qualtrics platform, an online survey tool.

While analysing the research questions, I manipulated not only the independent variable (type of entity), but also the moderator (type of position) with the intent to understand how that

manipulation would affect the dependent variable and explain those relationships. In this study, I measured the impact of two different mediators to explain the relationship between the independent and dependent variable. The mediators were the intellectual and social traits that each participant associated with the joining entity. This would help explain how each of the four different scenarios can be influenced or not by how people perceive the entity joining their workplace. Note that, both mediators are represented in the same exact question of the survey with trait options presented to participants in a table. For more information, check *Appendix 1: Survey*.

3.2. Participants

Participants were invited via word of mouth, social networks groups on Facebook, Survey Swap, and Survey Circle. The recommended minimum sample size for an experimental study is 30 participants per cell (Van Voorhis & Morgan, 2007) and when looking for participants, I aimed to, at least, get 50 participants per cell, i.e., a total of 200 participants across all scenarios, to increase the likelihood of incurring in type II error. A total of 220 valid answers out of 242 total answers were collected, for the four scenarios assuring more than the required minimum. Considering the relevant sample, 64,1% were female, and the average age was 29,34 years (SD = 11,97). Most participants have a bachelor's degree (40%), and even though the most common nationality was Portuguese (42,3%), there were still a big part of participants from both the United Kingdom plus Northern Ireland (20%) and the United States of America (10,5%). Most participants were still students (32,3%). Among participants that were currently working in a sector, most participants worked in the financial sector (10,9%), followed by the retail and consumption sector (10,5%). This can be seen somewhat when looking at the employment state of people, where almost the majority of people are still studying (49,5%), with 41,4% of participants being fully employed and 20% working part-time. For detailed information, please see *Appendix 2: Study Overview*.

3.3. Procedure

After reading the informed consent form, with minor details about the study to avoid biases, and accepting to proceed with the survey, participants were asked to rate their agreement to a set of statements, which was used to measure the PS each participant felt in their current work environment. If the participant was unemployed, as it could be a huge risk due to the ongoing

pandemic, or a student, we asked them to answer to this scale based upon their last interaction within a workplace (previous job).

For the study, the initial level of PS was measured in order to evaluate the impact that the randomly selected scenario, that were to follow, would have in this variable in this work environment. After the initial measurement of PS, we assigned each of the participants to one of the four total scenarios built, to measure their impact on their PS, also in their current workplace. The scenarios displayed were: having an addition of an AI or a human colleague to their work team or having a replacement of their team leader with an AI or human. The scenario was induced to each participant by showing them a diagram that demonstrated their hierarchical relationship with the entity added to their workplace. The participant was highlighted with a blue box in that diagram, while the entity added was highlighted with a red box. Therefore, they were asked to imagine themselves in that exact same situation and based upon the PS scale, tell us how they feel about their specific scenario, similar to the first scale.

After measuring the new value for PS, participants were asked to select from a list of traits (both related to intellectual and social sides) the ones they associated with the new entity and the ones they deemed necessary for the new entity to have. These questions were asked not only to observe if the different types of traits would have an impact on the overall value of PS and to observe if there is a discrepancy between what people tend to associate with humans and AIs at first sight and if they require the same necessities towards both these different entities. Within both PS evaluations, we included attention tests, to be able to identify careless answers within these scaled items questions (Meade & Craig, 2012).

At the end of the survey, we asked each participant to answer to demographic questions to have a better grasp of the kind of participants we are facing in our study, while also asking questions to understand if they consider this change to make sense in their current workplace and to understand if these would be realistic in their opinion. Finalizing the survey, we offered the option for the respondent to share with us, in more detail, how they feel about this change, followed by the debriefing and a final acknowledgment of the study's goals. For detailed information, please see *Appendix 1: Survey*.

3.4. Measurement Variables

3.4.1. Independent Variable – Entity Manipulation

Human vs. AI (Entity): For this study, we decided to manipulate the independent variable entity, where we would randomly select, in an even way between four scenarios all participants, where the entity that would be added to the company would be either human (control group) or AI (the experimental condition) with 50% of probability for each one – 2 out of 4 scenarios, each depending on the joint with the moderator variable. To submit participants to this manipulation we provided these different scenarios at random with even odds, where, at least, one person in a group of four would be submitted to each individual scenario separately. In each of these scenarios, the participant would be told that there would be an addition to their team of either an AI or human, depending on the chosen scenario, or a substitution to their team leader at their current workplace. The objective with this would be to observe if the manipulation of the entity variable would make any kind of difference to the PS of each participant. For more information, check *Appendix 2: Study Overview*.

3.4.2. Moderator Variable

Colleague vs. Leader (Position): Besides manipulating the entity of the new member in the participant's workplace, we also decided to manipulate the position of that entity in the company, colleague versus leader. This moderator helps us understand if there would a different impact if a normal human being (the participant) would feel safer in a situation where they would be under the influence of an AI or working alongside one. For more information, check *Appendix 2: Study Overview*.

3.4.3. Dependent Variable – Final Psychological Safety

Final Psychological Safety: In this experimental research design, we decided to evaluate participant's PS based upon one of the four randomly selected scenarios we mentioned previously, where each participant would be told that at their current workplace there would be either a substitution in the team's leadership or an addition of a new member in their team. They were asked to describe their level of conformity with seven different statements, taken from Amy Edmondson scale to measure PS (1999), which has been used across several studies (Bornemisza, 2013; Moreira, 2020). Based upon on each of the given answers on the Likert-scale, we made an average of the values from Completely Disagree to Completely Agree (respective numerical values: 1 to 7). Note that, when accounting the values of the negative statements, before including them in the average, we first recoded their values by inverting them as their original

values affect the PS value negatively (from 1 to 7, 2 to 6, and so on). For more information check *Appendix 2: Study Overview*.

3.4.4. Psychological Safety Scale

To be able to measure PS with certain precision, several scales were created based on a lot of research and practical implications (Edmondson 2004, p.42). In this thesis, we will be using the one that was built in 1999 by Amy Edmondson (*Appendix 6: Psychological Safety Scale*), which is composed by seven questions. Although, there are a lot of different scales that were built to measure PS, being one even more recent, we will be using this in specific as it tends to be the most reliable across several studies seen during our research, as it displays internal consistency reliability and was able to predict team learning behaviour and team performance (Edmondson 2004, p.11) within several studies (Carmeli, Brueller & Dutton, 2009; Carmeli, Reiter-Palmon & Ziv, 2010), showing its efficiency in evaluating how PS can have an actual impact on the company's overall performance, becoming a reliable scale for future measurements in our study.

3.4.5. Mediator Variables

Social & Intellectual Traits: In our model, we will have two different mediators based upon two types of traits (Social and Intellectual), that were acquired by asking each participant what kind of traits they associate with the entity added to their work environment. The reason behind this decision is the fact that even though people tell us no never judge a book by its cover, people often form opinions about the characteristics of other from single, static samples of their appearance, based upon the first thing they interact in any way possible with that entity, happening when and even before we meet that person (Hassin & Trope, 2000; Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz, 1996). These inferences occur spontaneously and rapidly (Ballew & Todorov, 2007; Bar, Neta, & Linz, 2006; Rule & Ambady, 2008; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006) and there is recent evidence that suggests these impressions tend to impact decision that people make facing different situations (Cit. in Olivola & Todorov, 2010). Based on this, we believe that the way our participants will associate certain traits with the entity they are faced in their particular scenario will have an influence in their PS and how they feel about them, influencing their answers.

In order to divide the traits into two groups, we used the two-dimensional configuration of 60 traits based upon social desirability and intellectual desirability (Rosenberg et al., 1968). It is

important to mention that most traits used for the study are not included in this two-dimensional axis, and to solve this matter we decided to associate them with other traits that were synonyms to define the location of each trait within the two-dimensional space. The authors also mention that there is a relationship between these two kinds of traits, where "a trait that was intellectually desirable (undesirable) was also likely to be judged as socially desirable (undesirable)", explaining why some traits associated to one group could also be in the other group. Therefore, we tried to associated them based on their exact position in the axis and closeness to each of them. Upon this, we have separated the 24 traits into the two groups, having a total of 14 Social and 10 Intellectual traits (examples: Helpful, Honest, Humble and Problem-Solver, Professional, Punctual, respectively). Take note that the twenty-fifth "trait" considered as "Other" for people to fill in, was also considered for this study, where we individually evaluated the most recommended category (social versus intellectual) for each of the recommended traits given by the participant.

3.4.6. Covariables

Initial Psychological Safety: To evaluate the total impact of the PS (final) after each scenario was induced to the participant and understand how big of an impact this addition would have in a person's overall PS, we decided to measure the PS of each participant before the new scenario was presented to the participants, the change in their workplace. Using the same seven-questioned scale built by Amy Edmondson (1999) that we used when measuring the Final PS, we measured the PS total value by making the average of the results, just as when measuring our dependent variable from our study. Similar to our dependent variable, we had to recode the negative statements, so their values would make sense when measuring the level of PS.

Change is Realistic & Change Makes Sense in Current Workplace: At the end of the survey, we asked participants to answer if this addition would be realistic in their current workplace and if it would make sense. The mindset behind these questions was to understand if by having people believe this would never become a reality in their environment, the condition they were submitted into would have no impact on their overall PS. Similar to the aspect of first impressions, mentioned in the mediators' variables chapter, where participant's perception of not seeing this as a reality, it was suspected such variables could impact their feelings about the topic. Thus, they were both considered as covariables.

Demographic Variables (Age; Gender; Education): When studying American attitudes and trends towards AI, Zhang and Dafoe (2019) examined support for developing AI on eleven different demographic variables, including age, gender, and education. Among these, seven of them were deemed as significant predictors of support for developing AI, which based on perception levels, might influence how psychological safe people might feel about possible inclusions of AI in their workplace. There were three that were more salient than the others, however, only two were obtained in our study: *Gender* (Male versus female) and *Education* (high school and less vs college graduation); and we also acquired information on significant demographic variables such as *Age* (Millennial and post-millennials vs. Gen. Xers and Baby Boomers). To be able to measure these covariables more efficiently, we will create dummy variables for each of them.

3.5. Exploratory Analysis

During this study, we had the opportunity to ask participants what kind of traits they would associate with the entity added to their work environment, but also the ones they would deem necessary for that new entity to have in order to make them feel safer in their workplace. In this exploratory analysis segment, we will look at the necessary traits, both social and intellectual, as done prior. The objective is to understand if there is a major difference on what they deem necessary for both these types of entities and positions. Not only that, we will further explore this analysis by comparing the effects of both these between-subject factors (*Position* and *Entity*), plus the type of traits we are facing (*Social* and *Intellectual Traits*) between the associated traits with the individual added to the workplace and the traits we deem necessary for that kind of scenario. To obtain these results, we will resort to an ANOVA Two-Way Repeated Measures, with within-subject factors being *Identification* (*Associated* versus *Necessary* traits) and *Type of Trait* (*Social* versus *Intellectual Traits*), and with between-subject factors being *Entity* and *Position*. The objective behind this is to understand if there is an actual difference between each kind of trait when we are identifying what we believe to be true (*Associated Traits*), and what we want to be true (*Necessary Traits*), depending on the kind of scenario and the type of traits.

4. Results

4.1. Data Preparation and Cleaning

220 participants fully answered the study. We excluded a total of 55 participants for failing so satisfy the attention checks. For detailed analysis, check section "*Attention Check*" and *Appendix 5: Attention Check*. Furthermore, we also decided to exclude two participants that were underaged (Ages: 12 & 17), as we considered them to not be legally able to work in any possible environment and, traditionally, in many Western societies, underaged people are often denied autonomy rights, as they are not yet regarded as having obtained full personhood (Spencer, 2000, as cited in Chabot, Shoveller, Spencer, & Johnson, 2012). Note however, that one of these had failed both attention checks, not being already a valid participant for our analysis, therefore, we excluded one more participant (*check Appendix 3: Demographics*).

For all the remaining 164 participants, all negative scored items were recoded (for more information check section "*Scale Reliability*"). Furthermore, I computed the average of all these items for each scale to obtain the total value of PS by doing the average of all scores per item. All considered *Intellectual and Social Traits* were also summed up together (all 25 traits listed when they had any value) and additional variables (dummy variables) were created to identify when facing a different kind of *Entity* (Human versus AI) and *Position* (Colleague versus Leader).

Note that despite having a lower number of viable participants to achieve the desired results, it still satisfies the recommended minimum sample size for an experimental study of 30 participants per cell (Van Voorhis & Morgan, 2007).

