iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Paying for Privacy in a Digital Age: Willingness to pay for attributes in a VPN (Virtual Private Network) service, and its relation to privacy literacy

Eduardo Pedro de Almeida Ribeiro Jardine Neto

Master in Economics

Supervisor:
PhD Ana Cristina Narciso Fernandes Costa, Assistant Professor, Iscte-Iul

PhD Nuno Manuel Mendes Cruz David, Associate Professor, Iscte-Iul

November, 2022

BUSINESS
SCHOOL

CIÊNCIAS SOCIAIS
E HUMANAS

Department of Economics

Department of Political Economy

Paying for Privacy in a Digital Age: Willingness to pay for attributes in a VPN (Virtual Private Network) service, and its relation to privacy literacy

Eduardo Pedro de Almeida Ribeiro Jardine Neto

Master in Economics

Supervisor:
PhD Ana Cristina Narciso Fernandes Costa, Assistant Professor, Iscte-Iul

PhD Nuno Manuel Mendes Cruz David, Associate Professor, Iscte-Iul

November, 2022

I want to, first and foremost, thank my parents for everything they've done for me. Both for my mom's emotional support and emphasis on my education (as the greatest investment one can make), and to my dad for his constant help, optimism, and advice towards the hardest challenges of life. A person I would also like give a special thank you to is my number one reviewer, my girlfriend Patricia (bless you). Thank you for keeping me sane throughout all of this and making me stay pragmatic and focused on my end goal. You're awesome. I would also like to extend my sincere gratitude to my supervisor Ana Costa and co-supervisor Nuno David, for without their help and extensive reviewing I'm sure this Thesis would not be what it is, and I would not have learned what I did. My sincere thanks to my aunt Sofia and uncle Diogo, for without their help this masters would not be possible at all, as well as my aunt Ana and aunt Rosario for all their support in these years. Lastly, thank all the people and best mates who helped me become who I am throughout the years (Fattha, Max, Michael, Mario, Diogo, Rodrigo, Tiago). And with this, I close another chapter.

Resumo

A presente Tese de Mestrado procura estimar a disponibilidade a pagar por atributos num serviço VPN (Virtual Private Network), e a sua relação com literacia no ramo da privacidade. Para o efeito, foi realizada uma experiência de escolha discreta (discrete choice experiment) acompanhada por um questionário de literacia na área da privacidade.

Os atributos e os níveis para a experiência foram selecionados através de uma pesquisa qualitativa que teve como base um *focus group*, entrevistas a especialistas, e uma revisão de literatura. Depois da seleção dos atributos e níveis, foram criadas 12 séries de escolhas, cada uma com duas alternativas possíveis a comparar dois serviços VPN diferentes. A experiência foi realizada com uma amostra de 84 participantes, para um total de 2016 observações e 1008 escolhas. Através do uso de software open-source (Rstudio), foram geradas regressões logísticas multinomiais, para estimar os atributos mais valorizados e a disponibilidades a pagar. Esta análise foi realizada para o total da amostra e para subgrupos com níveis de literacia de privacidade diferentes.

Os resultados obtidos trazem um entendimento acrescido sobre a valorização dos vários atributos num serviço VPN, assim como a disponibilidade a pagar por níveis diferentes de cada atributo, e como conhecimentos na área da privacidade afetam esta valorização. A análise contribuí para a literatura no ramo de disponibilidade a pagar por ferramentas para proteção da privacidade (privacy-enhancing tools) assim como para o debate geral sobre o valor da privacidade.

Palavras-Chave: Disponibilidade a pagar, serviço VPN, literacia de privacidade, experiência de escolha discreta, programação em R, regressão logística multinomial
Classificação JEL: C90; C83

Abstract

This Master Thesis focuses on estimating the willingness to pay for attributes in a VPN (Virtual Private Network) service, and how those estimates relate to privacy literacy. In this Thesis, I conducted a discrete choice experiment accompanied by a privacy literacy questionnaire.

Attributes and levels for the discrete choice experiment were selected through qualitative research encompassing a focus group, expert interviews, and a relevant literature review. After the attributes and levels were selected, 12 choice sets were created, with each choice set comparing two competing VPN services.

The experiment was carried out on a sample of 84 participants, for a total of 2016 observations with 1008 choices. Subsequent multinomial logit models were estimated based on collected data, through programming in R using open-source software (Rstudio), to estimate the most valued attributes and the respective willingness to pay estimates. This analysis was performed both for the general sample, and for subgroups with differing levels of privacy literacy.

The results bring additional insight into how individuals perceive the significance of various attributes of a VPN service, their willingness to pay for different levels of each attribute, and how privacy literacy impacts that valuation. The analysis extends the literature on willingness to pay for privacy-enhancing tools and contributes to the ever-evolving general debate on the worth of privacy.

Keywords: Willingness to pay, VPN service, privacy literacy, discrete choice experiment, programming in R, multinomial logit model
JEL Classification: C90; C83

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Introduction

The value individuals place on their privacy, and how much they are willing to pay to protect it is a subject that interests policymakers, businesses and researchers. On the side of businesses, consumer data has become a central aspect of many online firms. Firms like Google, Facebook and Twitter constitute quick examples of business models that derive substantial revenue from ads , and where user data has became a staple in their business models.

Despite the surface of significant advances in data protection legislation such as the GDPR (General Data Policy Regulation), users often rely on privacy-enhancing tools to retain a level of anonymity online. One of the most popular tools is a Virtual Private Network (VPN), normally delivered as a service by a VPN firm. This software can be used for several end goals, that range from greater online privacy to the access of geo-blocked content as well as bypassing of censorship.

I examine preferences for the attributes of such services, as well as how they relate to privacy literacy. Despite significant efforts to quantify willingness to pay to protect personal privacy in the relevant literature, attempts to quantify willingness to pay in regard to privacy-enhancing tools (PETs) have been scarce. However, since these technologies are often used to achieve greater levels of privacy and data protection, the preferences for this type of software can provide useful insights to the privacy debate. Moreover, knowing how differences in privacy literacy affect the preferences for attributes will tell us what privacy-conscious individuals are looking for in such software.

This research was inspired on the work done by Sombatruang et al. (2020) with notable differences. First off, while Sombatruang et al. (2020) looked into evaluating willingness to pay in a mobile scenario, I will evaluate the case for desktop versions of the software. I will also consider privacy literacy, and how that privacy literacy impacts the willingness to pay estimates for attributes in a VPN service.

The work is structured in the following way. In the first section of the literature review, I will highlight the major contributions and importance of privacy to economics. Following that brief context, I will mention studies that identified consumer privacy valuations, and the differences between them, as well as their impact in the privacy debate theme. The rationality of the consumer when making privacy related decisions will next be put in question, and the consequences that holds for policy, specifically in the case of the General Data Policy Regulation (GDPR). Following that context, I will go over some of the various privacy enhancing tools that exist in the market, with special emphasis on

Virtual Private Networks (VPN) services, that have gained great traction over the last years. The last section of the literature review evaluates research gaps and elaborates on the veins of future research and unexplored gaps I propose to contribute to.

The main contributions and hypotheses I venture forward are in Chapter 3 (Research Design). After, in Chapter 4 (Methodology), I will go over the methodological framework to answer the proposed hypotheses. This chapter includes a deviation in methodology of qualitative research over previous work (for example, Sombatruang et al. (2020) conducted a series of personal, one-on-one interviews with non-experts), focusing on expert interviews as well as a focus group with non-experts. This will allow the extraction of meaningful insights across the entire specter of privacy literacy, to understand better the factors and motivations that guide different individuals toward this service. I will also construct a questionnaire based on dimensions identified by Trepte et al. (2015) to assess privacy literacy scores to categorize participants into privacy literacy categories, for subsequent analysis. After a description over how the DCE will be deployed, I will end the chapter with the choice of models for analysis and econometric reasoning.

The results will be explored in Chapter 5 (Empirical Results), where participants' choices will serve to construct a multinomial logit model, traditionally used in discrete choice experiments literature, as well as the mixed multinomial logit model, which takes the heterogeneity of the population into account and relaxes strict assumptions in the standard multinomial model. Relevant additional models will be constructed to account for privacy literacy differences and evaluate key differences. These models will be programmed with R, using novel packages [1] from Croissant (2020) and Helveston (2022), and significantly expand on the practical use of Rstudio, an open-source software, for the purpose of discrete choice experiment analysis. These models will be tested accordingly regarding their goodness-of-fit, perceived accuracy and reliability in terms of statistical validity.

Finally, on Chapter 6, I will draw conclusions on the key results and suggest where future research could be applicable.

[1] R packages are extensions to the R statistical programming language. R packages contain code, data, and documentation in a standardised collection format that can be installed by users of R

CHAPTER 2

Literature Review

This literature review will attempt to summarize the various thoughts pertaining to economic behavior and privacy, particularly in the current European Law landscape and determine gaps of research and contributions to literature.

In an effort to empower consumers, recent data protection legislation was passed in 2018 in the European Union in the form of the GDPR (General Data Policy Regulation). This is the largest and most ambitious policy governing both people's rights towards their own data, and firms' responsibilities when keeping and processing said data, with dire consequences for firms who fail to uphold it (up until 4 per cent of turnover, or up until 20 million euros) (Art. 83 EU-GDPR). The effects of this regulation have been felt in many spheres - law, economics and behavioral economics, as well as marketing and computer science - and its apparent success has given strength to similar movements such as the The California Consumer Privacy Act in the United States and recent GDPR inspired legislation in certain South American Countries such as Brazil and Panama (Rodriguez and Alimonti, 2020).

Individuals have been bestowed with new or reshaped privacy rights, such as the right to be informed, the right of erasure, and the right of data portability. But to what extent is the consumer using these rights? And to what extent can the consumer rationally navigate the privacy landscape, if at all? For example, McDonald and Cranor (2008) estimated it would take on average 200 hours per year to read every privacy notice a consumer faces. Even despite the GDPR banning "long illegible terms of conditions", Facebook, Twitter and Google increased their policies an average of 1300 words since its implementation, with each policy signifying a 20-minute read (Coleman, 2018). This is only the time it takes to read the terms; let alone the time it takes to reflect on them and the consequences they might hold for an individual privacy preferences. The advancements in data processing do not help the readability, with the emergence of complicated machine learning models working in a "black box" fashion, where it is not possible to know how the data was analyzed and weighted but only the results. This widens the gap between how firms and consumers perceive information, especially if consumers have low technological literacy.

2.1. Privacy and economics

We have currently at our disposal many services that seem apparently free. They go from Google's email and search engine services that facilitate communication and give guidance to people in their daily life's, to Facebook's global social network that has connected millions and provides a source of daily joy for many. These services have frequently puzzled certain people as to how they make a profit despite not charging consumers for their services. The answer "Senator, We Sell Ads" Mark Zuckerbeg gave to Senator Orrin Hatch, who asked how it is possible to sustain a business model without the users paying, is not far from the truth (Stewart, 2018). Much of the revenue Facebook makes comes from ads (up to 98 per cent of total revenue according to Johnston (2021)), and consumers pay for these services with their attention and personal information (Shapiro, 2019). Recent events brought privacy issues to the spotlight, such as the Cambridge Analytica scandal, or the European Commission recent efforts in the form of the GDPR. The value of personal data, and the value individuals give to their personal data are complex topics that involve various trade-offs, (in)tangible economic losses and gains, and issues that affect both the individual and society as a whole (Acquisti et al., 2016). Data has been described as the "new oil" (Bhageshpur, 2019), and its impact with the advancement of new algorithms and data tools to process it can go from price personalization and discrimination, to societal benefits.

The question of how much individuals value their privacy is an elusive question, and there seems to be no unifying theory for privacy (Acquisti et al., 2016) with concerns about privacy arising from different contexts. To start with, individuals' conceptions of privacy are different, both within academia (De Capitani Di Vimercati et al., 2012) and in public awareness. Despite underlying differences in culture, age and individual perceptions, a fundamental aspect that has come to define privacy is it fundamentally concerns the limits between the private and public (Altman, 1976). I analyze privacy under the lenses of informational privacy, which has served as the primary focuses of privacy economics (Acquisti et al., 2016). There are direct consequences to the sharing of personal information - for example, a study by Edelman and Luca (2014) found widespread discrimination against African-Americans in the AirBnB platform, where guests with distinctively African-American-sounding names were decidedly less likely to be accepted than identical guests with white-sounding names. This followed a 2014 study by the same researchers where they found black hosts charged approximately 12 per cent less for rentals than non-black hosts (Edelman and Luca, 2014). When one shares information online the consequences of such disclosure are varied and often unpredictable and abstract. There are fears of price discrimination, data breaches, and even more drastic consequences such as identity theft or fraud. On the other hand, there are evident immediate benefits that go from allowing information location to better navigate through a foreign city, to personalized advertising that is not only more effective for firms but might better fit consumer needs and personality. Disclosing information can even in some scenarios be intrinsically rewarding (Tamir and Mitchell, 2012), such as sharing personal opinions and views on social network sites, and it can be done out of wide array of different motivations, from social engagement to altruism to personal gain (Oh and Syn, 2015).

At its core, the economics of privacy concerns these and other trade-offs associated with the balance of public and private spheres between individuals, organizations and governments (Acquisti, 2016).

2.2. Consumers valuation of privacy

Now that we established that personal data has value and its use economic consequences, the matter of quantifying said value is much more difficult. There is not an obvious way of valuing privacy and personal data. Should the reference point be the price one accepts for giving away the data, or the amount one pays to protect it? Should it be the cost/reward one gets from the exposure of personal data (estimating, for instances, a probabilistic expected value taking into account odds of events such as data breaches, wrongful use and immediate benefits) or should it be the expected value a firm can generate from that information? To add to this conundrum, there is not a market where data users can access and willingly sell their data (Acquisti, 2015), and they are often vulnerable to biases and heuristics present in decision making (Acquisti et al., 2013) (Acquisti and Grossklags, 2004) that pose issues for monetizing these amounts.

Willingness to pay (WTP) for privacy and willingness to accept (WTA) to give up privacy have served as the primary instruments in the literature for determining personal monetary valuations of privacy and personal information (Wagner et al., 2018). These usually take two primary forms – experiments such as discrete choice experiments, or laboratory and field experiments. As proposed by Wagner et al. (2018), who conducted the first extensive literature review on the subject, these valuations while usually significant, often differ tremendously due to different research methods and contexts. For instances in Europe, a recent study by Potoglou et al. (2017) found the willingness to pay for a monthly premium of privacy enhancing services (ISP hides information on users' online activity and warns user which websites do not meet desired level of privacy) varied between 3 EUR to 5000 EUR, depending on income.

Certain authors caution against taking willingness to pay and willingness to accept at face value. The literature is clear on the importance of context, and the danger of extrapolating a privacy valuation under a certain set of circumstances to another (Solove, 2020). Solove (2020) argues individuals' value different information differently, and they value their privacy by taking into account risks such as that of the receiver using their personal information against their interests. This calculus is both dependent on a range of factors (for example, who receives the data, what is the type of data, under what circumstances was it provided, and so on) and subject to decision making biases, such as the tendency to discount long-term risks disproportionately. Besides this, although data is often traded between firms, the consumer rarely has access to these markets, and they can not sell individually their own data directly to firms. In a study where individuals bid for a price for their location privacy when prompted by a fictional firm (Brush et al., 2010), they found people had a difficult time coming up with values, and often looked for social cues (such as wondering what value other people ask for). Beyond this, consumers might be placing a “moral” value to the data. Winegar and Sunstein (2019) measured willingness to pay and willingness to accept of individuals to look for an endowment effect, and in doing so found wide disparities within the WTP and WTA valuations. For example, certain individuals considered their willingness to accept to give up private information at over 1 million dollars. However, 14 per cent gave a willingness to pay for privacy value of zero – a value that could potentially serve as a sign of protest against paying to protect one’s own data. Perhaps one of the most significant aspect of these studies are not the absolute values found, that vary wildly between individuals, context, and methods of calculation, but rather insights they give into how consumers value different information. A study by Skatova et al. (2019) conducted in the UK, for example, found despite valuations varying between individuals, they were consistent in how they valued different “tiers” of information – with Banking Transactions and Medical Records at the top, followed by Browsing History, Social Media Data and Mobile Phone GPS. This falls in line with the analysis from Wagner et al. (2018) who found the strongest similarity between various willingness to pay and accept studies was the more sensitive the data was and the more easily people were identified, the higher price they attached to said data.

Another key element is often consumers place a low value on their data because strong regulation ensures firms handle it wisely – following the approach data is essentially a risk assessment calculus taken by the consumer, regulation such as the GDPR might lower the amount someone needs to give up this data. This might mean consumers take into account the fact firms treat their data with greater rigor and are subject to greater risks and consequences if they act in a nefarious way, and as such adjust their risk assessment calculus accordingly. Such evaluations taking into account the GDPR and effect of such regulation on willingness to pay or willingness to accept are missing from the literature.

In the precise context of the GDPR, there was a study conducted by Sobolewski et al. (2017) in the form of a discrete choice experiment run on Polish students that focused on measuring willingness to pay for each major point of the GDPR. This was done with the goal of determining the welfare gain from the policy. They found a combined welfare of of 6.5 EUR per capita per month, with the right of erasure of personal data valued at 1.4 EUR per month. As with the previously studies mentioned, this study too revealed a large amount of preference heterogeneity.

2.3. Privacy paradox, privacy beliefs and consumer behavior

While it’s generally undisputed individuals care about their privacy- (Kumaraguru and Cranor, 2005) provide a comprehensive summary of Westin’s surveys on privacy, spanning more than 30 years of analysis - individuals behavior often does not translate those same beliefs. In the literature, such similar situations have been popularized under the “Privacy Paradox” label (Acquisti, 2015). One of the first examples of the possible existence of a privacy paradox was observed in Spiekermann Grossklags (2001), where despite numerous surveys indicating people place a high value on their personal privacy, online shoppers quickly gave up various sensitive personal information when asked by an anthropomorphic 3-D shopping bot. Subsequent studies found similar dichotomies between beliefs and behavior, with (Brown and Muchira, 2004) coining the term. Similar studies followed, adapting insights from behavioral economics such as incomplete information, bounded rationality (Acquisti and Grossklags, 2004) and heuristics/biases present in decision making like endowment effect (Acquisti et al., 2013) hyperbolic discounting (Acquisti and Grossklags, 2003), overconfidence (Wagner and Mesbah, 2019) and others (Kolakis, 2017). These findings question consumer rationality, and set psychological factors that limits their ability to exercise privacy beliefs adequately. According to Acquisti et al. (2015), this heavily contributes to the privacy paradox.

There is a small minority that advocates consumers preferences for privacy are not very large, based on their behavior, in what is often called the revealed preferences argument (Solove, 2020). It is particularly important to mention this as it has severe consequences for privacy policy, for if behavior of consumers truly represents their true preferences, then that lessens the need for regulation. However, this argument taken from classical economic literature loses strength in the privacy debate, as highlighted by Solove (2020), for it is nearly impossible for the consumer to rationally behave in today's privacy landscape. Revealed preferences being different points more to the direction of a hard to manage privacy situation, rather than an outright disregard for privacy.

Despite many studies supporting the existence of a privacy paradox, they are often inconsistent in valuations, even within the same context (Kokolakis, 2017). They are dependent of privacy scenarios, but also of how the researcher considered the privacy paradox (specific concerns versus behaviors, intentions versus behaviors, ...) between various pairs of mental states and behaviors (Acquisti et al., 2020).

Even studies critical of the paradox (Solove, 2020) are unanimous in recognizing hurdles in the way of individuals manage privacy. The hypothetical perfectly rational consumer can hardly navigate the complex privacy maze present in today's world – knowing what information every company holds is not possible, even if he has the right of erasure and portability of said data. Reading every privacy notice is not feasible – even if one has interest in what level of protection firms offer. Even if privacy paradox has a number of sensible explanations that could translate it into more of a phenomenon, knowledge of decision making struggles should be taken into account when conducting policy (Acquisti et al., 2020), especially when taking into account matters like informed consent.

Summarizing, there are a couple of key explanations that account for some researchers finding a dichotomy in one scenario, and in another not. They stem from different definitions of the paradox, different methods of measuring it, uncertainty on individual privacy preferences, asymmetry of information between firms and consumers as to how data is processed and consequences of processing it (Hermstruwer, 2017), and behavioral biases and heuristics (Acquisti et al., 2015). The explanations are many, and as noted by Acquisti et al. (2020), not mutually exclusive. Notwithstanding , it is important to take the paradox as a valuable insight into how individuals might be limited in their search for ideal or quasi-ideal privacy, and as such delimit the role of policy in managing such struggles.

2.4. Consequences of the privacy paradox to policy – GDPR analysis

As mentioned earlier, the privacy paradox is becoming less of a paradox, since there are many reasonable explanations for its existence. Nevertheless, many of the issues the privacy paradox highlighted – from irrational decision making, to issues such as information overload and lack of personal control – should be taken into account when formulating policy (Acquisti et al., 2020). Moreover, insights from willingness to pay and willingness to accept studies can shed some light into how individuals value information in certain contexts, and the types of information and personal characteristics that increase these valuations.

There have been multiple attempts to provide consumers with more rights regarding their own privacy. In Europe, the most striking and recent example is the GDPR (General Data Regulation Policy), that takes an unprecedented approach to data privacy and data rights.

The right to privacy has been deemed important since as early the 1950's European Convention on Human Rights, but before the advent of the GDPR, the matter of privacy law was majorly reliant to each member state with the European Data Protection Directiveblishing minimum data privacy and security standards (Bhageshpur, 2020). The rapid changing internet and increasing power of personal information marked the need for privacy legislation to be more comprehensive and cover recent developments in how firms gather, keep and analyze data.

In 2016, the GDPR entered into force after passing the European Parliament, and as of May 25, 2018, all organizations needed to be compliant with it. As of now, it has become a staple name, with 69 per cent of residents in the European Union having heard of it (Wigand et al., 2020), and up to 144,376 complaints to data protection authorities recorded over the first year of its implementation (Center for Data Innovation, 2020). It has been described by the European Commission as a success, and it has been often quoted as an example of a good privacy policy (Wigand et al., 2020).