4.2. Scale Reliability

Although the scale used throughout this experiment has been confirmed as reliable by the literature (Edmondson, 1999; Moreira, 2020), I still conducted a reliability analysis by computing Cronbach's Alpha. Having obtained two different alpha values for each scale represented in the first scenario (before the change) and the second scenario (the agglomeration of all 4 different induced scenarios/changes), none of them was able to surpass the reasonable goal of 0.8 (Gliem & Gliem, 2003). Despite that, one of them achieved an acceptable score (≥ 0.7) with the main scenario, equal to all participants, scoring an alpha of 0.7. For the second scenario, however, a lower score of 0.653 was achieved, being a questionable score ($\geq 0{,}6$). Note that both scores can still be improved by removing one of the scale items. For the case of the first scenario, we can

improve the alpha by removing the statement "Members of the company are able to bring up problems and tough issues"; while for the case of the second scenario we would be removing the statement "People on this team sometimes reject others for being different.", although, the differences observed are not that impactful reaching values of 0.713 and 0.655, respectively. To avoid possible negative alpha values for each of the scales (due to negative covariances), the considered negative statements, previously mentioned in the "Dependent Variable – Final Psychological Safety" section, were used in their recoded version, instead of the original ones. For more information check *Appendix 4: Reliability Statistics*.

4.3. Attention Check

To examine the participants' attentiveness to the overall study, two attention checks were included in the study. Each one of these were implemented in each of the PS evaluations in the two scenarios each participant was submitted to (current PS level in their workplace and PS level evaluation with the change in their workplace). In both attention checks, participants were asked to answer the same exact option "Somewhat disagree" (*Appendix 1: Survey*). Participants that ended up failing at least one of the two attention checks, they would be considered to have failed this manipulation check. If participants managed to select the right option for both cases, then they would be considered to have passed this attention span check. All participants that ended up failing at least one of the attention checks, were excluded from the study as it was clear the intention behind this, as it was asked for the participant to choose a specific option, in order to observe if they were paying attention. Specifically, we asked the participants to "Please choose the «Somewhat disagree» option, so I know you are paying attention.", where the purpose behind the question is clear. Therefore, if they failed both tests it was obvious they were not paying attention; if they failed one test, it was considered that the other attention check that they managed to get correct was of the result of chance. For more information, check *Appendix 5: Attention Check.*

4.4. Hypothesis Testing

To test some of the previously mentioned hypotheses, I conducted a regression model built by Andrew F. Hayes and The Guilford Press (for detailed information check section "*Conceptual Model*"). The chosen model to better adopt the idea for this study was model 8, where we

included the dependent variable, plus an independent one, one moderator, two mediators and some covariables.

It's important to mention that when we look into this analysis, excluding the non-workers, there is no major difference between the results obtained for the exception of two that are related. When measuring the effects of *Final PS*, the regression model that includes non-workers has *Intellectual Traits* as marginally significant, while the one without it becomes no longer significant. The opposite can be said regarding the covariable *Age*, which, in our main model, is not statistically significant while in our model without non-workers, it becomes marginally significant. This is explained by the fact that in both models, when measuring for mediator *Intellectual Traits*, *Age* is either statistically significant or marginally significant. Knowing that most of the non-workers tend to be students, it makes sense that *Age* makes a difference in this aspect, however, not that significant for the overall conclusion of this study. For more information, check *Appendix 9: Regression Model without non-workers*.

4.4.1. AI vs Human on Psychological Safety

Hypothesis 1 predicted that participants facing a scenario where an AI would be added to their workplace would decrease their level of PS, when compared to another human being included.

As mentioned above, to be able to test this hypothesis, a regression analysis was carried out to predict participant's level of PS after the scenario was introduced to them. As a covariable, we included the level of PS before the induction of the scenario to be able to also compare these values, along with two other covariables, one in which we measured if participants deemed the scenario to be realistic (*Reality*) and another to understand if they considered to be applicable in their workplace (*Sense*) and some demographic variables (*Age*, *Gender* and *Education*). The results of the regression indicated that the whole model explained 68,34% of the variance and that it predicted the level of PS of a given participant after undergoing the randomly selected scenario, $R^2 = .6834$, $F(11, 152) = 12{,}11$, $p < .001$.

When looking into the diverse variables included in the model, we can observe that the type of Entity (AI vs human) is not statistically significant to explain the level of PS of an individual when facing the addition of a new member in their current workplace, independently of their position. Despite AI being supposedly negative for the overall level of PS with $\beta_1 = -.0508$, it is not a statistically significant result for our model: $t(152) = -.2861$, $p = .7752$. Therefore,

hypothesis 1 is not supported by our results. For more detailed information, check *Appendix 8: Regression Model*.

4.4.2. Leader vs Colleague on Psychological Safety

The second hypothesis focused on the aspect of the position of the given entity included in the workplace of the participant. Although there are a lot of different relationships we can have with another person at the company, we only observed two of them (Colleague vs. Leader), when trying to dive deeper into the research made by Moreira (2020). Given the results obtained from the exact same model, we can observe that this variable is also not statistically significant to predict our main variable: the level of PS after the scenario was induced to the participant. Looking at the results, being a leader compared to a colleague led to higher PS ($\beta_4 = .1687$), however, not significantly so, therefore our second hypothesis is not supported by our results: $t(152) = .9984$, $p = .3197$.

Another relevant aspect we should dive a little deeper is the conditional effect of this moderator in the model. When looking into it and observing the effects of both possible outcomes of the variable Position, we can observe that adding a colleague to a particular team (of both types of entities) does not lead to statistically significant differences, having no direct impact in our dependent variable: $t(152) = -.2861$, $p = .7752$. However, when looking into the effect of an entity being in a position of leadership towards the individual in study, we can observe that AI has a marginally significant negative impact ($\alpha_1 = -.3355$) on the individual's level of PS: $t(152) = -1.8462$, $p = .0668$. According to Pritschet, Powell, and Horne's (2016), marginally significant is associated with near-threshold p-values, these being slightly above the 0.05 p-value. Although this is considered to be a gray area, it might still be interpreted as equivalent to a significant result, despite some precautions still need to be done. Despite this, and considering this result to be significant, it replicates what Moreira (2020) achieved in her study by finding that the role of leadership has an impact on the overall levels of PS by itself.

Concluding, when looking at hypothesis 2, the results do not support this hypothesis fully. However, although having a new entity entering the workplace of an individual as a colleague has no significant impact in their level of PS, when they enter in a leadership position towards that individual, it will have a marginal negative impact on their PS level. For more detailed information, check *Appendix 8: Regression Model*.

4.4.3. Interaction between both Independent Variables (Entity x Position)

For the third hypothesis, we dive a little deeper in the analysis of our model regarding both independent variables. By inputting the variable Position as our moderator in this regression model, we are able to observe an interaction between our moderator *Position* and our independent variable *Entity*, mentioned in the two previous chapters, both not being statistically significant for this particular model. However, despite observing an even smaller p-value for this interaction when compared to both variables individually, we still do not find this to be statistically significant for our predicted model: t(152) = -1.2310, p = .2202. Therefore, the results do not support this hypothesis. For more detailed information, check *Appendix 8: Regression Model*.

4.4.4. Intellectual & Social Traits

The fourth hypothesis that we wanted to confirm was regarding the kind of traits people associate with the entity that has joined their workplace, by checking if there is a differentiation between humans and AI.

Looking at our main model obtained using the dependent variable as the outcome, we can observe that people associating social traits (our path *a* for this hypothesis) with the specific entity introduced in the different scenarios has statistically significant impact on their level of PS, and the more social traits they associate with it, the more psychologically safe they feel: β_2 = .0882, t(152) = 3.3159, p = .0011. On the other hand (path *b*), when we look at the effect of intellectual associated traits with the specific entity, we observe that, for this particular case, the impact is only marginally significant: t(152) = -1.7410, p = .0837. Despite it being only a marginal impact, we can still observe that unlike the associated social traits, the impact of intellectual traits has a negative impact on the level of PS: β_3 = -.0530. Looking upon these results, we can observe that for the fourth hypothesis, path *a* is fully supported by the results achieved, while path *b* is not fully supported as it does not have a positive impact on PS, but rather negative. For more detailed information, check *Appendix 8: Regression Model*.

While considering the hypothesis mentioned above and knowing that both kind of traits have a significant impact on the levels of PS of an individual, we can check if these are separately impacted by our independent variables (Entity & Position and also the interaction between those), which would allow us to conclude that in some way, both independent variables have an indirect impact on the overall level of PS.

Therefore, when we look at the specific regression model with the outcome variable being the associated social traits, we can observe that although both the type of position of the entity towards the participant and the interaction between both variables are not statistically significant (respectively): t(154) = -1.3188, p = .1892; t(154) = 1.2129, p = .2270; when we look at the type of entity we face in each of the scenarios, we can observe there is a statistical significant impact where we observe that when faced with AI, people tend to associate less social traits to that entity compared to humans, which goes to support our hypothesis mentioned in the beginning of this chapter that, in terms of being either AI or human, there is a statistically significant difference in the amount of social traits we associate with each of them: θ_1 = -1.9342, t(154) = -3.6675, p = .0003. For more detailed information, check *Appendix 8: Regression Model*.

When looking into the intellectual traits as an outcome model, we can see the exact same pattern as with the one focusing on social traits. Looking at *Appendix 8: Regression Model*, we can observe that out of the three main possible effects of these regression models with the mediators as the outcome variables, both the type of position of the individual entering the workplace and the interaction between the type of position and type of entity is also not statistically significant: t(154) = -.5463, p = .5856; t(154) = 1.0720, p = .2854; respectively. On the other hand, the type of entity entering our workplace does have a statistically significant impact on the intellectual traits we tend to associate with it: t(154) = 2.7382, p = .0069. This impact is also positive (λ_1 = 1.2618), meaning that we tend to associate higher intellectual traits with AI rather than other human beings, perhaps due to the fact that we believe machines to have the power to access information quicker due to modern technology.

4.5. Exploratory Analysis

To finalize the whole chapter relative to the results, we will now focus upon the results regarding the exploratory analysis. For this analysis, I wanted to observe if the type of entity and position influence the personality traits, we deem necessary for an individual in each of the four scenarios induced, similar to what we did to the traits we associate the individuals with. Furthermore, I also want to observe if there is a significant difference between the associated types of traits and the necessary ones.

Looking upon the table "*Test of Within-Subjects Effects*" conceived from this analysis (*Appendix 10: Exploratory Analysis)*, we can observe that the main effect in the type of *Identification* of the traits (*Necessary* or *Associated*) is statistically significant, meaning that there is an actual

difference when facing Associated or Necessary Traits: $F(1) = 5.826$, $p = .017$. Not only that, but the interaction between this variable and the type of *Entity*, is not statistically significant, while the interaction with *Position* is marginally significant: $F(1) = 0.021$, $p = .886$; $F(1) = 2.651$, $p = .105$; respectively.

When focusing upon the *Type of Trait*, similar to *Identification*, we observe that its main effect is also statistically significant: $F(1) = 18.223$, $p < .001$, as already observed in the model produced. As also seen in the regression model, we can observe that the main effect of the interaction between *Type of Trait* and *Entity* is statistically significant, while the interaction between *Type of Trait* and *Position* is not: $F(1) = 29.313$, $p <= .001$; $F(1) = .006$, $p = .94$; respectively.

More importantly and to better understand the difference in *Identification*, it's important to measure the main effect of the interaction between the *Identification* and the *Type of Trait*. Looking closely to these values, we can also observe that it is statistically significant: $F(1) = 136.214$, $p < .001$. Looking into the plot in the same appendix, the estimated marginal means for the *Type of Traits*, depends on the *Identification* of these traits, where they move differently from each other (*Social* versus *Intellectual Traits*). The same can be concluded when both variables also interact with the type of *Entity*: $F(1) = 26.421$, $p < .001$. While we cannot observe the main effect of both interactions in which we include *Position* and *Entity* plus *Position:* $F(1) = .811$, $p = .369$; $F(1) = .092$, $p = .762$; respectively.

Despite the interaction effect between both variables *Type of Trait* and *Identification* being statistically significant, we still cannot say the same thing regarding the whole main effect. To obtain more information and results regarding this, I decided to make two paired sample t-tests, with the intent to see if there are significant differences when facing different values for *Identification* and also for the *Type of Trait*.

When pairing by isolating *Identification*, we can observe that the difference between *Associated Social Traits* and *Necessary Social Traits* is statistically significant: $t(163) = -8.459$, $p <= .001$. The same can also be said regarding the difference between *Associated Intellectual Traits* and *Necessary Intellectual Traits*: $t(163) = 5.8624$, $p <= .001$. This then helps us conclude that there is a main effect for the *Identification* of traits. The same can also be said about the main effect of *Type of Trait* when pairing for *this variable*, with the difference between *Associated Social Traits* and *Associated Intellectual Traits* and the difference between *Necessary Social Traits* and

Necessary Intellectual Traits both being statistically significant*:* t(163) = -4.499, p <= .001; t(163) = 10.6461, p <= .001; respectively.