Perhaps the most telling aspect of the GDPR is the leaning of responsibilities to firms – where the data controller "is responsible for ensuring that data is processed in compliance with the principles of lawfulness, fairness, transparency, data minimization, accuracy, storage limitation, integrity, and confidentiality" (European Comission, 2020). It also gives the following rights to consumers, namely the right to be informed, right of access, right of rectification, right to erasure, right to restrict processing, right to data portability, right to object and rights pertaining to automated decision making and profiling.

Two studies were found that juxtapose the current GDPR law with the existence of behavioral biases and heuristics found from a behavioral economics standpoint (Hermstruwer, 2017; van Ooijen and Vrabec, 2019). It is difficult to find extensive literature on the subject due to how recent the legislation and its impacts are.

Never the less, Hermstruwer (2017) provides many interesting insights. It evaluates in particular the question of consent, and where the GDPR fails to address certain concerns.

A helpful takeaway the GDPR took from behavioral economics and decision making on consent notices is the idea of forbidding defaults (framing of the consent policy notice firms must have users accept). It has been established in the behavioral economic literature that consumers often do not change the default option (Löfgren et al., 2012), and by not allowing for silence or pre-ticket boxes to constitute consent, the GDPR acts to counter this bias.

There are also guidelines orienting consent policies to be transparent and easy to read. However, as noted by Hermstruwer (2017), it is difficult to balance clarity and transparency with high thresholds of information that must be provided. As such, individuals often have to resort to rules of thumb due to limited attention spans and attribute substitution to manage the overload of information consent notices have. Individuals often end up being overloaded and simply have what has been described as a digital resignation (Draper and Turow, 2019). To add to this, although privacy notices are required to state the purpose of processing and the recipients of the data, (Solove, 2020) notes that individuals often don't know what firms hold their data, which data they hold, and even to what they are using it for. Giving people more data rights, while important by itself, is not a fix for the asymmetry between firm and consumer knowledge (an often quoted cause of the privacy paradox).

2.5. Privacy enhancing technologies

Certain tools for privacy management exist. As exposed earlier, there are numerous hurdles with the effectiveness of consent and personal management of privacy, and as such some tools were proposed to better enable the individual to act in greater alignment with his privacy beliefs. For example, an AI program to manage smartphone privacy was developed by Liu et al. (2016) and was found to be easily used, although the participants were young and had a good level of technological literacy. There has been little attention, in part to the new emergence of some of these tools, to the willingness to pay for increased privacy mechanisms. To my knowledge, there is no tool in the market yet that can tell consumers what firms hold their data, or that can synthesize consent notices in a way that's easy to understand, and compare it with a constructed privacy preference profile. Advancements in machine learning should not be underestimated, especially to better serve consumers. Another area where some sort of tool is needed to better manage privacy is in the field of Internet of Things (IoT), where information travels seemingly to devices like smart watches, with no possibility to display lengthy privacy polices and as such make difficult the matter of informed consent. As of late, VPNs have also gained traction in the field of privacy enhancing tools.

2.5.1. VPN programs as privacy enhancing technologies

A popular, commonly advertised tool to aid those who seek privacy is a VPN (Virtual Private Network). VPN products create a secure connection, or a “tunnel”, to a secured server that in turn connects them to their intended destination (Ramesh et al., 2022). This tunnel provides extra encryption that acts as a protection from surveillance from the immediate networks, allowing the bypass of blocks as well as disguising the user’s IP address. These products are often easy to use, when compared to alternative approaches such as TOR[1] Moreover, as of late ads for this product have skyrocketed, with Akgul et al. (2022) estimating over 17.1K videos advertising the software, totaling 4.4B views in Youtube alone. As commercial VPNs reach wide use, being a 15 billion industry in 2018 and predicted to grow 20 % by 2022 Khan et al. (2018), questions come up regarding its effectiveness as a privacy tool, the transparency of the firms offering VPN services, as well as to what extend the average consumer can rationally navigate the multitude of different products currently in the market. The barrier of entry for a firm to enter the VPN is also relatively low - with a wide array of VPN services being offered to the consumer, at competing price points. The purpose of using a VPN is also not solely related to privacy. Namara et al. (2020) goes over the emotional and practical considerations that guide adoption and abandonment of VPNs as a privacy enhancing technology. A significant portion of VPN users are not interested in privacy at all, using them for practical reasons such as accessing geo-blocked content, evading censorship, or bypassing blocks set by their Internet Service Provider (ISP). On the other hand, those who adopt VPNs as a privacy tool do so for longer than those who do it for practical non-privacy reasons, and often do it for emotional reasons such as heightened privacy concerns, fear of internet surveillance, media attention and dislike of the current change or lack of privacy legislature (Namara et al., 2020).

2.6. Gaps of research

Based on previous research , I have identified the following research gaps.

First off, I highlight the dependency in context when it comes to evaluating factors influencing the demand of a VPN service. Geographical and cultural differences likely pose shifts to what drives the need of a VPN - for example, it is expected authoritative countries where internet use is more heavily regulated will find a greater percentage of users who seek to hide their activity of their ISP, or bypass blocks set by the state on certain external content. On the other hand, European countries covered by the GDPR might exhibit less distrust of privacy legislature, and as such discount the need to use a VPN for privacy reasons. Examining these cultural and geographical differences is a missing area of research, and consumers from different privacy landscapes might seek a different adoption of VPN services. The current research on willingness to pay for a VPN is also limited. The only literature found, Sombatruang et al. (2020), was directed to

[1] Tor, short for The Onion Router, is free and open-source software for enabling anonymous communication.

a VPN service in the context of a mobile app setting. VPN services are offered across many platforms, and so far no research examined the willingness to pay for a computer program. Besides this, the effect of privacy literature and preferences on the willingness to pay for such a VPN has been disregarded. Yet, certain characteristics of VPN services are likely to appeal directly to a more privacy conscious audience, and could result in different willingness to pay. Along with this, previous research suggests those who pay for emotional reasons to do for longer (Namara et al., 2020) but it's not clear to what that means monetarily. The fact VPN services often result in trade-offs such as a decrease of internet speed or less efficiency of services such as GPS navigation is likely to affect this estimate as well. To add to this, not a lot of research is conducted in the European Union, where significant shifts regarding privacy rights happened over the last years, and distrust of internet service provider might be lower than in other, less regulated markets.

2.7. Proposed contributions to existing literature

Throughout the review of the relevant literature, I highlighted potential contributions to the existing research the work I develop will contain. Those are:

(1) Use of a Discrete Choice Experiment (DCE), which are grounded in economic theory that have a long-standing, well-tested theoretical basis in random utility theory.
(2) Assign a monetary valuation to the various attributes pertaining to the adoption of a VPN service.
(3) Conduction of several interviews to experts in the field of Informatics, Information Security and Data Privacy Law to gather a better understanding of the usage, views, and significant attributes of relevance in the choice of a VPN service.
(4) Conduction of a focus group with non-experts, with the emphasis on a dynamic discussion and answering of semi-open questions on the use of streaming platforms, experience with VPNs, valued attributes and other privacy related questions.
(5) Relate, using the DCE, how privacy literacy, which encompasses knowledge of several dimensions identified by Trepte et al. (2015) affects the willingness to pay for attributes in a VPN service.
(6) Conduction of a discrete choice experiment with either entirely free software (Rstudio[2]), or software easily available to students using student licenses (JMP[3] and Conjointly[4]).

[2] RStudio is an integrated development environment for R, a programming language for statistical computing and graphics

[3] JMP is a suite of computer programs for statistical analysis developed by JMP, a subsidiary of SAS Institute.

[4] Conjointly is an all-in-one survey research platform that focuses on discrete choice experiments, with easy-to-use advanced tools for pricing research

CHAPTER 3

Research Design

The need for use of a VPN service, as will be further elaborated by expert and non-experts throughout the qualitative research, is not always in line with a need of greater privacy. As we can see, the reasoning for adoption of a VPN service is varied and goes from wishing to navigate the internet more anonymously, but also to access region-blocked content from streaming platforms, bypass website blocks put forth by the Internet Service Provider or evade censorship. It would be an error to assume VPN services as a proxy for privacy, yet it is commonly its most advertised advantage.

The main goal of this study is to ascertain the willingness to pay for attributes within a VPN service, and to estimate how those change depending on various covariates, such as a person's gender, income, age, education and knowledge of online privacy and data protection (measured as privacy literature).

I highlight in this chapter the key hypotheses and research design, connecting them when relevant with the aforementioned literature review.

3.1. Hypotheses

The hypotheses ventured forward relate to the value individuals are willing to pay for a VPN service. Based on existing research, it is expected consumers will exhibit a preference towards free products over paid ones. According to Sombatruang et al. (2020), who conducted a study on the willingness to pay of a VPN app, with all other attribute levels being equal, the marginal probability (MP) of participants choosing a free app was 0.65 (United Kingdom) and 0.48 (Japan). Besides this, the qualitative research established conducted in this work showed certain individuals had a strong tendency towards free alternatives, and proved reluctant to pay for a VPN app. The matter of the probability of choosing a paid service over a free one warrants discussion and leads to the following null hypothesis:

Hypothesis 1. *Probability of choosing a free VPN app over a paid one, all attributes constant, is equal to zero.*

3.1.1. Attributes affecting willingness to pay for a VPN service

The previous research on willingness-to-pay for a VPN service that takes into account trade-offs between attributes is scarce. Sombatruang et al. (2020) found consumers taken from a UK and Japanese population were willing to pay on average 3.05 £ more if (all other attributes held equal) the baseline app customer rating would have a good review rating. This work focused on a mobile app, which is usually presented to consumers in a "playstore" scenario. It is therefore relevant to ascertain attributes that affect the willingness to pay for a VPN program in a computer setting, as opposed to a smartphone one. In this regard, I draw on existing research, as well as on qualitative research in order to develop the following null hypotheses directed related to the explanatory variables at hand:

Hypothesis 2. *Average monthly cost has no impact on the likelihood of choosing a VPN service*

Hypothesis 3. *Recommendations from friends/family have no impact on willingness to pay for a VPN service*

Hypothesis 4. *Rating from specialized magazines and consumers has no impact on willingness to pay for a VPN service*

Hypothesis 5. *Reduction of internet speed associated with the VPN service has no impact on willingness to pay for a VPN service*

Hypothesis 6. *Logging has no impact on willingness to pay for VPN service*

Hypothesis 7. *Number of countries with servers the VPN service offers has no impact on willingness to pay for the VPN service*

3.1.2. Covariates affecting willingness to pay for a VPN service

The main focus of the introduction of these covariates is to account for preference heterogeneity across different consumers. Covariates (or characteristics of the decision maker), unlike attribute levels, do not change during the course of the DCE, but rather they change between each participant. For the purpose of my analysis, I divided these covariates in two broads terms: a) demographic characteristics covariates and b) privacy literacy covariates.

Demographic characteristics covariate

These covariates will be analyzed, in order to characterize the sample accordingly and conclude appropriately on the results.

Privacy Concerns and Technological Literacy covariates

Hypothesis 8. *Level of privacy literacy does not have an impact on the preferences for a VPN service.*

Will be analyzed in order to ascertain impact of privacy literacy on valuation of attributes in a VPN service.