5. Discussion

5.1. Research Findings and Main Conclusions

The intention behind this study was to assess how humans' PS at work can be impacted differently by a human against an AI. Therefore, four different effects were investigated. I will start by discussing the effects of having an AI versus a human joining a person's workplace, followed by having someone at a leadership position versus a peer relationship towards that same person. Afterwards, the joint effect of the interaction between both previous cases will be discussed. And finally, I will discuss the effects of perception levels by investigating the impact of associated traits on PS and how the type of entity can have an indirect impact on PS, resulting from the impact mentioned.

5.1.1. AI vs Human on Psychological Safety

Hypothesis 1 predicted that having an AI as an addition to our work environment compared to another human being would impact negatively a person's PS. As concluded in the "Hypothesis Testing" section where the results for this hypothesis are provided, this study's findings do not support the hypothesis. This result was not expected. With AI being a recent technology and seen by some people as a threat to humanity not only in survival terms, but also related to job loss, we expected people to feel safer working alongside humans compared to this technology. Not only that, but the fact of being from the same species, would supposedly make a distinct difference in the levels of PS. This study demonstrates that all these assumptions are not true and it can be the result from people becoming more open-minded towards this scenario and less subject to the fear of the unknown with these entities. To strengthen this point, there were some participants in this study that shared their opinion on this matter at the last step of the survey (Check "Procedure" Section). Some examples: "I would be very curious and open to work with them."; "At first it would be uncomfortable since is not a human being. Although I think it would be very interesting in the long run, since this new member could do the most mechanical work and perhaps some analytical processes helping a lot the 'normal' members.".

5.1.2. Leader vs Colleague on Psychological Safety

For the second hypothesis regarding the position of the entity added to the company, although the hypothesis was not supported by the results, having no significant difference between having a new addition of an individual as a leader or a peer towards a person's PS, there stills seems to be a significant impact of a Leadership position towards PS levels, individually. When tending to a person who is superior to us from a hierarchical point of view, people tend to feel more exposed to interpersonal risks, however, this aspect is not correlated with having someone at the same hierarchical stage, as towards peers, we tend to feel indifferent. This makes sense as people at work tend to feel more at risk when facing a person that has relative power over us at the company than someone who has not.

5.1.3. Interaction between both Independent Variables (Entity x Position)

Looking at our third hypothesis, resulting from the combination of the previous two, the results tend to be the same, where the hypothesis is not supported. This finding demonstrates that having AI at a leadership position does not affect differently a person's PS levels compared to the other three possible scenarios. This can be the result of a peer position and AI itself not having a direct impact on the levels of PS, demonstrating that all four scenarios seem to have the exact same kind of impact towards a person's psychological safety.

5.1.4. Intellectual & Social Traits

Unlike all other hypotheses, the fourth and final one, was supported by the results achieved, especially hypothesis *4a*, where it was concluded that having more social traits associated with the entity impacts positively the levels of PS. For the case of hypothesis *4b*, although the hypothesis was not fully supported, we concluded that instead of having a positive impact on psychological safety, having more intellectual traits impacts it negatively, which was unexpected. Being PS associated with a person feeling at ease and not exposed to possible interpersonal risks at work, humans tend to feel safer towards these risks when in an environment in which they feel welcomed. Therefore, by being associated with more positive social traits, participants felt more psychological safe when considering a scenario with a more social entity than with a less social one.

The difference in terms of how certain types of associated traits can affect the levels of PS, can be explained by how people perceive themselves against other individuals and how this affects someone in specific to their PS and state of mind. Nowadays, especially with the rise of social

media, we currently live in a society that is focused on constant comparison with other people because of the ideal image that is shown to us (Fardouly & Vartanian, 2016, cit. in Monks et al., 2020). Although, this tends to be focused on appearance and not on psychological terms, we still strive to become the best and we tend to compare ourselves with others in every aspect. Intellectual traits are a huge example of this type of scenario, where we belittle ourselves when comparing to others that are better than us. This effect is called social comparison and has been shown to mediate depression (Swallow & Kuiper 1988) and can be commonly found to happen within the work group which negatively impacts levels of PS (Edmondson, 1999; Nembhard & Edmondson, 2006). These lower levels of PS towards intellectual traits can be also result from the possibility of us feeling at risk if being evaluated worse at our jobs due to comparisons with someone that is considered to be intellectually superior to us.

5.1.5. Exploratory Analysis

Regarding the exploratory analysis, what I wanted to observe was also achieved with the results obtained from the ANOVA and paired sample t-tests. Not only we were able to observe similar effects regarding the Associated Social and Intellectual Traits included in our regression model, I was also able to achieve results that support that there is a significant difference between Associated and Necessary Traits, in general, but also between *Social* and *Intellectual Traits* within the necessary ones, when facing different types of *Entity* and *Position*.

As observed in the section regarding the results from the exploratory analysis, both the individual main effects of the variables *Identification* and *Type of Traits* are statistically significant. Meaning, that there is a differentiation between Associated and Necessary Traits, and the same for Social and Intellectual, the latter explaining why the type of entity might have different impacts for each of these. The difference of having an AI or Human, at our company, in terms of the personality traits we associate them with can once again be seen when we observe its interaction with the variable *Entity*, proving the same conclusion from the regression model: when facing either AI or human, there is a significant difference on how the type of traits are affected.

When looking deeper into the fact that whether people tend to deem the traits necessary to an individual comparing to what they tend to associate with the person, we find something quite interesting. Looking into both repeated measures variables' individual interaction with the type of

Position and the type of *Entity* introduced to the company, we can observe that while with the *Type of Trait*, the *Position* of the entity towards the individual at the company is not statistically significant, the same cannot be said regarding the type of Entity, that is statistically significant. This explains a lot, as we tend to speak differently about the *Type of Traits* identified into an individual when facing different types of entities, independently of the position that they work in the company. Therefore, the way we perceive them to be depends totally on whether it's a human or AI and not on the position, as already concluded with the main model of our study.

On the other hand, the same cannot be said about the variable *Identification*, as there is no main effect regarding the interaction with the type of *Entity*, being irrelevant the fact that we are facing an AI or human, but there is actually a marginally significant interaction between this variable and *Position*, which further defends the idea and results from Furnham (2002), that tells us there is a difference on how we perceive our peers and our leaders and the kind of personality we expect them to have and associate with.

Despite this, there are still some effects that are interesting and should be further explored, and that is the fact that when observing the interaction of both variables *Identification* and *Type of Traits*, its interaction effect is statistically significant, meaning that there is an interaction effect between both these variables. The same can also be said when adding this interaction with the type of *Entity*. Being statistically significant it adds to the finding that the type of Entity does have somewhat of an impact on how these variables combine, strengthening the conclusions from the model.

Another peculiar thing regarding this exploratory analysis that should be further explored is when we look at the plot of the estimated marginal means. Looking individually at each Type of Trait, we can observe that for the case of Social Traits we tend to associate less of these positive type of traits to an individual, although we tend to deem them more necessary, while Intellectual Traits follow the other way around. This is such an interesting fact since why would we tend to consider someone to possess a smaller number of traits when we consider these to be so important to us, or even the other way around? Although this was not a topic that was mentioning throughout this paper as it was not something considered as a variable for this study, Emotional Intelligence could be a huge factor towards this effect as it is related to how we feel towards ourselves and others Salovey & Mayer (1990).

5.2. Academic and Managerial Relevance

Overall, this dissertation provides relevant results and findings both to academic but also managerial contexts. While there has been a lot of studies on the workplace environment at companies and how people feel within those considering certain changes (such as the Hawthorne Effect; Sedgwick & Greenwood, 2015), and a lot of studies focusing on AI and its development, not many have been focused on the effects of a workplace from a potential future where there is involvement of AI amongst humans. Concentrating specifically on how the levels of PS are impacted by AI either in a position of leadership or in a similar position to the individual considered, this study can be beneficial for business leaders that are looking into the future of work environments and how to adapt to them and researchers not only of AI and PS, but also researchers on the effects of changes at the workplace to the employees residing in it.

The present research findings defend that AI has no significant impact on PS when included in a person's workplace compared to another human being. Not only that, but it also defends that no matter what type of individual is added to the workplace, being AI, human, the type of position it joins the company does not significantly impact PS, although, when an individual joins at a leadership position it has a direct effect on PS, unlike joining as a colleague, that has no effect at all. This is interesting and important, as it allows companies that wish to become a hybrid type of company where humans work alongside AI to better reach that goal. Companies should be mindful that the reaction to AI might depend on its hierarchical position. AI should be considered as a key component for the future of companies as they tend to be associated with higher productivities and capabilities in intellectual terms, and has no difference on the impact provoked to current employees' psychological safety levels, allowing to create new compositions within the company and perhaps achieve better results.

Furthermore, the present findings also suggest that the traits that us humans tend to perceive and associate with a given entity impact PS as a whole. These traits, in turn, are impacted by the type of entity we are facing (AI or human). When comparing humans to AI, humans tend to be considered more social entities while AI are considered to be more intellectual, and this is an important aspect when measuring the levels of PS due to the individual impact of each kind of trait towards PS. In this study, I found that beings with higher number of social traits impact positively the levels of PS, while beings with higher number of intellectual traits have the opposite effect. This key finding tells us that companies that want to achieve hybrid

environments with humans working alongside AI, should invest in the development and promotion of machines that are able to provide positive social environments at their workplace to promote safety to current employees and, in turn, result in higher productivity, as the exposure to interpersonal risks diminishes.

5.3. Limitations and Future Research

Although this dissertation offers valuable insight for future research in terms of a futuristic environment on how PS levels will be affected by the rise of AI in companies' daily work, this study still has its own limitations and aspects that could be further analysed.

One of the main limitations is related to the scale used throughout the survey built by Amy Edmondson (1999). Despite it having great scale reliability in other studies (e.g., Moreira, 2020) and being considered a good scale to measure the levels of PS as shown before, when measuring the scale reliability for this study, it failed to reach the desired level of .8, which although it is not as detrimental as it could be, since a level of at least .6 was achieved in this study for the scale in general, it is still not optimal to apply this type of scale for future research.

A key component that could also be added to future research, would be the concept of emotional intelligence. Emotional Intelligence has been researched to have connections and a certain degree of impact on the levels of PS (Harper & White, 2013) and could also be a way to explain how the different type of traits identified in this study could affect differently the levels of PS. Meaning, the levels of emotional intelligence could explain how differently people tend to associate different types of traits with both humans and AIs and how impactful it could be in this decision and in the way on how they perceive these entities. A good recommendation to evaluate the levels of emotional Intelligence could be the Mayer-Salovey-Caruso Emotional Intelligence Test or MSCEIT (Mayer et al., 2003; Brackett & Salovey, 2006). Furthermore, by considering the component of emotional intelligence we should also take into account how we perceive these machines to be, as we tend to associate them as "not" emotionally intelligent beings, explaining the effects obtained in this study.

Another concept that could be focused on this type of study would be to apply the concept of fear as a possible mediator. Being associated a lot with AI and the futuristic environment involving them, it could a potential mediator between both the Entity variable described in this study and how differently both can affect the levels of PS. Fear is also a concept that is really associated

nowadays with the workplace (as mentioned in the Literature Review), where some people tend to fear losing their job, especially during the current times we are facing of the pandemic. The finding that AI agents are perceived as possessing more intellectual traits than humans, can lead the fear of job loss by having people perceive themselves as the underdog compared to AI, being afraid of eventually being substituted by this technology.

It would also be interesting to further explore the variable Position in this study, by including one more possible category for this case: having a person join our workplace in a position that is lower than us in the hierarchy of the company, that is, as a subordinate, similar to the idea behind Furnham's research where they found significant differences especially between boss and subordinate (2002). Although, at first glance, we could hypothesize for it to be non-significant in terms on how it might affect the level of PS of the individual in question, especially when comparing both AI and Human, we could still be impressed by it having some impact on these kinds of scenarios. We could observe and hypothesize that maybe it would have a different impact and be statistically significant unlike the colleague position investigated in this study. Further exploration of this concept could be done by attributing different types of personalities to machines at different levels, as there has been research that suggested that a machine capable of agreeableness could help reduce interpersonal distance between humans and robots (Takayama & Pantofaru, 2009), while in other research, extroverted participants preferred to talk with an extroverted computer (Nass & Lee, 2001). This could also be further explored if added a mediator "trust" as it is key for positive work relationships (Pratt & Dirks, 2007).

Last, but not least, being AI already a force existing in modern days and becoming a part of companies slowly, there are different levels to their autonomy, and this could have different degrees on how this might end up affecting the levels of PS of individuals at their workplace.

6. Conclusion

This dissertation aimed to identify the impact of a futuristic environment on the levels of PS of people that currently work in a company's environment. In a long-term perspective, it is only a matter of time until technology has developed to reach the stage this study takes into consideration. With this dissertation, I was able to identify and conclude that, despite not having direct contribution to how psychological safe people will feel in these scenarios, AI still has an influence on how much exposure people feel towards interpersonal risks at the company.

Resulting from the simple fact that, despite similarities and advances, we still tend to perceive humans differently to machines. This perception then becomes the major cause on why people will eventually feel unsafe in their workplace when this scenario becomes a reality.

7. References

Baer, M., & Frese, M. (2002). Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior*, *24*(1), 45–68. https://doi.org/10.1002/job.179

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278. https://doi.org/10.1037/1528-3542.6.2.269

Bornemisza, A. (2013). *Creativity and Psychological Safety*. Lund University, DEPARTMENT OF PSYCHOLOGY.

Brackett, Marc A., & Salovey, Peter (2006). Measuring emotional intelligence with the Mayer-Salovery-Caruso Emotional Intelligence Test (MSCEIT). Psicothema, 18( ),34-41.[fecha de Consulta 8 de Julio de 2021]. ISSN: 0214-9915. Available at: https://www.redalyc.org/articulo.oa?id=72709505

Bunderson, J. S., & Boumgarden, P. (2010). Structure and Learning in Self-Managed Teams: Why "Bureaucratic" Teams Can Be Better Learners. *Organization Science*, *21*(3), 609–624. https://doi.org/10.1287/orsc.1090.0483

Carmeli, A. (2007). Social Capital, Psychological Safety and Learning Behaviours from Failure in Organisations. *Long Range Planning*, *40*(1), 30–44. https://doi.org/10.1016/j.lrp.2006.12.002

Carmeli, A., Brueller, D., & Dutton, J. E. (2009). Learning behaviours in the workplace: The role of high-quality interpersonal relationships and psychological safety. *Systems Research and Behavioral Science*, *26*(1), 81–98. https://doi.org/10.1002/sres.932

Carmeli, A., & Gittell, J. H. (2009). High-quality relationships, psychological safety, and learning from failures in work organizations. *Journal of Organizational Behavior*, *30*(6), 709–729. https://doi.org/10.1002/job.565

Carmeli, A., Reiter-Palmon, R., & Ziv, E. (2010). Inclusive Leadership and Employee Involvement in Creative Tasks in the Workplace: The Mediating Role of Psychological Safety. *Creativity Research Journal*, *22*(3), 250–260. https://doi.org/10.1080/10400419.2010.504654

Chabot, C., Shoveller, J. A., Spencer, G., & Johnson, J. L. (2012). Ethical and Epistemological Insights: A Case Study of Participatory Action Research with Young People. *Journal of Empirical Research on Human Research Ethics*, *7*(2), 20–33. https://doi.org/10.1525/jer.2012.7.2.20

Collins, C. J., & Smith, K. G. (2006). Knowledge Exchange and Combination: The Role of Human Resource Practices in the Performance of High-Technology Firms. *Academy of Management Journal*, *49*(3), 544–560. https://doi.org/10.5465/amj.2006.21794671

Collins, T., & Prigg, M. (2017, August 1). Facebook shuts down chatbots after they make own language. *Mail Online*. https://www.dailymail.co.uk/sciencetech/article-4747914/Facebook-shuts-chatbots-make-language.html

Cook, T., & Emler, N. (1999). Bottom-up versus top-down evaluations of candidates' managerial potential: An experimental study. *Journal of Occupational and Organizational Psychology*, *72*(4), 423–439. https://doi.org/10.1348/096317999166770

Detert, J. R., & Burris, E. R. (2007). Leadership Behavior and Employee Voice: Is the Door Really Open? *Academy of Management Journal*, *50*(4), 869–884. https://doi.org/10.5465/amj.2007.26279183

Detert, J. R., & Edmondson, A. C. (2011). Implicit Voice Theories: Taken-for-Granted Rules of Self-Censorship at Work. *Academy of Management Journal*, *54*(3), 461–488. https://doi.org/10.5465/amj.2011.61967925

Dutton, J., & Heaphy, E. 2003. The power of high quality connections. In K. Cameron, J. Dutton, & R. E. Quinn (Eds.), Positive organizational scholarship: Foundations for a new discipline: 263-278. San Francisco: Berrett- Koehler.

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, *44*(2), 350. https://doi.org/10.2307/2666999

Edmondson, A. C. (2004). Psychological safety, trust, and learning in organizations: A group-level lens. In R. M. Kramer & K. S. Cook (Eds.), *Trust and distrust in organizations: Dilemmas and approaches (pp. 239 –272)*. New York, NY: Sage

Fardouly, J., & Vartanian, L. R. (2016). Social Media and Body Image Concerns: Current Research and Future Directions. *Current Opinion in Psychology*, *9*, 1–5. https://doi.org/10.1016/j.copsyc.2015.09.005

Ferris, G. R., Liden, R. C., Munyon, T. P., Summers, J. K., Basik, K. J., & Buckley, M. R. (2009). Relationships at Work: Toward a Multidimensional Conceptualization of Dyadic Work Relationships. *Journal of Management*, *35*(6), 1379–1403. https://doi.org/10.1177/0149206309344741

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2016). Psychological Safety: A Meta-Analytic Review and Extension. *Personnel Psychology*, *70*(1), 113–165. https://doi.org/10.1111/peps.12183

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Furnham, A. (2002). Rating a boss, a colleague and a subordinate. *Journal of Managerial Psychology*, *17*(8), 655–671. https://doi.org/10.1108/02683940210450475

Furnham, A., & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology*, *67*(1), 57–67. https://doi.org/10.1111/j.2044-8325.1994.tb00549.x

Furnham, A., Forde, L., & Ferrari, K. (1999). Personality and work motivation. *Personality and Individual Differences*, *26*(6), 1035–1043. https://doi.org/10.1016/s0191-8869(98)00202-5

Furnham, A., & Stringfield, P. (1998). Congruence in Job-Performance Ratings: A Study of 360° Feedback Examining Self, Manager, Peers, and Consultant Ratings. *Human Relations*, *51*(4), 517–530. https://doi.org/10.1177/001872679805100404

Gliem, J. and Gliem, R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Paper presented at the 2003 Midwest Research to Practice Conference in Adult, Continuing and Community Education, the Ohio State University, Columbus, Ohio*. Retrieved from

https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf?sequence=1.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, *61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Harper, S. R., & White, C. D. (2013). The Impact of Member Emotional Intelligence on Psychological Safety in Work Teams. *Journal ofr Behavioral and Applied Management*. Published. https://doi.org/10.21818/001c.17933

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837–852. https://doi.org/10.1037/0022-3514.78.5.837

Kahn, W. A. (1990). Psychological Conditions of Personal Engagement and Disengagement at Work. *Academy of Management Journal*, *33*(4), 692–724. https://doi.org/10.5465/256287

Langer, B., & de Certeau, M. (1988). The Practice of Everyday Life. *Contemporary Sociology*, *17*(1), 122. https://doi.org/10.2307/2069486

Lee, M. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society, 5(1), 1-16. doi: 10.1177/2053951718756684

Liang, J., Farh, C. I. C., & Farh, J. L. (2012). Psychological Antecedents of Promotive and Prohibitive Voice: A Two-Wave Examination. *Academy of Management Journal*, *55*(1), 71–92. https://doi.org/10.5465/amj.2010.0176

Malhotra, N. K., Nunan, D., & Birks, D. F. (2017). Marketing research: An Applied Approach (Fifth Edit). Harlow: Pearson Education Limited. Retrieved from https://www.pearson.com

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, *3*(1), 97–105. https://doi.org/10.1037/1528-3542.3.1.97

McCarty, G. J. (1967). Personal and Organizational Change through Group Methods: The Laboratory Approach by Edgar H. Schein and Warren G. Bennis. *International Journal of Group Psychotherapy*, *17*(1), 105–106. https://doi.org/10.1080/00207284.1967.11642993

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Monks, H., Costello, L., Dare, J., & Reid Boyd, E. (2020). "We're Continually Comparing Ourselves to Something": Navigating Body Image, Media, and Social Media Ideals at the Nexus of Appearance, Health, and Wellness. *Sex Roles*, *84*(3–4), 221–237. https://doi.org/10.1007/s11199-020-01162-w

Moreira, P. (2020, June). *Artificial Intelligence Leadership: How Trust and Fairness Perceptions impact Turnover Intentions through Psychological Safety*.

Morrison, E. W., & Milliken, F. J. (2003). Speaking Up, Remaining Silent: The Dynamics of Voice and Silence in Organizations. *Journal of Management Studies*, *40*(6), 1353–1358. https://doi.org/10.1111/1467-6486.00383

Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, *72*, 432–440. https://doi.org/10.1016/j.chb.2017.02.067

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171–181. https://doi.org/10.1037/1076-898x.7.3.171

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior*, *27*(7), 941–966. https://doi.org/10.1002/job.413

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315–324. https://doi.org/10.1016/j.jesp.2009.12.002

Pratt, M. G., & Dirks, K. T. 2007. Rebuilding trust and restoring positive relationships: A commitment based view of trust. In J. E. Dutton, & B. R. Ragins (Eds.), Exploring positive relationships at work: 117-136. Mahwah, NJ: Lawrence Erlbaum.

Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally Significant Effects as Evidence for Hypotheses. *Psychological Science*, *27*(7), 1036–1042. https://doi.org/10.1177/0956797616645672

Roberts, B. W., & Hogan, R. (2001). *Personality Psychology in the Workplace (Decade of Behavior)* (1st ed.). American Psychological Association (APA).

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283–294. https://doi.org/10.1037/h0026086

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology*, *44*(4), 1100–1105. https://doi.org/10.1016/j.jesp.2007.12.001

Ryan, K. D., & Oestreich, D. K. (1998). *Driving Fear Out of the Workplace: Creating the High-Trust, High-Performance Organization* (2nd ed.). Jossey-Bass.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. Imagination, Cognition, and Personality, 9, 185–211.

Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y. C., & Hancock, P. A. (2011). A Model of Human-Robot Trust: Theoretical Model Development. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1432–1436. https://doi.org/10.1177/1071181311551298

Sarder TV [Russel Sarder]. (2015, June 22). *What is interpersonal risk? by Amy Edmondson, Author of Teaming* [Video]. YouTube. Available at: https://www.youtube.com/watch?v=bC3Qdhmg_9Y&ab_channel=RussellSarder

Schaubroeck, J., Lam, S. S. K., & Peng, A. C. (2011). Cognition-based and affect-based trust as mediators of leader behavior influences on team performance. *Journal of Applied Psychology*, *96*(4), 863–871. https://doi.org/10.1037/a0022625

Schein EH. (1993). How can organizations learn faster? The challenge of entering the green room. *Sloan Management Review*, 34, 85–92.

Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. *BMJ*, h4672. chttps://doi.org/10.1136/bmj.h4672

Sharan, N. N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon*, *6*(8), e04572. https://doi.org/10.1016/j.heliyon.2020.e04572

Siemsen, E., Roth, A. V., Balasubramanian, S., & Anand, G. (2009). The Influence of Psychological Safety and Confidence in Knowledge on Employee Knowledge Sharing. *Manufacturing & Service Operations Management*, *11*(3), 429–447. https://doi.org/10.1287/msom.1080.0233

Simplilearn. (2020, May 12). How AI Will Impact The Future | Rise of AI (Elon Musk,Bill Gates,Sundar Pichai,Jack Ma) |Simplilearn [Video]. YouTube. https://www.youtube.com/watch?v=uz8PSSOB-4E&ab_channel=Simplilearn

Skinner, J. (2004). Temporal strategies and fear in the workplace. *Journal of Sociology*, *40*(4), 417–431. https://doi.org/10.1177/1440783304048384

Spencer, G. E. (2000). Review : Children's competency to consent: an ethical dilemma. *Journal of Child Health Care*, *4*(3), 117–122. https://doi.org/10.1177/136749350000400305

Swallow, S. R., & Kuiper, N. A. (1988). Social comparison and negative self-evaluations: An application to depression. *Clinical Psychology Review*, *8*(1), 55–76. https://doi.org/10.1016/0272-7358(88)90049-9

Takayama, L., & Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Published. https://doi.org/10.1109/iros.2009.5354145

Temin, P. (1997). Two Views of the British Industrial Revolution. *The Journal of Economic History*, *57*(1), 63–82. https://doi.org/10.1017/s0022050700017927

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating Faces on Trustworthiness After Minimal Time Exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Tucker, A. L., Nembhard, I. M., & Edmondson, A. C. (2007). Implementing New Practices: An Empirical Study of Organizational Learning in Hospital Intensive Care Units. *Management Science*, *53*(6), 894–907. https://doi.org/10.1287/mnsc.1060.0692

Van Voorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. Tutorials in Quantitative Methods for Psychology, 3 (2), 43-50. https://doi.org/10.20982/tqmp.03.2.p043

Willis, J., & Todorov, A. (2006). First Impressions. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wiseman, Y. (2021). Autonomous Vehicles. *Encyclopedia of Information Science and Technology, Fifth Edition*, 1–11. https://doi.org/10.4018/978-1-7998-3479-3.ch001

Zebrowitz, L. A. (1996). Physical appearance as a basis for stereotyping. In M. H. N. McRae & C. Stangor (Eds.), Foundation of Stereotypes and Stereotyping (pp. 79–120). New York, NY: Guilford Press.