CHAPTER 4

Methodology

4.1. Introduction to DCE

In defining the methodology to use, I took into consideration the various methods used in the relevant privacy literature for measurement of the WTP/WTA, since all hypothesis require this estimation. There are two major methodologies used – traditional conjoint analysis (CA) and discrete choice experiments (DCE) - (Wagner et al., 2018) reviews several papers and considered these two methods as the primary ways of eliciting WTP/WTA through non-direct surveys in the field of privacy. The chosen method was the DCE, for as exposed by Louviere et al. (2010), conjoint analysis is generally inconsistent with economic demand theory. On the other hand, DCEs have a long-standing tradition in fields such as health care economics, transport economics, and environmental economics.

Discrete Choice Experiments take away from many different disciplines, from axiomatic conjoint measurement and information integration theory in psychology, random utility theory-based discrete choice models in economics, and discrete multivariate models for contingency tables and optimal experimental design in statistics (Hoyos, 2010). At its core, a discrete choice experiment involves presenting a respondents with different alternatives to pick from, with each alternative having different attributes and levels to force consumers to make trade-offs taking into account their preferences. By taking into account an attribute of cost or price, its possible to convert marginal utility into willingness to pay (WTP) estimates.

4.2. Design of a DCE

As summarized by Pérez-Troncoso (2020), a DCE usually has three parts, 1) a introduction, 2) the DCE itself and 3) respondent information. The main purpose of the introduction is to guide participants and instruct them on what they are responding and how to do it correctly. This step is of crucial importance in any DCE, but particularly in our case since for many participants a VPN service is a new program, whose attributes and advantages might be difficult to grasp.

As such, in the introduction to the experiment the following was explained: a) general scope of the work as part of a master thesis experiment, b) what type of experiment it is, and how it is structured, c) how a VPN functions, with a short description in a 2 minute video, d) how the data was to be handled and need of consent and e) demographic questionnaire. This introduction and demographic questionnaire was fine-tuned from pilot tests of the experiment, taking insights to try to assure participants came prepared for the taking of the DCE while trying to avoid subjecting them to an “information dump” and put too much cognitive strain before the DCE.

The second part contains the DCE itself. In this, correspondents were presented with a series of choices between hypothetical scenarios. These are known as choice sets, and contain 2 or 3 alternatives with varying attributes and levels.

The decision of attributes relevant to decision making is crucial to the strength of the model estimations and conclusions reached. After deciding the respective attributes, a number of decisions also need to be carried out in this part, such as: (1) the use of labeled or unlabeled alternatives; (2) number of attribute levels; (3) range of the attribute levels; (4) balance of the attribute-levels (Hoyos, 2010). Following the selection of attributes and levels, it is necessary to figure out what experimental design to apply. Experimental design is the sample from all possible combinations of attribute levels to construct the choice alternatives and assign them to choice sets. At this stage, we are presented with a basic problem: if we were to take all possible combinations of attributes and levels and construct from these a “full factorial design” consisting of all possible combinations of levels for all attributes, doing so would give a hard to wield design. It is standard to use a “fractional fractional” design, that is, a sample from the full factorial design to estimate merely the parameters of interest (usually main effects and as many interaction effects as possible) (Lancsar and Louviere, 2008).

The process of designing the DCE is a cyclical process, where 1) attributes and levels are defined; 2) experimental design is chosen; 3) questionnaire is developed; 4) sampling strategy is defined (Hoyos, 2010). The subsequent data collected is analyzed under specific econometric models that take into account different assumptions, which will be elaborated in the section of Choice of Models. The choice of model poses consequences to the design of the survey and subsequent analysis of the data, and the process of designing the DCE changes as new information is gathered from focus groups, interviews, and pilot studies (Hoyos, 2010).

4.3. Choice of attributes

The first stage of a DCE, following the structure lined up by Reed Johnson et al. (2013), involves the identification of the relevant attributes and corresponding levels relevant to the research question. In this case, this translates to identifying the attributes used to describe and differentiate each VPN program from its alternatives, as well as in what levels those attributes will vary. This is a stage with great importance, since erroneous/superfluous characteristics will undermine the validity of the DCE (Pérez-Troncoso, 2020). In general, this stage is conducted based on expert interviews and literature reviews (Pérez-Troncoso, 2020). Other authors such as Coast et al. (2012) recommend the use of focus groups as well.

I carried out the qualitative research on three general guidelines:

(1) Literature review.

(2) Interviews with experts in the field of informatics, security and data protection law.
(3) Focus Group with non-technical participants.

This approach differs from Sombatruang et al. (2020), whose work focused on willingness to pay for a VPN service in the form of a mobile application and conducted multiple interviews with non-technical participants to extract attributes. The reason for the different approach lies with the hypotheses set in this work that investigate the possible effects of privacy and technological literacy on the adoption of a VPN program. It was important to ensure the attributes on the alternatives given in the choice sets were also valued by experts alike, and we did not lose complete relevance of the attributes as privacy and technological literacy went up. If the attributes in the DCE are not relevant to those with higher degrees on privacy literacy, the choices made lose significance. Moreover, I wished to contrast the different reasoning and motivations by experts and non-experts for use of a VPN service, since it can be used for a variety of reasons and goals as exhibited by Khan et al. (2018).

I also kept the attribute recollection with non-technical participants since it is expected those will form the bulk of the study participants. This stage is important, so we can distinguish what attributes are valued in a Portuguese population (for instances, (Sombatruang et al., 2020) focused solely on UK and Japanese populations) as well as obtain additional insights that motivate the use of VPN programs, both in general and as a privacy enhancing tool.

4.3.1. 1) Literature review and product search

A search of the top 20 VPNs by webpage visits as well as a literature review was conducted to identify the broad scope of possible attributes and characteristics that were identified in VPN programs. This allowed us to construct a very expansive list that ranged from non-technical to technical attributes. The list was discussed at the end with experts and non-experts to spark any final discussion. We also took into account how much these attributes varied among different VPNs, in order to be able to focus our attributes in the choice alternatives presented as trade offs that are likely to happen in a online scenario. A search was also completed of the top 20 VPN firms by webpage visits to make sure the attributes were relevant.

The levels were also selected by means of a literature review and product search – especially the extensive work done by Khan et al. (2018) – since it was important to select levels that allowed both for variation between choice sets while keeping realism (and as such avoid setting participants up with unrealistic scenarios). This means that even if certain participants or experts considered a attribute and its levels relevant, if the vast majority of VPNs does not take into account these levels we would not be presenting participants with a real choice making scenario.

4.3.2. 2) Expert interviews

A series of online interview were conducted with various experts in the fields of Information Technology, Information Security and Data Protection Law. Five semi-structured interviews of 1-hour each were conducted with the purpose of achieving a clearer grasp of personal VPN use among experts and find insights concerning valued attributes and levels both in previously used VPN software, as well as key points one would look for in changing or adopting a new VPN service provider.

The interviews took place online in videocalls, and four were fully recorded and transcribed to extract key insights, with one expert preferring not to be recorded. For that expert, written notes were taken during the process and particularly at the end of the interview, clarifying additional points and ending remarks when necessary. Consent was obtained for each recording, with the terms highlighted and the recording eliminated before the date established with each correspondent.

The set of open-ended questions was summarily divided into two parts, the first of which focused on previous and current personal VPN service usage, as well as reasons, motivations, and perceived risks that identified the need (or not need) to use such software. The second part focused on attributes and characteristics that were valued in a personal VPN service, considering views, experiences, and other insights towards why a VPN service was chosen in the past, or what they consider important to differentiate among the appeal of different VPN services.

Besides these goals, correspondents were encouraged to elaborate on past experiences and to expand on how their trust in aspects such as security of public WiFis, trust of internet service providers, and other dangers that would justify the use of a personal VPN service. Awareness and importance of the GDPR was also questioned, for it is the largest and most expansive legislative effort for personal data protection so far and sets different boundaries and securities for how personal data can be used by firms established inside the European Union. These geographical differences were sought particularly since previous studies on willingness to pay for a VPN service are not always bound to Europe.

List of experts interviewed:

(1) Expert 1 - Security and informatics expert
(2) Expert 2 - Informatics and data protection expert
(3) Expert 3 - Informatics expert
(4) Expert 4 - Data protection law and privacy law expert
(5) Expert 5 - Informatics expert

The following key points were extracted from the analysis of the recordings:

- All experts agreed on a high risk of public WiFi's, and that consumers are dangerously unaware of such risks.
- All experts knew how a VPN functioned.

- Three experts used a personal VPN service, with all three knowing colleagues who did and who did not use one.
- Two experts did not use a personal VPN service, and showed either distrust of providers or did not give much thought to acquire one. They used other means to avoid public WiFi risk.
- All experts had considerable confidence in their Internet Service Provider to manage their personal data.
- Opinion about logging varied among experts, with two experts who used VPN services not particularly concerned about it, and one concerned with it. Of the two who did not use VPN services, one had a neutral approach while the other was distrustful of any VPN service for this exact reason.
- Certain users with very high technological literacy do not pay for a firm provided VPN program because they mount their own. Although to the average consumer this is likely an extreme case, two experts mentioned they knew coworkers who had completed such an endeavor.

Usage of personal VPN services among the interviewed

Experts had different opinions and habits when it came to the use of VPN services. While all experts exhibited experience in the use of VPN services for professional use, only three used them for personal use.

Some experts considered VPN use essential. For instances, the expert in the field of data protection law and privacy law has used two different VPN services for the past 7 years, having first used a standard VPN when recommended by an American coworker, and making a switch by a recommendation of another coworker. Currently, the expert uses an Antivirus that also offers VPN services, particularly since it provided a "2 in 1 package":

"I think the issue here in Portugal is people are not sensibilized towards how exposed they are in a network, and they are not aware of the risks of such exposure – the same happened with HIV, whoever had sexual relations with someone with HIV was not aware they were exposed and could transmit the disease without proper protection. It's the same logic. Here in Portugal, there is a big gap in knowledge of those themes, even in conversations with friends, when those themes are approached, they ask me "what can I do", and one of the things I say always is to use a VPN. It's a good practice" – Expert 4.

On the other hand, some experts might not even rely on a commercial VPN program. For instances, expert 5 and 1 mentioned a possibility of certain individuals mounting their own VPN service:

"Many professionals don't depend on a server (of a VPN service) and mount their own VPN – meaning, they mount a VPN server at home and regardless of service, use their house VPN to anonymize the traffic. They need only to pay for the server, pay for a computer, which can be a Raspberry Pi of 40 or 50 dollars, have access to the internet, and install required software." - Expert 1

Risks identified with online activity

All experts identified risks in online activity, with the most prevalent one being public WiFi risk. Distrust of the ISP (Internet Service Provider) on the other hand, despite being a common selling point in VPN programs, proved moot with all experts. The presence of the GDPR as a limiting force on the possibility of the ISP misusing, selling, or taking advantage of the individuals traffic or data in a way that goes against their interest remains a relevant contextual factor. Expert 5 puts it into words with *"The detail is I don't use a VPN service so much for the privacy, but for the security. This means, I'm not worried of my data being used by firms because there's legislation for that, but rather against ill-intentioned people. The VPN is important to protect data against identities that are not ruled by the GDPR, but rather are criminals."*.