Zhang, B., & Dafoe, A. (2019). Artificial Intelligence: American Attitudes and Trends. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.3312874

8. Appendices

Appendix 1: Survey

Introduction

Welcome and thank you for participating in this experiment on psychological safety. I am conducting this experiment as part of my Master Thesis. The study consists of answering to multiples questions with scales, where you will base your answer regarding a specific scenario. The purpose is to gain insight into how individuals react to certain scenarios and if it influences how psychologically safe they feel. It will take about 5 minutes to complete. Please answer as honestly as possible. All answers will be kept strictly confidentially and are anonymous. This means that there will be no way to link your responses to your identity. The data collected will be used for research purposes only. Your participation will contribute to research on psychological safety. If you have any questions about this study, please email Tomás Bandeira (152119211@alunos.lisboa.ucp.pt). By continuing you agree to participate. Thank you!

First Scenario

First Scenario

Q1: Imagine yourself in each of the following scenarios at your current work environment (or at your last one; in case you are not working at the moment), we would like you to answer on how much you agree with each statement:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
If you make a mistake in the company, it is often held against you. (1)	○	○	○	○	○	○	○
Members of the company are able to bring up problems and tough issues. (2)	○	○	○	○	○	○	○
It is safe to take a risk on this company. (3)	○	○	○	○	○	○	○
My unique skills and talents are valued and utilized when working for the company. (4)	○	○	○	○	○	○	○
Please choose the "Somewhat disagree" option, so I know you are paying attention. (5)	○	○	○	○	○	○	○
People on this team	○	○	○	○	○	○	○

sometimes reject others for being different. (6)							
No one on this team would deliberately act in a way that undermines my efforts. (7)	○	○	○	○	○	○	○
It is difficult to ask other members of this team for help. (8)	○	○	○	○	○	○	○

Second Scenario

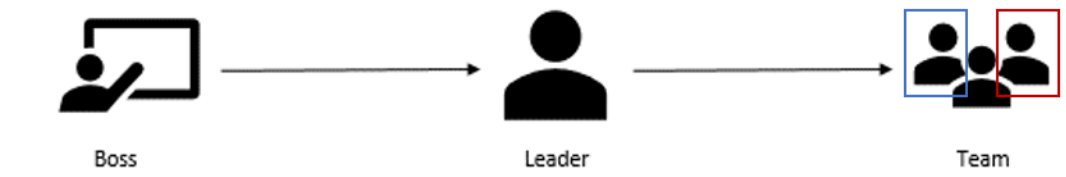
Note: All scenes differ only on the description of the scenario and the image shown, therefore, all questions associated were only listed once in this appendix.

Scene 1: Human Colleague

Now imagine that one day, you were told by your team leader (boss) that a colleague of yours is resigning and they would be adding a new person to your team, starting tomorrow.

To provide you with a better scope of the relationship built in the company, down below you can observe a pictogram describing the relationship with this new individual. Notice that you belong in the section team (highlighted by the blue box), while the individual added is represented by the figure highlight by the red box.

Image 1:

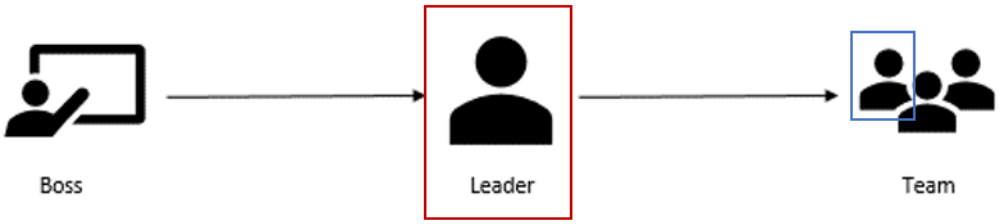


Scene 2: Human Leader

Now imagine that one day, you were told by your team leader (boss) that they will be resigning and they would be adding a new person as your boss, starting tomorrow.

To provide you with a better scope of the relationship built in the company, down below you can observe a pictogram describing the relationship with this new individual. Notice that you belong in the section team (highlighted by the blue box), while the individual added is represented by the figure highlight by the red box.

Image 2:



Scene 3: AI Colleague

Now imagine that one day, you were told by your team leader (boss) that a colleague of yours is resigning and they would be adding an Artificial Intelligence* Robot (commonly known as android) to your team, starting tomorrow.

To provide you with a better scope of the relationship built in the company, down below you can observe a pictogram describing the relationship with this new individual. Notice that you belong in the section team (highlighted by the blue box), while the individual added is represented by the figure highlight by the red box.

*Artificial intelligence (AI) refers to the simulation of human intelligence on machines that are programmed to think and behave like humans. This term can also be applied to any machine that demonstrates traits associated with the human mind such as, learning and problem solving.In other words, an Artificial Intelligence system is a computer program that has the ability to think and learn intelligently, just like humans. Usually these agents are known as computers or intelligent robots. More specifically, an artificial intelligence colleague is a robot that is constantly learning, adapting and developing, just like a human being once it has intelligence. This intelligent robot is programmed to be able to guide, evaluate, structure and divide work, make tasks easier, provide feedback and improve interactions within a group or company.

Image 3

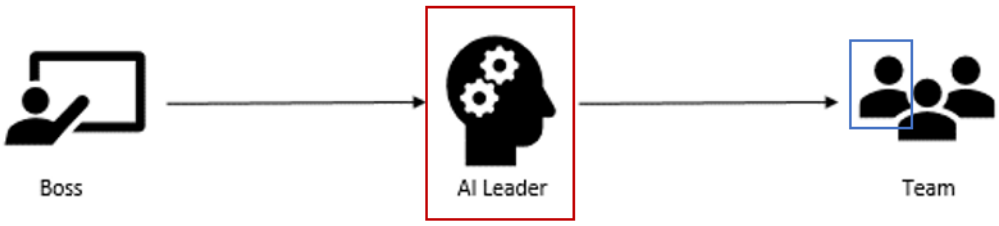
Scene 4: AI Leader

Now imagine that one day, you were told by your team leader (boss) that they are resigning and they would be adding an Artificial Intelligence* Robot (commonly known as android) as your boss, starting tomorrow.

To provide you with a better scope of the relationship built in the company, down below you can observe a pictogram describing the relationship with this new individual. Notice that you belong in the section team (highlighted by the blue box), while the individual added is represented by the figure highlight by the red box.

*Artificial intelligence (AI) refers to the simulation of human intelligence on machines that are programmed to think and behave like humans. This term can also be applied to any machine that demonstrates traits associated with the human mind such as, learning and problem solving.In other words, an Artificial Intelligence system is a computer program that has the ability to think and learn intelligently, just like humans. Usually these agents are known as computers or intelligent robots. More specifically, an artificial intelligence manager is a robot that is constantly learning, adapting and developing, just like a human being once it has intelligence. This intelligent robot is programmed to be able to guide, evaluate, structure and divide work, make tasks easier, provide feedback and improve interactions within a group or company.

Image 4



Q2: Once again, imagine yourself in each of the following scenarios, but this time imagining yourself under the condition mentioned in the beginning of this page. Answer to each statement carefully:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
If you make a mistake in the company, it is often held against you. (1)	○	○	○	○	○	○	○
Members of the company are able to bring up problems and tough issues. (2)	○	○	○	○	○	○	○
It is safe to take a risk on this company. (3)	○	○	○	○	○	○	○
My unique skills and talents are valued and utilized when working for the company. (4)	○	○	○	○	○	○	○
Please choose the "Somewhat disagree" option, so I know you are paying attention. (5)	○	○	○	○	○	○	○
People on this team sometimes reject others for being different. (6)	○	○	○	○	○	○	○

No one on this team would deliberately act in a way that undermines my efforts. (7)	○	○	○	○	○	○	○
It is difficult to ask other members of this team for help. (8)	○	○	○	○	○	○	○

Q3: When you imagine this specific new member, what kind of traits do you associate with this new member? (In this question you are free to select more than one option)

- ☐ Accepting (1)
- ☐ Adaptable (2)
- ☐ Attentive (3)
- ☐ Collaborative (4)
- ☐ Creative (5)
- ☐ Energetic (6)
- ☐ Goal-Oriented (7)
- ☐ Good Communicator (8)
- ☐ Hard-Working (9)
- ☐ Helpful (10)
- ☐ Honest (11)

- [ ] Humble (12)
- [ ] Inspirational (13)
- [ ] Kind-Hearted (14)
- [ ] Open to Feedback (15)
- [ ] Optimistic (16)
- [ ] Organized (17)
- [ ] Passionate (18)
- [ ] Problem-Solver (19)
- [ ] Professional (20)
- [ ] Punctual (21)
- [ ] Reliable (22)
- [ ] Resourceful (23)
- [ ] Respectful (24)
- [ ] Other (Please Specify) (25)

_____

Q4: What kind of traits would be necessary for you to be more comfortable around this new member? (In this question you are free to select more than one option)

- ☐ Accepting (1)
- ☐ Adaptable (2)
- ☐ Attentive (3)
- ☐ Collaborative (4)
- ☐ Creative (5)
- ☐ Energetic (6)
- ☐ Goal-Oriented (7)
- ☐ Good Communicator (8)
- ☐ Hard-Working (9)
- ☐ Helpful (10)
- ☐ Honest (11)
- ☐ Humble (12)
- ☐ Inspirational (13)
- ☐ Kind-Hearted (14)
- ☐ Open to Feedback (15)
- ☐ Optimistic (16)

- [ ] Organized (17)
- [ ] Passionate (18)
- [ ] Problem-Solver (19)
- [ ] Professional (20)
- [ ] Punctual (21)
- [ ] Reliable (22)
- [ ] Resourceful (23)
- [ ] Respectful (24)
- [ ] Other (Please Specify) (25) _____

Personal Information

To finish off, we would kindly ask you to answer to the following demographic and study questions.

Q5: How old are you?

_____

Q6: What is your gender?

- ○ Male (1)
- ○ Female (2)
- ○ Other (Please Specify) (3) _____

Q7: In which country do you currently reside?

▼ Afghanistan (1) ... Zimbabwe (1357)

Q8: Are you currently working, studying or both?

- ☐ Employed full time (1)
- ☐ Employed part time (2)
- ☐ Unemployed looking for work (3)
- ☐ Unemployed not looking for work (4)
- ☐ Retired (5)
- ☐ Student (6)

Q9: What is your highest level of education?

- ○ Less than high school (1)
- ○ High school graduate (2)
- ○ Some college credit, no degree (3)
- ○ Bachelor's Degree (4)
- ○ Master's Degree (5)
- ○ Professional Degree (6)
- ○ Doctorate/PhD (7)

Q10: In what sector do you currently work in?

- ☐ I'm just a Student (1)
- ☐ Agricultural (2)
- ☐ Automotive (3)
- ☐ Retail & Consumer Goods (4)
- ☐ Education (5)
- ☐ Financial Services (6)
- ☐ Health Services (7)
- ☐ Manufacturing (8)
- ☐ Public Service (9)
- ☐ Real Estate (10)

- [ ] Tourism (11)
- [ ] Telecommunications, Media, Technology (12)
- [ ] Utilities/Energy (Water; Gas; Eletricity) (13)
- [ ] Other (Please Specify) (14) _____

Q11: Lastly, does the addition of this new member make sense in the working environment you currently belong to?

- ○ Yes (1)
- ○ No (2)
- ○ Not Sure (3)

Q12: Do you consider this addition to be realistic?

- ○ Yes (1)
- ○ No (2)
- ○ Not sure (4)

Q13: If you wish to share with us in more detail, how you would feel working alongside an individual like this one, feel free to answer and explain in the box below (optional):

_____

Conclusion

Thank you for your participation in this study.

The purpose behind it, is to better understand, in a hypothetical future, if people would still be able to feel psychological safe, being comfortable to share opinions, feedback, among others, with the addition of an Artificial Intelligence in their work environment and if there is a main difference when comparing it with the addition of a human being instead.

To achieve this purpose, we separated the total number of participants in 4 different scenarios, where they would be facing either an addition of a new member to the team or a change of their team leader. This change/addition would be done with either a human being or an Artificial Intelligence.

Appendix 2: Study Overview

Study Sample Size

	Valid		Invalid		Total	
	N	%	N	%	N	%
Answers	220	90,91%	22	9,09%	242	100,00%

Scenarios Frequency

	n	%
Human Colleague	57	25,91%
Human Leader	53	24,09%
AI Colleague	54	24,55%
AI Leader	56	25,45%

Scales Used

Likert Scale	Psychological Safety Scene 1	Psychological Safety Scene 2
1	Strongly Disagree	Strongly Disagree
2	Disagree	Disagree
3	Somewhat Diasgree	Somewhat Diasgree
4	Neither Agree nor Disagree	Neither Agree nor Disagree
5	Somewhat Agree	Somewhat Agree
6	Agree	Agree
7	Strongly Agree	Strongly Agree

Note: Psychological Safety Scene 2 represents all scales used for each of the four different induced scenarios.

Variables in the Associated and Necessary Traits

#	Trait Variables
1	Accepting
2	Adaptable
3	Attentive
4	Collaborative
5	Creative
6	Energetic
7	Goal-Oriented
8	Good Communicator
9	Hard-Working
10	Helpful
11	Honest
12	Humble
13	Inspirational
14	Kind-Hearted
15	Open to Feedback
16	Optimistic
17	Organized
18	Passionate
19	Problem-Solver
20	Professional
21	Punctual
22	Reliable
23	Resourceful
24	Respectful
25	Other (Please Specify)

Note: All Trait Variables seen in this table are used throughout the survey when asked to select the traits we deemed necessary to a certain individual and the ones we associate with it for every scenario out of the four induced.

Appendix 3: Demographics

Samples Demographic Characteristics

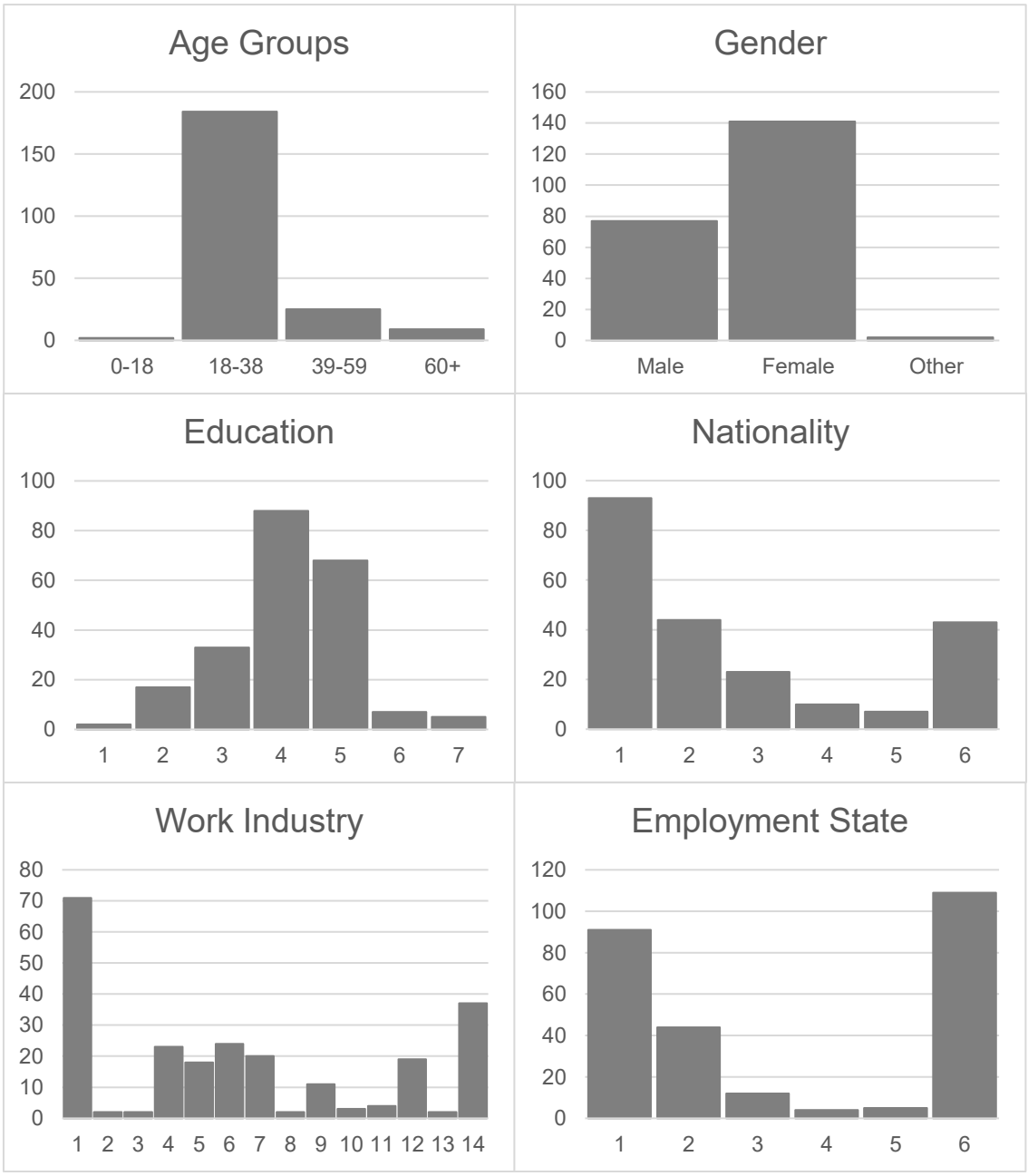
Characteristics	Total	
	N	%
Age		
0-18	2	0,91%
18-38	184	83,64%
39-59	25	11,36%
60+	9	4,09%

Gender		
Male	77	35,00%
Female	141	64,09%
Other	2	0,91%
Education		
Less than high school (1)	2	0,009091
High school graduate (2)	17	0,077273
Some college credit, no degree (3)	33	0,15
Bachelor's Degree (4)	88	0,4
Master's Degree (5)	68	0,309091
Professional Degree (6)	7	0,031818
Doctorate/PhD (7)	5	2,27%
Nationality		
Portugal (1)	93	42,27%
United Kingdom (2)	44	20,00%
United States of America (3)	23	10,45%
Germany (4)	10	4,55%
Netherlands (5)	7	3,18%
Others (6)	43	19,55%
Work Industry		
I'm just a Student (1)	71	29,83%
Agricultural (2)	2	0,84%
Automotive (3)	2	0,84%
Retail & Consumer Goods (4)	23	9,66%
Education (5)	18	7,56%
Financial Services (6)	24	10,08%
Health Services (7)	20	8,40%
Manufacturing (8)	2	0,84%
Public Service (9)	11	4,62%
Real Estate (10)	3	1,26%
Tourism (11)	4	1,68%
Telecommunications, Media, Technology (12)	19	7,98%
Utilities/Energy (Water; Gas; Eletricity) (13)	2	0,84%
Other (Please Specify) (14)	37	15,55%
Employment State		
Employed full time (1)	91	34,34%
Employed part time (2)	44	16,60%
Unemployed looking for work (3)	12	4,53%

Unemployed not looking for work (4)	4	1,51%
Retired (5)	5	1,89%
Student (6)	109	41,13%

Note: $N = 220$.

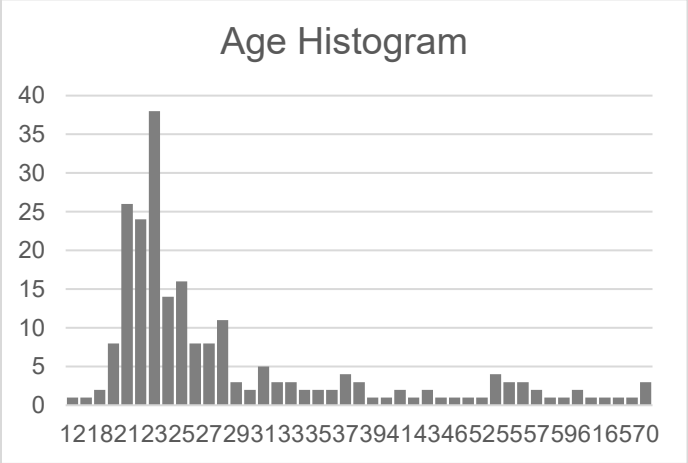
Barcharts:



Demographic Descriptive Statistics

	Valid		Missing		Total	
	N	%	N	%	N	%
Age	220	90,91%	22	9,09%	242	100%
Gender	220	90,91%	22	9,09%	242	100%
Nationality	220	90,91%	22	9,09%	242	100%
Education	220	90,91%	22	9,09%	242	100%

Note: Working Industry and Employment State not included as participants were able to answer to more than one question, therefore, despite only being correspondent to having 220 valid answers and a total of 22 missing answers, they were still able to get more than 220 valid outcomes.



Other Descriptive Statistics

Addition	Yes		No		Not Sure		Total	
	N	%	N	%	N	%	N	%
Is Realistic	115	52,27%	61	27,73%	44	20,00%	220	100%
Makes Sense	106	48,18%	67	30,45%	47	21,36%	220	100%

Descriptive Statistics for both measured PS

							Skewness		Kurtosis	
	N	Mean	Median	SD	Min	Max	Statitstic	Std. Error	Statitstic	Std. Error
Initial Psychological Safety	225	4,6495	4,71	0,9497	1,71	7	-0,0654	0,1622	-0,2587	0,3231
Final Psychological Safety	220	4,4584	4,43	0,928	1,57	7	0,0025	0,164	0,1077	0,3266

Independent sample t-test Gender Binary and level of PS

	Male			Female			t-test		Levene's test	
	n	M	SD	n	M	SD	t(216)	p	F	p
Initial PS	77	4,52	0,9690	141	4,71	0,9257	-1,489	0,138	0	0,998
Final PS	77	4,39	0,9828	141	4,50	0,8957	-0,805	0,421	0,466	0,495

Note: Participants who mentioned having a non-binary gender, were not taken into account in the analysis above. These results suggest that women, on average, tend to possess higher levels of PS than men.

Participants average level of PS for each component of the PS scale

	Initial Psychological Safety		Final Psychological Safety	
	M	SD	M	SD
(1) If you make a mistake in the company, it is often held against you.	4,25	1,6139	5,20	1,5597
(2) Members of the company are able to bring up problems and tough issues.	4,78	1,4771	4,56	1,5786
(3) It is safe to take a risk on this company.	4,21	1,4223	4,13	1,5332
(4) My unique skills and talents are valued and utilized when working for the company.	4,71	1,6555	4,60	1,6161
(5) People on this team sometimes reject others for being different.	3,33	0,8743	4,72	1,6849
(6) No one on this team would deliberately act in a way that undermines my efforts.	4,77	1,7300	4,47	1,6498
(7) It is difficult to ask other members of this team for help.	4,64	1,6187	4,75	1,6029

Descriptive Statistics for Traits

	Traits	N	Mean	Median	SD	Min	Max	Skewness		Kurtosis	
								Statitstic	Std. Error	Statitstic	Std. Error
Associated	Social	184	3,61	2	2,6955	1	14	1,57	0,1791	2,72	0,3564
	Intelectual	212	3,80	3	2,2048	1	10	0,73	0,1671	0,00	0,3326
Necessary	Social	215	4,84	4	3,0422	1	14	1,02	0,1659	0,61	0,3303
	Intelectual	183	3,14	2	2,2778	1	10	1,35	0,1796	1,31	0,3573
Difference	Social	180	1,34	2	2,8993	-7	10	0,37	0,1811	0,72	0,3602
	Intelectual	178	-0,72	-1	2,5284	-8	7	-0,07	0,1821	1,00	0,3622

Appendix 4: Reliability Statistics

Initial Psychological Safety Cronbach's Alpha

Construct	# Items	Cronbach's Alpha	Items	If item Deleted
Initial PS	7	0,700	If you make a mistake in the company, it is often held against you.	0,690
			Members of the company are able to bring up problems and tough issues.	0,713
			It is safe to take a risk on this company.	0,657
			My unique skills and talents are valued and utilized when working for the company.	0,650
			People on this team sometimes reject others for being different.	0,644
			No one on this team would deliberately act in a way that undermines my efforts.	0,672
			It is difficult to ask other members of this team for help.	0,635

Final Psychological Safety Cronbach's Alpha

Construct	# Items	Cronbach's Alpha	Items	If item Deleted
Final PS	7	0,653	If you make a mistake in the company, it is often held against you.	0,624
			Members of the company are able to bring up problems and tough issues.	0,652
			It is safe to take a risk on this company.	0,597
			My unique skills and talents are valued and utilized when working for the company.	0,583
			People on this team sometimes reject others for being different.	0,655
			No one on this team would deliberately act in a way that undermines my efforts.	0,631
			It is difficult to ask other members of this team for help.	0,572

Note: This Cronbach's Alpha, representing the Final PS scale reliability, is the joint combination of all four randomly selected scenarios induced to the participants.

Appendix 5: Attention Checks

Attention Check

Scene	Passed		Failed		Total	
	n	%	n	%	n	%
Initial Scene	174	79,09%	46	20,91%	220	100,00%
Scene with Change:	180	81,82%	40	18,18%	220	100,00%
Human Colleague	47	21,36%	10	4,55%	57	25,91%
Human Leader	40	18,18%	13	5,91%	53	24,09%
AI Colleague	46	20,91%	8	3,64%	54	24,55%
AI Leader	47	21,36%	9	4,09%	56	25,45%
Both Scenes	165	75,00%	55	25,00%	220	100,00%

Note: Participants who failed at least one of the two attention checks were assigned to the failed group, summing to a total of 165 passed participants and 55 failed.