The need for data protection and security when navigating on public WiFis and the existence of *man in the middle attacks* was deemed a significant threat, but different experts handled it in different manners, with Experts 2 and 3 not having ever used a personal VPN at all and instead relying on other methods to safeguard their data and/or privacy. Expert 2, for instances, instead of relying on VPN services would rather rely on using a *hotspot*, using mobile data from his Internet Service Provider, avoiding having to have his data through a server hosted by some VPN service, and to avoid connecting to a public WiFi connection in the first place.

A divisive topic was that of logging. The matter of how much client information the VPN provider can access, keep, and potentially misuse remains a point of contention, with different VPN services offering differing levels of assurance.

For Expert 4, the matter of a VPN service keeping logs is an unavoidable yet calculated risk (*"I'd rather be unprotected versus one agent - the VPN provider - than being unprotected against many, and as such I'll run the risk of having it (the VPN service provider) possibly having some access to personal information"*). The matter of logging can also serve as an important deciding factor to choose from different VPN providers, with Expert 1 denoting the following *"Some services I would not acquire, for matter of keeping logs"* . On the far end of the scale, Expert 2 exhibited distrust from personal VPN services, and preferred to rely on his ISP to provide him with mobile data whenever he needed, than to use a public WiFi or a personal VPN to secure the connection.

Expert 3 displayed the importance of logging for the average consumer who did not understood what it meant technically: *"Logging as a matter of control of level of service, people are more familiarized, for example, when they buy a service to a phone operator – people are more sensitive to the matter of the contract. Basically, how they check if the contract is being kept"*. The matter is logging concerns fundamentally a more ample characteristic that is hard to verify in VPN providers, and which is also true for other services, which is transparency.

Other ways to mitigate risks

Possibly as technological and privacy literacy goes up, since as previously seen individuals might be more likely to choose alternative methods to protect their privacy that do not involve purchasing a personal VPN service. This should be considered – willingness to pay for a VPN service could be low not because the risks are perceived as insignificant, but because an expert would rather execute that service himself or seek alternative ways to obtain anonymity. Nevertheless, preferences explored in these interviews could serve as indicators of different willingness to pay estimates among different privacy literacy categories.

Attributes identified by the experts as significant:

(1) Location of servers (mentioned by nearly all of them – one found it not very significant unless in China or Russia).

(2) Location of VPN service firm (most common division was Europe – Out of Europe).

(3) Data protection laws of the country the VPN service firm operates on (the affirmation "nothing that comes from the United States or China" was common).

(4) Whether the service can be "audited" – that is, there's a way to check if the terms of the contract between VPN user and VPN service firm are being upheld (logging).

(5) Price.

(6) Reviews from a specialized magazines (two experts mentioned to give a high importance on consumer defense magazines e.g. "DECO Proteste").

(7) Advice and opinion from friends and specialists (mentioned by 2/5 experts, or 2/3 who used a personal VPN).

(8) Effect of VPN service on connection speed - Expert 1 highlighted the importance of this attribute for both technical and non-technical people. Particularly, since often free VPN services put a cap on speed, offering a premium version to remove said cap could be enticing.

(9) Possibility to choose the server to connect to from a list (one expert noted they liked their service because it allowed them to choose a server – and the correspondent "speed" of the server would be attached to the choice; another mentioned they also did not consider the speed component to be very relevant because their VPN provider allowed them to choose the server with the highest speed).

(10) Promotions in price.

(11) Possibility to pay only once for the service.

4.3.3. 3) Focus group

A focus group with 8 participants was conducted. This focus group was chosen from participants from a non-technical background, from various universities and areas of study and work. They were all young, with a median age of 24 years and generally classified as WEIRD (Western, educated, industrialized, rich and democratic). Half of the participants were female and half male, to ensure gender differences were accounted for in focus group. The interviews took place online through Zoom. There was one moderator, and since one participant did not concede to being recorded, one independent individual who did not participate and whose sole role was to take notes. These notes were later compared with the moderator own notes to ascertain the key insights. The set of open-ended questions was structured the following way:

In the first 30 minutes, the participants were asked a series of ice breaker questions, mostly focused on their online experiences with VPNs as well as use of streaming services. The reason for such questions had to do with VPN services popularity as a tool to access foreign region-blocked content.

Following this discussion, there I explained the general uses, goals, and way of functioning of a VPN. The possibility of using a VPN to access foreign region-blocked content was addressed, as well as privacy and security benefits such software could provide.

1. Usage of and knowledge of personal VPN services among the interviewed
   - Only P1 had used a VPN for personal use before, while the other participants were still unaccustomed to this type of product from a personal setting standpoint. However, all the participants had heard of VPNs being used in a professional setting.
   - P1 had used a free VPN service (that boosted a premium version he did not use) in order to bypass website blocks his ISP had set.
   - Two participants (P1 and P4) had seen ads of this type of software before.
   - P4 and P8 mentioned knowing friends who used VPNs for personal use. P4 described his friend who used the VPN for personal use as “not tech savy” and his main purpose was to access geo-blocked content. P8 mentioned having several friends, who worked in the IT field and used VPNs for a variety of reasons such as accessing geo-blocked content, accessing blocked websites, and overall online privacy and protection. Despite P8 never having found a compelling use to install a VPN, he believed VPN use to be widespread within his circle of friends.
2. Desire to watch geo-blocked content:
   - P1 mentioned he knew people who had an interest in consuming certain media as soon as it was released in other countries, and for such the use of a VPN could be deemed useful. For himself he felt the case did not apply.
   - All users used streaming services, with two having admitted to the use of piracy software in the past to access content.
   - P2 mentioned when he did not have subscriptions for a service, he resorted to piracy – this was a sentiment echoed by other participants. However, commodity was important for this participant, and he would prefer to access the content over a streaming service. If it was a significant gap in quality and commodity, he would consider purchasing a VPN service.

- P3 mentioned buffering speed and experience with lack of quality with piracy services.

3. Privacy
   - P4 mentioned he felt search engines kept data but did not consider it a strong concern. He felt during certain searches he wished for the possibility to keep more privacy.
   - P2 mentioned concerns when planning trips, and whether cookies and other aspects could affect the search results.
   - P1 mentioned when it comes to payments, specially in bank apps, he assumed they were secure. The use of a Portuguese service called MBway to create virtual cards took away any danger he felt towards others online payments.
4. Attributes identified as important in such a service
   - P1 considered the presence of a "free alternative" inside the same VPN as relevant, being a way to try the product.
   - P1 considered the interface to be relevant. P7 in this topic added the possibility to have it as a Chorme/Browser extension would be very appreciated. When P7 was questioned as follow up on whether that service would be considered desirable as part of a "anti-virus + VPN" pack, she mentioned it seemed more convenient than as a separate program.
   - P1 has used a free VPN before to download content from a blocked website, and he felt speed was not important to him – since he never used it for streaming, but rather to download material to see later. For him, a free service even with subpar speed was enough.
   - P1 and P4 mentioned they had seen ads for VPNs by youtubers they followed.
   - Amount of information asked by the VPN provider – P6 mentioned when too much information is asked, she gives up purchasing the service. The method of payment is also relevant as well.
   - Opinions of friends was mentioned as a relevant factor – for instances, P6 mentioned how when she went to purchase a computer, she consulted an opinion of a friend in the area.
   - P8 also mentioned the importance of an easy-to-use interface, how quick the program functions. He believed the use of a VPN for privacy or security was only for a specific kind of privacy concerned person, who was a minority of the population.
   - Promotions in price was mentioned by P8 as being a significant aspect of why his friends chose their VPNs.

4.3.4. Analysis of interviews and attribute identification

On the matter of valued attributes, we compared the insights obtained in the literature review with the attributes identified by experts and non-experts as significant. The goal was to discern the attributes that were relevant for the decision making from those that actually differed between different VPNs. For example, certain attributes that experts who have used a VPN in the past unanimously considered important - such as possibility to choose server and location of server - were discarded for the vast majority of VPN services since they already allow for this functionality. As such, it's unlikely they would form the core of the decision making process but rather serve as a means to exclude outliers.

On the other hand, factors such as an easy to use interface, policies and information regarding transparency subjects (such as logging), origin and strength of recommendations, type of program (as a stand alone program, a feature of an anti-virus, or a browser add on), location of countries with servers, rating or impact on internet speed are offered in varying combinations throughout the market, and all were considered for analysis. The wide array of attributes identified in the focus group and interviews with experts will be juxtaposed with the relevant literature to arrive at the final list of attributes, and I will provide reasoning as to why certain attributes were included or excluded.

4.3.5. Included attributes

Monthly average cost

Price is the traditional attribute common to most discrete choice experiments, especially those with a focus on computing willingness to pay for a product or service. Besides serving as the strongest point of comparison between VPNs, it is the introduction of a price attribute that allows the researcher to express trade-offs not only in terms of utility, but of cost and as such determine how variations in attributes are valued monetarily.

I chose to focus on the average monthly cost for such services. The wording is relevant in the sense we are looking for the average cost a consumer will pay per month, regardless of subscription type (paid monthly, quarterly, yearly or other combination). This stems from the consideration subscription models vary substantially within the same VPN and across different VPNs as well. Work by Khan et al. (2018) found a monthly subscription to be the most common, however many VPNs offered quarterly, 6 months, and annual subscriptions, with some going as far as offering a "life-time" offer. This means individuals can pay a higher amount a single time, and as such obtain a "bulk discount". I chose to focus solely on the average monthly cost (ignoring subscription type) to avoid both setting too large of a cognitive burden on participants, but also to avoid multicollinearity (if monthly cost and subscription model are both treated as independent variables, it is likely they are highly correlated. This will undermine the statistical significance of the independent variables themselves).

Monthly Average Cost - Levels

This decision was left out of the focus group and expert interviews, since I wanted to use price points that reflected the present VPN environment, while allowing for participants to engage on the relevant trade-offs. Just in terms of monthly payments, three points were set by Khan et al. (2018), with a minimum of 99 cents ($), average of 10.10 ($) and maximum of 30 ($). Since Khan (2018) did not specify which prices were the most frequent, I supplemented the analysis by looking into the monthly cost of the top 20 VPNs currently in the market by webpage visits, both offered in a monthly and yearly subscriptions. I did not consider the 30 $ price point, since it likely embodies an extreme that is to be applied to a small subset of consumers. By choosing to add a very high level to the questionnaire, we risk obtaining many cases where the high-cost alternative is rejected near unanimously, and as such not provide us with valuable information regarding trade-offs between prices (levels closer in range could then provide more valuable information). The following price levels were included:

- **Level 1** - Free level was included given about 40 % of the top 20 VPNs by page view offer such a alternative. Participants in the focus group also, unlike the expert interview, displayed a strong preference towards free VPN services. Offering a free alternative is paramount to evaluate *Hypothesis 1*, since probability of choosing a free alternative over a paid one requires such free alternative to be presented to participants in the first place.
- **Level 2** - I included the price point 3,99 € since this is close to the average of the cheapest offered price point when we take into account payments in yearly subscription modes.
- **Level 3** - Chosen since it compromises a mid point between level 2 and level 4. 18 % of the top 20 VPNs analyzed offer a monthly price point between 6.99 and 9.99. This value is also close to the max monthly rate for yearly paid subscription.
- **Level 4** - 54 % of monthly prices are between 9,99 € and 11,99 € (with 82 % between 7,99 € and 11,99 €). Another methodological advantage of choosing 11,99 € was to keep values between levels equal.

Recommendation of friends/family

The attribute "Recommended by friends/family" was frequently mentioned among participants in the focus group. Participants often mentioned the difficulty in choosing software and hardware without the help of knowledgeable friends/family. Parallels to other products and services, such as purchase of a computer or of a Anti-virus program were drawn.

Recommendation of friends/family - Levels

Two levels were considered, yes and no, with no signifying no recommendation was made (and not that someone recommended against it).

The reasoning for just two levels lies in maximizing the trade-offs of extra options versus greater model and choice set complexity. Although VPNs publicized by VPN comparing websites could be seen as a form of recommendation, the attribute "Rating" provides weight of those recommendations in way of a star rating, with a greater degree of granularity.

Rating

I considered rating for a couple of reasons. First off, it was pointed in literature by Sombatruang et al. (2020) as an important attribute, and according to (Ramesh et al., 2022), users seem to lean towards search engines and recommendation sites (61,1 % and 56,5 %, respectively) over methods such as word of mouth. VPN comparison websites provide its own industry of advice, and often share much of the transparency issues VPN services do. Testing how users value those ratings could provide some clarity into the pulling effect these websites have towards a certain VPN service.

Secondly, experts identified the importance of opinion of consumer defense magazines. Although consumers in the focus group did not mention expert rating *per se* specifically, they also placed emphasis on qualified user opinions. User rating could serve as important attribute, together with magazine and website recommendations rating, to guide consumers.

The main topic of discussion here was to either include rating as a conjoint attribute comprising of both expert rating and user rating, or whether to desegregate both in different attributes. Ultimately, I decided to group both ratings into one, losing differences between both in the process, but avoiding subjecting the participants to a too high cognitive load. Giving them 7 attributes to trade off would likely undermine the trade-offs with the other 6, also relevant, attributes.

Rating - Levels

I chose two interval levels, of 3 to 4 stars and 4 to 5 stars. Lower quality ratings were not chosen, since they would add complexity to the DCE that could be better used for other attributes. Comparison websites, from my own research, rarely include VPNs services with ratings below 3 stars, serving this as a bare “minimum”.

Reduction in internet speed

Using a VPN service has direct consequences to internet speed, since this means internet traffic is going through the VPN server, which adds a extra step in the process. This insight was highlighted in the expert interviews and focus groups. While the matter of reduction in internet speed is relevant to both experts and non-experts, experts have a deeper knowledge of what causes it, and what factors affect it. In essence, it is not possible to know the exact effect a certain VPN service has on internet speed *a priori*, but only after experiencing said VPN service. The reason is the effect on internet speed will depend on what server the person connects to, with closer and faster servers offering different effects on internet speed, as well as someone’s geographical location and own internet characteristics (for example, if someone is experiencing throttling[1] from his internet provider, a VPN service can help alleviate it). Additional factors apply to each VPN service, with VPN servers that are “overloaded” with users, as often is the case for

[1] Throttling is defined as the intentional slowing or speeding of an internet service by an Internet service provider (ISP). It is a reactive measure employed in communication networks to regulate network traffic and minimize bandwidth congestion

free services, experiencing bigger impacts. The consequence is most VPN firms offering their services are vague on the terms of their effect on internet speed - often marketing themselves as the “VPN with the highest speed connection” - dodging the buzzword reduction/decrease entirely. However, firms often can and will place caps on internet speed when it comes to free versions. This means they often specify those limits, as a way to draw users to premium versions of their products.

Reduction in internet speed - Levels

I considered three ranges. The reasoning is meant to capture the scope of experience VPN users when they adopt the service. These vary in degree of intensity, and are capped at moderate to severe impact, since I considered above this point users are unlikely to opt for a VPN service. The use of the service only makes sense if it allows users to achieve their goal - be it browse the internet with greater security and anonymity, or watch geo-blocked streaming content - and if VPN service does not allow them to do that enjoyably, it loses significance as a service.

Logging

As noted in the experts interviews, logging is a divisive and important topic to consider when choosing VPN service. This attribute fundamentally concerns a problematic of the current VPN service market - which is that of transparency. For a service that bolsters anonymity and security, VPN firms are often illusive in what data they are keeping of their users, with Khan et al. (2018) estimating up to 25% of the 200 VPNs analyzed did not have a link to their privacy policy, and only 45 of those 200 claiming a ”no-logs” policy. As choosing a VPN service for privacy related reasons often equates trusting the VPN firm to handle the users internet traffic with discretion, this attribute is of crucial relevance to the DCE.

Logging - Levels

I considered three levels, each embodying a increasing degree of transparency. I included the first two levels mainly to distinguish between preferences for a service who did not refer the data it keeps, and a firm who does - in order to infer preferences of participants towards the minority of VPNs who address this issue. The last level is narrower case, but that has gained popularity as of late. Firms such as Nord VPN and ExpressVPN hired firms to audit their privacy policies over the last years, in order to ensure their policies are compliant with the data they actually keep of their users.

Number of countries with servers

Number of countries serves to quantify two major points of the expert interviews - location of VPN servers, and number of servers. The issue of data protection law, connection speed, and access to geo-blocked content are indirectly represented in this attribute. Experts showed a greater sensitivity to these matters than the participants of the focus group, and evaluating the relative coefficient of this attribute among participants with different literacy levels poses a important analysis.

Number of countries with servers - Levels

To determine the levels, I conducted a search of the top 20 VPNs by webpage visits. The first level captures 18 % of the offers below 50 countries, while the second and third level capture both 41 % of the population in analysis. The reasoning for the different percentage in the first level, is it allows to equalize the second and third levels in percentage of VPN offerings. It is also likely the first level is underrepresented in the top 20 VPNs, and as we go further down in popularity, number of countries offered will likely decrease. There is also incentive towards adding a first level that is low enough to capture preferences. Ranges are kept similar amongst levels as is recommended in the relevant literature.

4.3.6. Excluded attributes

I found important to add this section, to justify why certain characteristics that could be seen as attributes were excluded, despite mentions in focus groups, expert interviews, or literature research.

Interface - Mentioned in focus group

This attribute was removed since it is not directly testable, - despite it being important and mentioned in the focus group, we considered the VPNs being offered to participants as being equal in user friendliness - which is the case for most popular VPN services.

Headquarters of VPN location - Mentioned by experts

The matter of headquarters, despite relevant, seems to serve more as a reason to excluded outliers than gauge preferences. Most experts had aversion to certain locations (such as the United States or China), which are often countries with either lax laws on data protection or oppressive government surveillance. Nevertheless, participants did not show a strong preference towards headquarters in specific countries.

Number of servers - Mentioned by experts

The attribute of number of countries with servers servers the purpose of this topic.

Location of servers - Mentioned by experts

The attribute of number of countries with servers servers the purpose of this topic. Moreover, with a increase in number of countries, the chances of being able to choose a server in a preferred location increases.

Type of program - Mentioned in focus group

Was excluded since type of program is not exactly a attribute of a program. Often, being offered as a package of a anti-virus does not tell anything of the VPN itself - it has more to do with the way it is marketed towards consumers. By itself, the VPN and the anti-virus are still two separate programs - so even if it is more likely to get people to acquire that particular service, they are not engaging on attribute trade offs, but rather being herded by the antivirus. Browser VPNs, on the other hand, since they only work while using the browser, also add unnecessary complexity to the model. This is on account of not redirecting all traffic to the VPN server, but rather only the browser traffic, constituting therefore a similar, yet different service.

Number of downloads - Mentioned by (Sombatruang et al., 2020)

Excluded since it is not directly observable in a VPN program for computer use. The issue of number of downloads is something directly comparable among different VPN providers in an app store, but in an online setting where each VPN service has their own website the matter is complicated

We can see the full list of chosen attributes and respective levels down below:

Table 4.1. Attributes and levels selected for the DCE

Puchase of a VPN Program	
Attribute	Levels
Average Monthly Cost	Free 3.99 € 7.99 € 11.99 €
Recommended by friends/family	Yes No
Rating	Between 3 and 4 stars Between 4 and 5 stars
Reduction in internet speed	None to little reduction on internet speed Little to moderate reduction on internet speed Moderate to severe reduction on internet speed
Logging	Does not mention what data it keeps from user Mentions what data it keeps from user in its website or privacy policy Mentions what data it keeps from user in its website or privacy policy which was confirmed by an independent firm
Number of countries with servers	Less than 50 countries Between 50 and 80 countries More than 80 countries

4.4. Design of the choice sets

Based on the previous discussion of relevant attributes and levels, I constructed 12 choice sets using a fractional factorial design, taking into account three properties: Orthogonality, balance and minimal overlap among attributes. Designing an efficient experimental design is a subject that has evolved through out the literature, and often can be a challenging procedure. There is often a trade-off between orthogonality and balance, with measures to determine efficiency such as use of a measure called D-effiency. I focused on maintaining the properties of orthogonality, balance and minimal overlap, while limiting the number of choice sets each participant has to fill in - the literature states 18 as a practical limit up until the point boredom starts to set in (Mangham et al., 2008). With this, we are able to achieve a design that is robust and allows us to gather data efficiently to later construct our model. I resorted to the program JMP since it allowed to construct these sets taking into account our specifications.

No choice alternative

For each choice set, I did not include an additional, third option to choose neither of the first two alternatives. The reason for not providing this option has to do with a few insights. The first of which is to avoid participants giving up on effectively considering the trade-offs, and selecting the no-choice option since it provides the least cognitive burden. Secondly, adding this option would result in a larger complexity of the models to be estimated, and since each "no choice alternative" presents no information on trade offs, it would effectively require either a much larger sample size, or a larger increase of choice sets presented to participants. Such sample size is not obtainable in the scope of this work, and increasing choice sets could present serious hurdles in terms of participant attention and focus (especially since we are already including two questionnaires in addition to the basic DCE). Nevertheless, to add a level of control over the quality of the choices participants took, three questions were included after the DCE, with the relevant possible answers below:

Question 1. For you, was it difficult to choose between the presented options?

- (very hard, hard, neither hard nor easy, easy, very easy)

Question 2. In the choices you made previously, did you take into account all attributes (average monthly cost, recommendations of friends/family, rating, reduction on internet speed, logging and number of countries with servers)?

- (yes, no)

Question 3. If no, what attributes did you value the most (select at most 3)?

- (average monthly cost, recommendations of friends/family, rating, reduction on internet speed, logging, number of countries with servers)

4.5. Questionnaire

Two questionnaires were developed, with the goal to gather information on the demographic characteristics covariates and on the privacy literacy covariates.

Demographic Questionnaire

For development of the demographic characteristics questionnaire, participants were asked to fill their gender, age, education level and monthly individual income.

Privacy Literacy Questionnaire and Literacy Score

To evaluate privacy literacy covariates, I developed a questionnaire to measure privacy concerns and technological literacy, based on the work done by Trepte et al. (2015), who identified 5 key dimensions for online privacy literacy. Nearly all dimensions were kept intact in their meaning, with the exception of dimensions 2 (that evaluated knowledge of data protection rights in Germany) and 4 (that evaluated knowledge European directives regarding data protection), that were joined together since Trepte et al. (2015) precedes the GDPR, and the current paradigm in European data law does not justify evaluating rights in Portugal separately from the general European law.