Psychological Safety Sample Mean and Standard Deviation

	Passed			Failed		
	n	M	SD	n	M	SD
Initial PS	174	4,697241	0,974671	46	4,426304	0,800911
Final PS	180	4,503389	0,963799	40	4,257	0,787941

Appendix 6: Psychological Safety Scale by Amy Edmondson 1999

Edmondson's Psychological Safety Scale (1999):

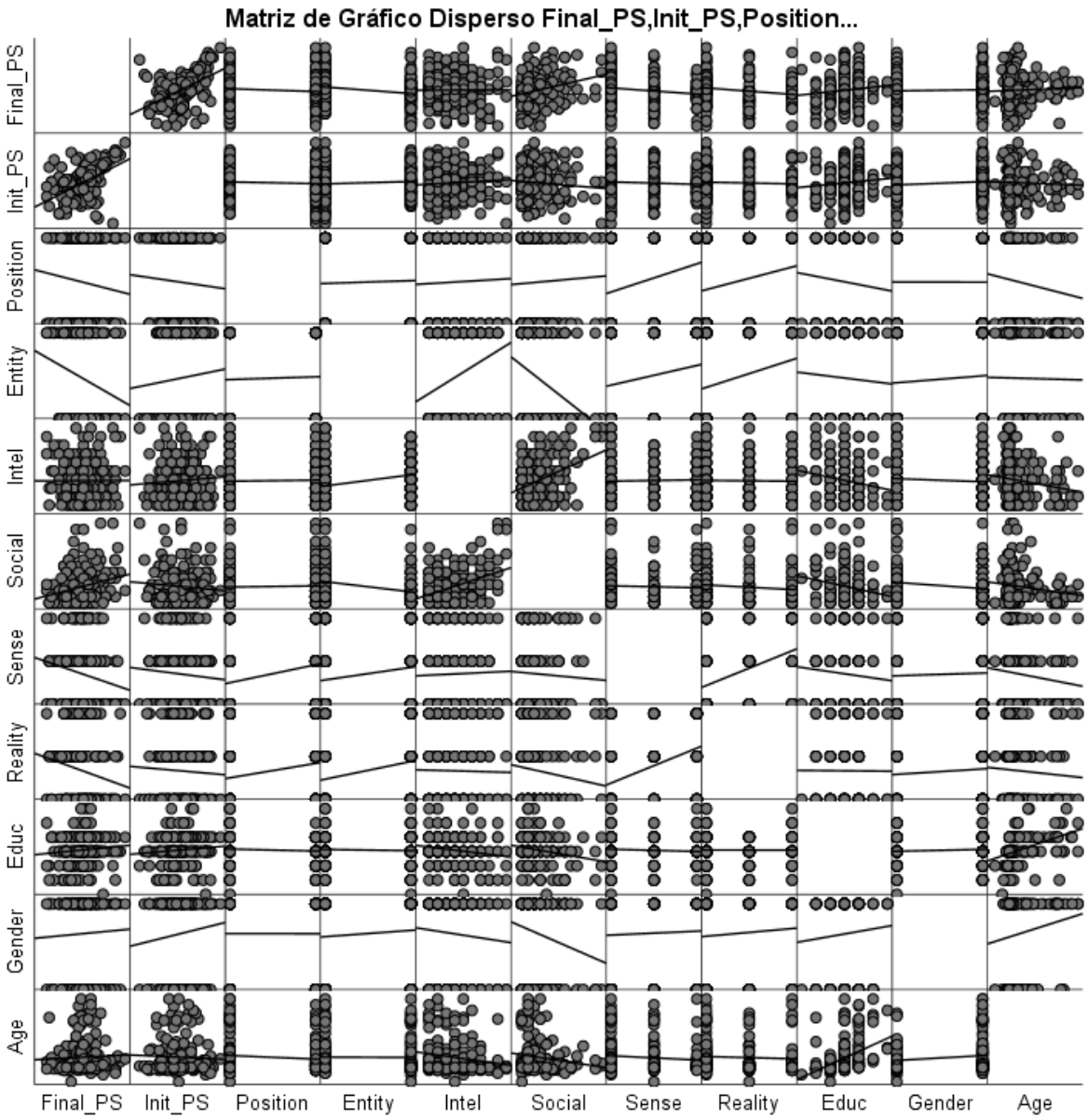
- If you make a mistake on this team, it is often held against you.
- Members of this team are able to bring up problems and tough issues.
- People on this team sometimes reject others for being different.
- It is safe to take a risk on this team.
- It is difficult to ask other members of this team for help.
- No one on this team would deliberately act in a way that undermines my efforts.
- Working with members of this team, my unique skills and talents are valued and utilized.

Appendix 7: Correlations

Spearman's Correlations of all Variables Included in the Model

Variables		1	2	3	4	5	6	7	8	9	10	11
1. Final PS	Correlation	1	,542**	-,136*	-,165*	-0,012	,184*	-,203**	-,247**	0,107	0,055	,137*
	N	220	220	220	220	212	184	220	220	220	220	220
2. Initial PS	Correlation	,542**	1	-0,086	0,061	0,110	-0,067	-0,066	-0,079	0,086	0,077	0,000
	N	220	225	225	225	212	184	220	220	220	220	220
3. Position	Correlation	-,136*	-0,086	1	0,108	0,063	0,018	,259**	,206**	-0,076	-0,047	-0,029
	N	220	225	242	242	212	184	220	220	220	220	220
4. Entity	Correlation	-,165*	0,061	0,108	1	,326**	-,363**	,156*	,239**	-0,039	0,084	-0,019
	N	220	225	242	242	212	184	220	220	220	220	220
5. Intelectual Traits	Correlation	-0,012	0,110	0,063	,326**	1	,201**	0,041	-0,034	-0,076	-0,014	-0,117
	N	212	212	212	212	212	178	212	212	212	212	212
6. Social Traits	Correlation	,184*	-0,067	0,018	-,363**	,201**	1	-0,131	-,210**	-0,129	-,194**	-0,085
	N	184	184	184	184	178	184	184	184	184	184	184
7. Change makes Sense	Correlation	-,203**	-0,066	,259**	,156*	0,041	-0,131	1	,446**	-0,088	0,051	-0,060
	N	220	220	220	220	212	184	220	220	220	220	220
8. Change is Realistic	Correlation	-,247**	-0,079	,206**	,239**	-0,034	-,210**	,446**	1	-0,030	0,069	-,133*
	N	220	220	220	220	212	184	220	220	220	220	220
9. Education	Correlation	0,107	0,086	-0,076	-0,039	-0,076	-0,129	-0,088	-0,030	1	0,110	,498**
	N	220	220	220	220	212	184	220	220	220	220	220
10. Gender	Correlation	0,055	0,077	-0,047	0,084	-0,014	-,194**	0,051	0,069	0,110	1	0,042
	N	220	220	220	220	212	184	220	220	220	220	220
11. Age	Correlation	,137*	0,000	-0,029	-0,019	-0,117	-0,085	-0,060	-,133*	,498**	0,042	1
	N	220	220	220	220	212	184	220	220	220	220	220

*Note: **. A correlação é significativa no nível 0,01 (2 extremidades). *. A correlação é significativa no nível 0,05 (2 extremidades).*



Appendix 8: Regression Model

This regression model is a process procedure for SPSS written by Andrew F. Hayes with the Guilford Press. The conceptual model chosen for this study is model 8. The chosen variables for this model were: *Final PS* as the Dependent Variable (X); *Entity* as the Independent Variable (Y); *Social and Intellectual Traits* as the two Mediators (M); and *Position* as our Moderator (W); as Covariates, the variables chosen were: *Initial PS*, *Change is Realistic*, *Change makes Sense*, *Education*, *Gender* and *Age*. For this model, the sample size was constituted by exactly 164 participants.

Outcome Variable - Social Traits:

Model Summary

R	R-sq	MSE	F	df1	df2	p
0,4131	0,1706	5,7007	3,5202	9	154	0,0005

Model

	coeff	se	t	p	LLCI	ULCI
Constant	6,0729	1,4844	4,0913	0,0001	3,1406	9,0053
Entity	-1,9342	0,5274	-3,6675	0,0003	-2,976	-0,8923
Position	-0,725	0,5497	-1,3188	0,1892	-1,8109	0,361
Position x Entity	0,9123	0,7521	1,2129	0,227	-0,5736	2,3981
Initial PS	0,0999	0,1959	0,5101	0,6107	-0,287	0,4868
Change is Realistic	-0,5868	0,2631	-2,2303	0,0272	-1,1066	-0,067
Change makes Sense	0,2592	0,268	0,9672	0,335	-0,2703	0,7887
Education	-0,3605	0,195	-1,8491	0,0664	-0,7457	0,0247
Gender	-0,1657	0,3849	-0,4305	0,6675	-0,926	0,5946
Age	-0,0108	0,0163	-0,6626	0,5086	-0,0431	0,0215

Outcome Variable - Intellectual Traits:

Model Summary

R	R-sq	MSE	F	df1	df2	p
0,4272	0,1825	4,3522	3,8203	9	154	0,0002

Model

	coeff	se	t	p	LLCI	ULCI
Constant	4,2678	1,297	3,2906	0,0012	1,7057	6,83
Entity	1,2618	0,4608	2,7382	0,0069	0,3515	2,1721
Position	-0,2624	0,4803	-0,5463	0,5856	-1,2112	0,6864
Position x Entity	0,7045	0,6572	1,072	0,2854	-0,5938	2,0027
Initial PS	0,242	0,1711	1,4142	0,1593	-0,0961	0,5801
Change is Realistic	-0,3812	0,2299	-1,658	0,0994	-0,8353	0,073
Change makes Sense	-0,0468	0,2342	-0,1997	0,842	-0,5094	0,4159
Education	-0,173	0,1704	-1,0152	0,3116	-0,5095	0,1636
Gender	-0,011	0,3363	-0,0327	0,974	-0,6753	0,6534
Age	-0,0263	0,0143	-1,8414	0,0675	-0,0545	0,0019

Outcome Variable – Final PS:

Model Summary

R	R-sq	MSE	F	df1	df2	p
0,6834	0,467	0,5323	12,1094	11	152	0

Model

	coeff	se	t	p	LLCI	ULCI
Constant	1,7102	0,4825	3,5443	0,0005	0,7569	2,6635
Entity	-0,0508	0,1777	-0,2861	0,7752	-0,402	0,3003
Social Traits	0,0882	0,0266	3,3159	0,0011	0,0357	0,1408
Intelectual Traits	-0,053	0,0304	-1,741	0,0837	-0,1132	0,0071
Position	0,1687	0,1689	0,9984	0,3197	-0,1651	0,5024
Position x Entity	-0,2847	0,2312	-1,231	0,2202	-0,7415	0,1722
Initial PS	0,5452	0,0602	9,0507	0	0,4262	0,6642
Change is Realistic	-0,1346	0,0819	-1,6434	0,1024	-0,2964	0,0272
Change makes Sense	-0,105	0,0823	-1,2764	0,2038	-0,2675	0,0575
Education	0,0442	0,0603	0,7341	0,464	-0,0748	0,1633
Gender	0,1364	0,1177	1,1589	0,2483	-0,0961	0,3689
Age	0,0058	0,005	1,1529	0,2508	-0,0042	0,0158

Conditional direct effect(s) of X on Y:

Position	Effect	se	t	p	LLCI	ULCI
0	-0,0508	0,1777	-0,2861	0,7752	-0,402	0,3003
1	-0,3355	0,1817	-1,8462	0,0668	-0,6945	0,0235

Note: The level of confidence for all confidence intervals in output is 95%.