The five dimensions to capture in the questionnaire were the following:

- (D1) Dimension 1: Pratices of institutions and online service provider

  (D1.1) Knowledge: Internet Service Provider (ISP) practices

  (D1.2) Knowledge: Online service providers practices
- (D2) Dimension 2: Knowledge about the technical aspects of online privacy and data protection

  (D2.1) Knowledge: Privacy Enhancing Technologies (PETs)

  (D2.2) Knowledge: How internet functions/technical infrastructure
- (D3) Dimension 3: Knowledge about the laws and legal aspects of data protection in Portugal and the European Union

  (D3.1) Knowledge: GDPR (data protection and privacy legislation)
- (D4) Dimension 4: Data protection strategies for individual self-control

The privacy literacy questionnaire itself was composed of three parts. The first two parts represent a self-assessment, where participants were asked to share their perceived knowledge on a multitude of topics, relevant to each dimension. While the first part encompasses dimension 1,2 and 3, the second part focuses solely on evaluating self-assessed privacy protecting habits on dimension 4. The reasoning was that unlike with other dimensions where the general question could be understood, the matter of whether a individual employs strategies to protect their online privacy is a vague topic, and a participant could employ strategies but fail to recognize them as such. To add to this point, the levels are not a matter of extensive or poor knowledge, but rather a matter of how often the participant employs these strategies (always, frequently, sometimes, rarely or never).

The last part of the questionnaire focused on a set of true and false questions with the goal of validating previous self-assessed knowledge. All dimensions were covered, and this served to counter the Dunning-Kruger effect, where people with low privacy literacy levels could be overestimating their knowledge of the subject.

The version in English and Portuguese can be found in list in the appendix B.3. and figures on B.4., respectively.

4.6. Categorization of participants

Based on the results of the privacy literacy questionnaire, each participant result will be converted into a score by means of a POMP score (Cohen et al., 1999). A POMP score is a score that follows the formula:

$$[(observed - minimum)/(maximum - minimum)]100 \tag{4.1}$$

Where observed = the observed score for a single case,
minimum = the minimum possible score on the scale, and
maximum = the maximum possible score on the scale.

A minimum score will be established to classify participants into either a “Advanced Privacy Literacy” or “Basic privacy literacy” category.

After this preliminary categorization, participants with a advanced level of privacy literacy will be subjected to a validation threshold, calculated in terms of percentage of correctly answered questions in the last part of the questionnaire. This serves to counter a possible dunning-kruger effect, where participants have disparaging differences between self-assessed privacy literacy levels and actual privacy literacy levels. The population will then be sorted into subgroups, and results will be compared between models estimated.

4.7. Deployment of DCE

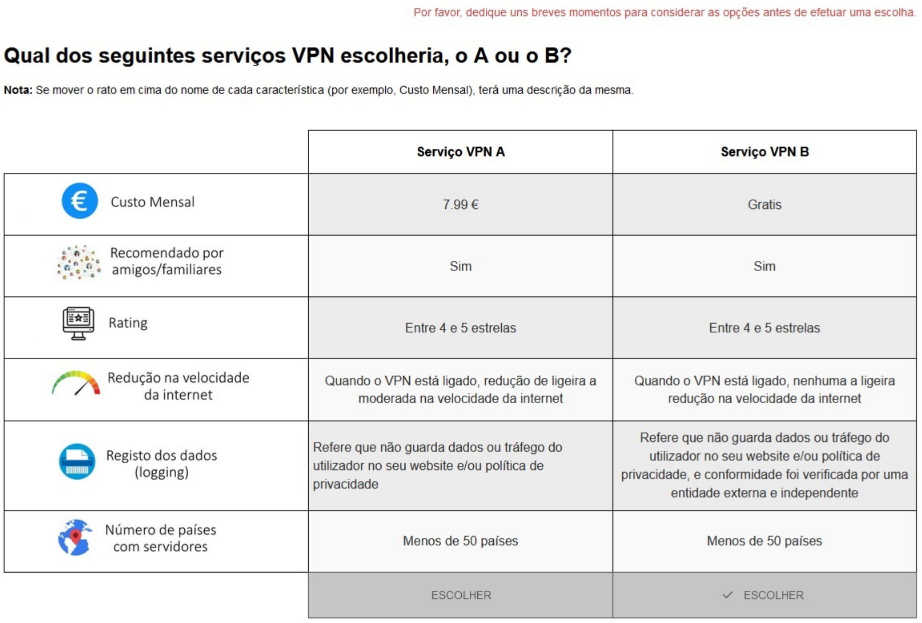
Having already established sufficient information for design of the DCE, it is necessary to embed the experiment in a interface with which participants can interact and that can be easily distributed.

After the choice sets were designed, they were exported from JMP to a online platform called Conjointly. This platform allowed to construct the following aforementioned parts of the experiment:

(1) An introduction
(2) Consent form for the data collected
(3) A demographic questionnaire
(4) The Discrete Choice Experiment, where participants had to choose between a series of two alternative VPN programs, A and B.
(5) A privacy literacy questionnaire

The platform also allowed to set a minimum timer of 3 seconds per choice, unblocking the “choice” option only after those seconds had passed, to set a block to participants who tried to skip randomly through the DCE.

The experiment was distributed in a link to a sample of participants, collected through personal circles, social network websites, and university circles. Despite efforts to make this distribution as random as possible, participants were likely college educated, from the center of Lisbon, and with a relative degree of privacy literacy. The relevant demographic characteristics are set in the appendix table B.1, and it is important to be cautious with extrapolating findings to the general population, since this experiment's data and key findings are likely to differ to groups with different demographic make ups. The full set of prints of the deployed experiment is available in figure D.1 available in the appendix. We can see in figure 4.1 an example of a choice designed in Conjointly.



Por favor, dedique uns breves momentos para considerar as opções antes de efetuar uma escolha.

Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
Custo Mensal	7.99 €	Gratis
Recomendado por amigos/familiares	Sim	Sim
Rating	Entre 4 e 5 estrelas	Entre 4 e 5 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente
Número de países com servidores	Menos de 50 países	Menos de 50 países
	ESCOLHER	✓ ESCOLHER

FIGURE 4.1. Example of a choice designed in Conjointly

4.8. Choice of Models

I estimated a series of econometric models to explore plausible explanations for the series of choices obtained while carrying out the experiment, and test the relevant hypothesis set in the Research Design chapter.

After briefly explaining the basis of the random utility model, the first model to run is the multinomial logit model, which is not only the most used model in discrete choice experiments but also represents the most constricted model in terms of hypothesis and assumptions that must be followed. A secondary model that takes into account the heterogeneity of the population was deployed, the mixed logit model, in a effort to relax strict hypotheses such as the IID (independent and identically distributed errors) and allowing for participants preferences to vary from one another.

4.8.1. Random utility model

Discrete choice experiments are an attribute-based stared preference valuation technique. They draw on Lancaster characteristics of demand, where consumers have preferences for and derive utility from underlying attributes, rather than from goods *per se*. Choices made in the DCE are analyzed using Random Utility Theory (RUT), which proposes that utility U_l can be decomposed into an explainable or systematic component V_l and a non-explainable or random component ϵ_l:

$$U_l = V_l + \epsilon_l \tag{4.2}$$

Economists see this error component as unobservable or unobserved attributes, unobserved preference variation, specification error and/or measurement error. Moreover as j are not observed, choices can only be modeled in terms of probabilities from the researcher point of view. Alternative l is therefore chosen if $\epsilon_l < (V_l - V_j) + \epsilon_{jl} \forall l \neq j$. The probability of choosing this alternative then is

$$\mathrm{P}\left(\epsilon_1 < V_l - V_1 + \epsilon_l, \epsilon_2 < V_l - V_2 + \epsilon_l, \ldots, \epsilon_J < V_l - V_J + \epsilon_l\right) \tag{4.3}$$

Denoting F_{-l} the cumulative density function of all the ϵs except ϵ_l, this probability is:

$$\left(\mathrm{P}_l \mid \epsilon_l\right) = F_{-l}\left(V_l - V_1 + \epsilon_l, \ldots, V_l - V_J + \epsilon_l\right). \tag{4.4}$$

The unconditional probability is obtained by integrating out the conditional probability using the marginal density of ϵ_l, denoted f_l:

$$\mathrm{P}_l = \int F_{-l}\left(V_l - V_1 + \epsilon_l, \ldots, V_l - V_J) + \epsilon_l\right) f_l\left(\epsilon_l\right) d\epsilon_l. \tag{4.5}$$

4.8.2. Multinomial logit model

The multinomial logit model represents a special case of the model developed in the previous section, and its basic formulation is the following:

$$P_l = \frac{e^{V_l}}{\sum_{j=1}^{J} e^{V_j}} \tag{4.6}$$

This model is based on three key hypotheses. They are i) independence of the errors, ii) the assumption that each ϵ follows a Gumbell distribution and iii) the assumption that the errors are identically distributed. The last assumption serves essentially as homoscedasticity hypothesisis.

IIA property of the Multinomial Logit Model

IIA (independence of irrelevant alternatives) relies on the hypothesis that the errors are identical and independent. This assumption states characteristics of one particular choice alternative do not impact the relative probabilities of choosing other alternatives. In our particular scenario, if this is valid, this means that participants are not being presented with alternatives (VPN services) that are very similar among each other, and probabilities of choosing one VPN over another are not being overestimated/underestimated, on account of some VPN services being indistinguishable among each other. This is particularly significant for our work in the sense that if relevant attributes are being omitted, this hypothesis is violated in the basic multinomial model.

Interpretation of results

Unlike a linear model, coefficients represent marginal utilities that can not be interpreted directly. It is by transforming these coefficients we are able to obtain meaningful results. This is expressed in the calculation of marginal rates of substitution, that can be obtained with the ratio of coefficients:

$$-\left.\frac{dx_2}{dx_1}\right|_{dV=0} = \frac{\beta_1}{\beta_2} \tag{4.7}$$

In our case, we can take the average monthly cost of a VPN service (x_2), and find out the marginal ratios of substitution by dividing the associated β_2 with the β_1. It is with these transformations we are able to obtain willingness to pay estimates.

Another alternative to find willingness to pay estimates is to model the multinomial model in the WTP-space as opposed to the preference space (Helveston, 2022). For example, consider the following function where utility for a VPN is given by the following model:

$$u_j = \alpha p_j + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3} + \beta_4 x_{j4} + \varepsilon_j \tag{4.8}$$

We can estimate this model in the preference-space as follows:

$$\begin{aligned} u_j = \alpha p_j + \beta_1 x_j^{\text{recomSim}} + \beta_2 x_j^{\text{rating4 to 5}} + \beta_3 x_j^{\text{speedmod a sev}} + \beta_4 x_j^{\text{speed lig to mod}} \\ + \beta_5 x_j^{\text{logref}} + \beta_6 x_j^{\text{logref e audit}} + \beta_7 x_j^{\text{countryless50}} + \beta_8 x_j^{\text{countryplus80}} + \varepsilon_j \end{aligned} \tag{4.9}$$

We can estimate this same model in the WTP-space as follows:

$$\begin{aligned} u_j =& \lambda \Bigg(\omega_1 x_j^{\text{recomSim}} + \omega_2 x_j^{\text{rating4 to 5}} + \omega_3 x_j^{\text{speedmod a sev}} + \omega_4 x_j^{\text{speed lig to mod}} \\ & + \omega_5 x_j^{\text{logref}} + \omega_6 x_j^{\text{logref e audit}} + \omega_7 x_j^{\text{countryless50}} + \omega_8 x_j^{\text{countryplus80}} - p_j \Bigg) + \varepsilon_j \end{aligned} \tag{4.10}$$

We can take ω as being established in monetary units, with λ being the scale parameter.