Appendix 9: Regression Model without non-workers

Model Summary

R	R-sq	MSE	F	df1	df2	p
0,4817	0,2321	5,9007	4,1633	9	124	0,0001

Model

	coeff	se	t	p	LLCI	ULCI
Constant	9,5208	1,7375	5,4795	0	6,0818	12,9598
Entity	-1,7655	0,5782	-3,0532	0,0028	-2,91	-0,621
Position	0,0668	0,644	0,1038	0,9175	-1,2077	1,3414
Position x Entity	-0,2141	0,8596	-0,249	0,8038	-1,9155	1,4874
Initial PS	-0,2478	0,2323	-1,0667	0,2882	-0,7075	0,212

Change is Realistic	-0,3408	0,3126	-1,09	0,2778	-0,9596	0,278
Change makes Sense	-0,2329	0,3211	-0,7253	0,4696	-0,8684	0,4026
Education	-0,3264	0,2131	-1,5315	0,1282	-0,7482	0,0954
Gender	-0,8892	0,428	-2,0776	0,0398	-1,7364	-0,0421
Age	-0,0252	0,0191	-1,3207	0,189	-0,0629	0,0125

Outcome Variable:
Intelectual Traits

Model Summary

R	R-sq	MSE	F	df1	df2	p
0,3334	0,1112	4,6942	1,7231	9	124	0,0905

Model

	coeff	se	t	p	LLCI	ULCI
Constant	5,4538	1,5497	3,5192	0,0006	2,3864	8,5212
Entity	1,019	0,5157	1,9758	0,0504	-0,0018	2,0398
Position	0,1779	0,5744	0,3097	0,7573	-0,9589	1,3147
Position x Entity	0,0778	0,7667	0,1015	0,9193	-1,4398	1,5954
Initial PS	0,0044	0,2072	0,0211	0,9832	-0,4057	0,4144
Change is Realistic	-0,2968	0,2789	-1,0643	0,2892	-0,8487	0,2551
Change makes Sense	0,1842	0,2864	0,6431	0,5214	-0,3827	0,751
Education	-0,015	0,1901	-0,0787	0,9374	-0,3912	0,3613
Gender	-0,5839	0,3818	-1,5296	0,1287	-1,3396	0,1717
Age	-0,0372	0,017	-2,188	0,0305	-0,0708	-0,0035

Outcome Variable:
Final PS

Model Summary

R	R-sq	MSE	F	df1	df2	p
0,6668	0,4446	0,5626	8,8799	11	122	0

Model

	coeff	se	t	p	LLCI	ULCI
Constant		0,5993	2,2893	0,0238	0,1856	2,5583
Entity	0,0257	0,1966	0,1308	0,8961	-0,3634	0,4148
Social Traits	0,1051	0,0323	3,2556	0,0015	0,0412	0,169
Intelectual Traits	-0,0434	0,0362	-1,2003	0,2323	-0,1151	0,0282
Position	0,1723	0,1989	0,8661	0,3881	-0,2215	0,5661
Position x Entity	-0,3444	0,2656	-1,2969	0,1971	-0,8702	0,1813
Initial PS	0,5757	0,0722	7,9759	0	0,4328	0,7186
Change is Realistic	-0,0756	0,0971	-0,7786	0,4377	-0,2679	0,1167
Change makes Sense	-0,0482	0,0999	-0,4826	0,6303	-0,246	0,1496
Education	0,0302	0,0666	0,4542	0,6505	-0,1016	0,1621
Gender	0,0254	0,1346	0,1889	0,8505	-0,241	0,2919
Age	0,0103	0,006	1,7244	0,0872	-0,0015	0,0222

Appendix 10: Exploratory Analysis

Two-Way Repeated Measures ANOVA:

Within-Subject Factors

Identification	Type of Trait	Depedent Variable
Associated	Social	Associated Social
	Intellectual	Associated Intellectual
Necessary	Social	Necessary_Social
	Intellectual	Necessary_Intellectual

Between-Subject Factors

Variable	Values	N
Entity	Human	82
	AI	82
Position	Colleague	85
	Leader	79

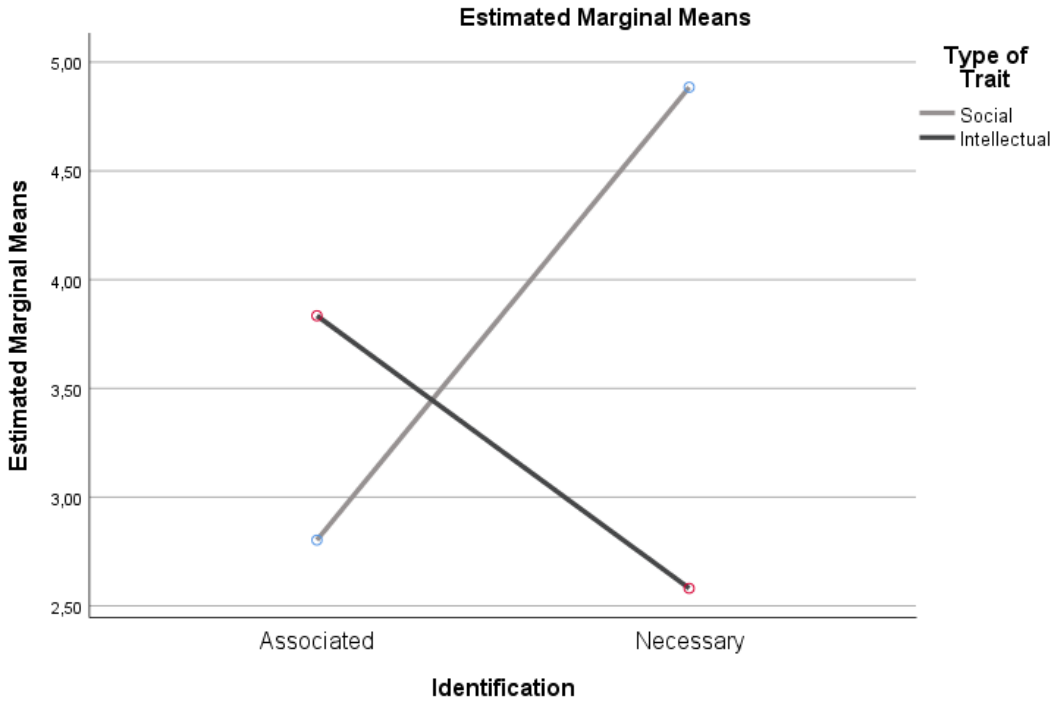
Descriptive Statistics

Dependent Variable	Entity	Position	M	SD	N
Associated_Social	Human	Colleague	3,98	2,7491	44
		Leader	3,18	2,7099	38
		Total	3,61	2,7431	82
	Leader	Colleague	1,95	2,0242	41
		Leader	2,10	2,1308	41
		Total	2,02	2,0666	82
	Total	Colleague	3,00	2,6186	85
		Leader	2,62	2,4719	79
		Total	2,82	2,5483	164
Associated_Intellectual	Human	Colleague	3,27	2,4147	44
		Leader	2,87	1,7884	38
		Total	3,09	2,1442	82
	Leader	Colleague	4,44	2,3564	41
		Leader	4,76	1,7997	41
		Total	4,60	2,0897	82
	Total	Colleague	3,84	2,4439	85
		Leader	3,85	2,0197	79
		Total	3,84	2,2427	164
Necessary_Social	Human	Colleague	4,89	3,0970	44
		Leader	4,92	2,5296	38
		Total	4,90	2,8311	82
	Leader	Colleague	4,39	2,8185	41
		Leader	5,34	3,4323	41
		Total	4,87	3,1574	82
	Total	Colleague	4,65	2,9589	85
		Leader	5,14	3,0202	79
		Total	4,88	2,9896	164
Necessary_Intellectual	Human	Colleague	2,66	2,2407	44
		Leader	2,39	1,9388	38
		Total	2,54	2,0976	82
	Leader	Colleague	2,24	2,4164	41
		Leader	3,02	2,9282	41
		Total	2,63	2,6966	82
	Total	Colleague	2,46	2,3225	85
		Leader	2,72	2,5061	79
		Total	2,59	2,4088	164

Test of Within-Subjects Effects

Source		Type III Sum of Squares	df	MSE	F	p	Partial Eta Squared
Identification	Sphericity Assumed	28,079	1	28,079	5,826	0,017	0,035
	Greenhouse-Geisser	28,079	1	28,079	5,826	0,017	0,035
	Huynh-Feldt	28,079	1	28,079	5,826	0,017	0,035
	Lower-bound	28,079	1	28,079	5,826	0,017	0,035
Identification * Entity	Sphericity Assumed	0,100	1	0,100	0,021	0,886	0,000
	Greenhouse-Geisser	0,100	1	0,100	0,021	0,886	0,000
	Huynh-Feldt	0,100	1	0,100	0,021	0,886	0,000
	Lower-bound	0,100	1	0,100	0,021	0,886	0,000
Identification * Position	Sphericity Assumed	12,777	1	12,777	2,651	0,105	0,016
	Greenhouse-Geisser	12,777	1	12,777	2,651	0,105	0,016
	Huynh-Feldt	12,777	1	12,777	2,651	0,105	0,016
	Lower-bound	12,777	1	12,777	2,651	0,105	0,016
Identification * Entity * Position	Sphericity Assumed	0,231	1	0,231	0,048	0,827	0,000
	Greenhouse-Geisser	0,231	1	0,231	0,048	0,827	0,000
	Huynh-Feldt	0,231	1	0,231	0,048	0,827	0,000
	Lower-bound	0,231	1	0,231	0,048	0,827	0,000
Type of Trait	Sphericity Assumed	66,237	1	66,237	18,223	0,000	0,102
	Greenhouse-Geisser	66,237	1	66,237	18,223	0,000	0,102
	Huynh-Feldt	66,237	1	66,237	18,223	0,000	0,102
	Lower-bound	66,237	1	66,237	18,223	0,000	0,102
Type of Trait * Entity	Sphericity Assumed	106,546	1	106,546	29,313	0,000	0,155
	Greenhouse-Geisser	106,546	1	106,546	29,313	0,000	0,155
	Huynh-Feldt	106,546	1	106,546	29,313	0,000	0,155
	Lower-bound	106,546	1	106,546	29,313	0,000	0,155
Type of Trait * Position	Sphericity Assumed	0,021	1	0,021	0,006	0,940	0,000
	Greenhouse-Geisser	0,021	1	0,021	0,006	0,940	0,000
	Huynh-Feldt	0,021	1	0,021	0,006	0,940	0,000
	Lower-bound	0,021	1	0,021	0,006	0,940	0,000
Type of Trait * Entity * Position	Sphericity Assumed	0,021	1	0,021	0,006	0,940	0,000
	Greenhouse-Geisser	0,021	1	0,021	0,006	0,940	0,000
	Huynh-Feldt	0,021	1	0,021	0,006	0,940	0,000
	Lower-bound	0,021	1	0,021	0,006	0,940	0,000
Identification * Type of Trait	Sphericity Assumed	454,992	1	454,992	136,214	0,000	0,460
	Greenhouse-Geisser	454,992	1	454,992	136,214	0,000	0,460
	Huynh-Feldt	454,992	1	454,992	136,214	0,000	0,460
	Lower-bound	454,992	1	454,992	136,214	0,000	0,460

Identification * Type of Trait * Entity	Sphericity Assumed	88,254	1	88,254	26,421	0,000	0,142
	Greenhouse-Geisser	88,254	1	88,254	26,421	0,000	0,142
	Huynh-Feldt	88,254	1	88,254	26,421	0,000	0,142
	Lower-bound	88,254	1	88,254	26,421	0,000	0,142
Identification * Type of Trait * Position	Sphericity Assumed	2,707	1	2,707	0,811	0,369	0,005
	Greenhouse-Geisser	2,707	1	2,707	0,811	0,369	0,005
	Huynh-Feldt	2,707	1	2,707	0,811	0,369	0,005
	Lower-bound	2,707	1	2,707	0,811	0,369	0,005
Identification * Type of Trait * Entity * Position	Sphericity Assumed	0,307	1	0,307	0,092	0,762	0,001
	Greenhouse-Geisser	0,307	1	0,307	0,092	0,762	0,001
	Huynh-Feldt	0,307	1	0,307	0,092	0,762	0,001
	Lower-bound	0,307	1	0,307	0,092	0,762	0,001



Paired Samples T tests:

Pairing by Identification – Associated versus Necessary:

Paired Sample Statistics

	Pairs	M	N	SD
Pair 1	Associated Social	2,82	164	2,5483
	Necessary Social	4,88	164	2,9896
Pair 2	Associated Intellectual	3,84	164	2,2427
	Necessary Intellectual	2,59	164	2,4088

Paired Samples Correlations

Pairs		M	Corr	p
Pair 1	Associated Social & Necessary Social	164	0,370	0,000
Pair 2	Associated Intellectual & Necessary Intellectual	164	0,306	0,000

Paired Sample Tests

Pairs	Paired Differences					t	df	p
	Mean	SD	Std. Mean Error	95% CI of the Difference				
				Lower	Upper			
Associated Social - Necessary Social	-2,067	3,1294	0,2444	-2,5496	-1,5845	-8,4590	163	0,000
Associated Intellectual - Necessary Intellectual	1,256	2,7439	0,2143	0,8330	1,6792	5,8624	163	0,000

Pairing by Type of Trait – Social versus Intellectual:

Paired Sample Statistics

Pairs		M	N	SD
Pair 1	Associated Social	2,82	164	2,5483
	Associated Intellectual	3,84	164	2,2427
Pair 2	Necessary Social	4,88	164	2,9896
	Necessary Intellectual	2,59	164	2,4088

Paired Samples Correlations

Pairs		M	Corr	p
Pair 1	Associated Social & Necessary Social	164	0,264	0,001
Pair 2	Associated Intellectual & Necessary Intellectual	164	0,493	0,000

Paired Sample Tests

Pairs	Paired Differences					t	df	p
	Mean	SD	Std. Mean Error	95% CI of the Difference				
				Lower	Upper			
Associated Social - Associated Intellectual	-1,024	2,9159	0,2277	-1,4740	-0,5748	-4,4990	163	0,000
Necessary Social - Necessary Intellectual	2,299	2,7652	0,2159	1,8724	2,7252	10,6461	163	0,000