4.8.3. Mixed logit model

The IIA/IID assumption underlying the multinomial logit also assumes survey respondents have the same preferences and/or that unobserved variation around those preferences are similar. Models such as the mixed multinomial (MMNL) allow β to vary randomly across individuals, by considering the β as random draws from a distribution whose parameters are estimated. This relaxes the IIA hypothesis, and translates into a model that takes the heterogeneity of the population into account.

The combination of random-attribute coefficients and extreme-value (Gumbell) distributed errors result in a complex K-dimensional integral which cannot easily be solved analytically, but is rather approximated with a simulations with a finite number of draws:

$$P_{il} = \mathrm{E}\left(P_{il} \mid \beta_i\right) = \int_{\beta_1} \int_{\beta_2} \dots \int_{\beta_K} \left(P_{il} \mid \beta\right) f(\beta, \theta) d\beta_1 d\beta_2 \dots d\beta_K \tag{4.11}$$

CHAPTER 5

Empirical Results

The experiment was released officially on 5th of November, and for the following 2 weeks, 84 participants took part. After the responses were gathered, they were exported to Excel through the platform. The analysis of demographic characteristics and privacy scores of each individual was conducted on Excel, but the statistical and econometric analysis of the models was programmed in R, with the use of two major packages:

- mlogit package (Croissant, 2020) - Package to analyze R multinomial logit models and tests.
- logitr package (Helveston, 2022) - A recent, faster package that generates multinomial and mixed logit models in two spaces: Preference Space and Willingness to Pay Space.

5.1. Analysis of participants - demographic and privacy literacy

I analyzed participants' data with the goal of better understanding their demographic characteristics and privacy level, compounded by means of a score using the methodology explained in section 4.5.

Demographic Characteristics

I obtained a near equal amount of male and female correspondents, with 50 % male, 49 % female, and 1 % who preferred not to answer. The most common range for age was between 18 to 24, with 39 % of correspondents within this age bracket, followed by 25 to 34 with 23 % and 45-54 with 14%. The population under analysis is young, with 73 % of participants with an age below 44 (below the average age in Portugal of 46 years).

Individual income was below 500 € a month for 26 % of participants, with participants with incomes between 1000 € and 1500 € representing 23% and 26 % of participants displaying individual income above 2000 € a month. When it comes to level of education, 86 % of the sample had a bachelor's or above, with 14 % holding a doctorate degree. This contrasts heavily with the 14 % who only have a high school education or below, although it should be noted only one participant was below 18, which can justify this difference.

As such, this sample displays a above average level of education, is distributed equally in gender, with differing levels of income and distributed unevenly across all age brackets, with a particular focus on younger participants. The methods of distributing the quiz (through social networks, university networks, and public online forums) are likely the major factor contributing to this. The full analysis of demographic characteristics is available in the appendix in table B.1.

Privacy Literacy and Privacy Literacy Score

For evaluating the privacy literacy level of participants, I calculated POMP scores for each participant across the various dimensions identified by Trepte et al. (2015), measured in the self-assessment questionnaire. Participants were preliminarily categorized into two categories, "Basic Literacy", if their score was below 40 %, and "Advanced Literacy" if their score was above 40 %. The reasoning behind choosing 40 % lies both in the fact the average self-assessed privacy score was 41 %, and that choosing a higher minimum score for advanced privacy literacy would yield very few responses and drastically unequal sample sizes, which would harm any conclusion taken from model comparison. I divided across merely two categories since I valued lower standard deviations and greater significance of results over less meaningful results across a wider array of privacy literacy categories.

After this preliminary categorization, participants with an advanced level of privacy literacy were subjected to a validation threshold, calculated in terms of the percentage of correctly answered questions in the last part of the questionnaire, which comprised of a sequence of true or false questions.

In this regard, since the average quiz scoring was 63 %, in order to accept the participant in the category "Advanced Privacy Literacy", they had to score at least above 63 %, displaying therefore above average knowledge. After running this analysis, 11 participants were passed to basic literacy, since they did not pass the validation test and scored below 63 %. This yields a final score of 52 basic literacy participants, and 32 advanced literacy participants. As with other aspects of this study, this analysis would benefit greatly from a larger sample size. This would allow setting a larger amount of categories to capture different privacy literacy categories and as such examine how different brackets of higher privacy literacy scores impact valuations in a VPN service.

5.2. General

Given each participant answered 12 choice sets with two alternatives each, and I gathered 84 responses, the combined total observations under analysis are 2016 (8x2x12), with 1008 choices made. This allowed me to estimate the following models:

5.2.1. Multinomial logit model

In table 5.1 is the multinomial logit model (abbreviated to MLN onwards), estimated on the preference space through the use of the logitr package (Helveston, 2022).

Table 5.1. MNL - Preference Space

	MNL - Preference Space		
	Estimate	Std. Error	
price	-0.24	0.02	***
recomyes	2.57	0.42	***
rating 4 to 5	-0.14	0.16	
speedred mod to sev	-0.79	0.19	***
speedred none to lig	0.59	0.16	***
logref	1.24	0.14	***
logref e audit	1.7	0.21	***
countryless50	-0.11	0.19	
countryplus80	0.61	0.14	***
Log Likelihood	-452.77		
null.logLik	-698.69		
AIC	923.78		
BIC	967.78		
R^2	0.35		
Adj. R^2	0.34		
nobs	1008		

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

I incorporated the relevant attributes in the model as dummy variables, with *k -1* variables for every level. Since the model in table 5.1 is estimated in the preference space, the coefficients can not be interpreted directly given they are expressed in terms of marginal utility. As such, we need to convert them as previously mentioned into willingness to pay estimates, by taking into account the marginal rate of substitution in regards to price.

Doing so, we arrive at the estimates in table 5.2:

Table 5.2. MNL - Willingness to Pay Estimates (MRS)

	MNL - willingness to pay (MRS)				
	Estimate	Std. Error	z-value	$Pr(>\|z\|)$	
recomyes	10.72	1.98	5.39	6.93E-08	***
rating4 to 5	-0.59	0.67	-0.88	0.3809754	
speedred mod to sev	-3.3	0.88	-3.78	0.0001577	***
speedred none to lig	2.47	0.67	3.69	0.0002234	***
logref	5.18	0.57	9.06	< 2.20E-16	***
logref e audit	7.11	0.81	8.79	< 2.20E-16	***
countryless50	-0.47	0.79	-0.61	0.5460303	
countryplus80	2.55	0.62	4.1	4.21E-05	***

We can also estimate the multinomial model through the WTP-Space, as exposed in our methodology in table 5.3:

Table 5.3. MNL - WTP Space

	MNL - WTP Space		
	Estimate	Std. Error	
scalePar	0.23	0.02	***
recomyes	10.7	1.96	***
rating 4 to 5	-0.59	0.66	
speedred mod to sev	-3.11	0.86	***
speedred none to lig	2.48	0.65	***
logref	5.18	0.57	***
logref e audit	7.11	0.8	***
countryless50	-0.48	0.78	
countryplus80	2.6	0.61	***
Log Likelihood	-452.77		
null.logLik	-698.69		
AIC	923.78		
BIC	967.78		
R^2	0.35		
Adj. R^2	0.34		
nobs	1008		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Which will yield us similar results, as noted in the comparison table 5.4:

Table 5.4. Comparison of WTP Estimates - MRS and WTP-Space

	pref (MRS)	wtp	difference
scalePar	0.24	0.24	9.12E-06
recomyes	10.72	10.72	-4.80E-04
rating4 to 5	-0.59	-0.59	2.64E-04
speedred mod to sev	-3.31	-3.31	1.24E-03
speedred none to lig	2.47	2.48	3.82E-04
logref	5.18	5.18	-3.36E-04
logref e audit	7.12	7.12	-3.40E-04
countryless50	-0.48	-0.48	-4.94E-04
countryplus80	2.55	2.55	-6.03E-04
logLik	-452.77	-452.77	2.55E-06

5.2.2. Mixed logit model

The mixed multinomial model (MMNL from this point onwards) goes beyond the multinomial model, and allows random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time. By not requiring IIA, I am able to derive effects taking into account the heterogeneity of the sample, and as such reach conclusions that do not rely on the strict assumptions kept in the standard multinomial logit model. This was accomplished with simulation in Rstudio. The model is computationally generated and was programmed in R, with 200 random draws. Table 5.5 shows the computed model:

Table 5.5. MMNL - Preference Space

	MMNL - Preference Space		
	Estimate	Std. Error	
price	-0.29	0.03	***
recomyes	2.59	0.69	***
rating 4 to 5	-0.17	0.21	
speedred mod to sev	-0.76	0.21	***
speedred none to lig	0.71	0.21	***
logref	1.52	0.19	***
logref e audit	1.96	0.29	***
countryless50	-0.21	0.21	
countryplus80	0.62	0.14	***
sd_recomyes	-0.25	2.85	
sd_rating 4 to 5	1.11	0.3	
sd_speedred mod to sev	-0.07	0.27	
sd_speedred none to lig	-0.02	0.33	
sd_logref	0.55	0.28	
sd_logref e audit	-1.45	0.38	
sd_countryless50	-0.03	0.36	
sd_countryplus80	0.03	0.22	
Log Likelihood	-452.7		
null.logLik	-698.69		
AIC	923.78		
BIC	967.78		
R^2	0.35		
Adj. R^2	0.34		
nobs	1008		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

5.2.3. Willingness-to-pay estimates

Taking the implied WTP from the preference model, I arrive at the following willingness to pay estimations for the general sample in table 5.6:

Table 5.6. MMNL model - willingness to pay estimates

	MMNL - willingness to pay (MRS)				
	Estimate	Std. Error	z-value	$Pr(>\|z\|)$	
scalePar	0.29	0.03	10.76	2.20E-16	***
recomyes	8.95	2.54	3.52	0.0004	***
rating4 to 5	-0.57	0.76	-0.76	0.4485	
speedred mod to sev	-2.62	0.82	-3.20	0.0014	**
speedred none to lig	2.46	0.69	3.54	0.0004	***
logref	5.26	0.56	9.38	2.20E-16	***
logref e audit	6.79	0.93	7.32	2.39E-13	***
countryless50	-0.73	0.73	-1.00	0.3196	
countryplus80	2.14	0.56	3.84	0.0001	***
sd_recomyes	-0.86	10.00	-0.09	0.9313	
sd_rating4	3.82	0.97	3.92	0.0001	***
sd_speedred mod to sev	-0.25	0.94	-0.26	0.7925	
sd_speedred none to lig	-0.06	1.17	-0.05	0.9565	
sd_logref	1.89	0.91	2.07	0.0387	*
sd_logref e audit	-5.00	1.19	-4.21	2.60E-05	***
sd_countryless50	-0.09	1.25	-0.07	0.9406	
sd_countryplus80	0.11	0.77	0.14	0.8904	

5.3. Privacy literacy subgroup models

5.3.1. Willingness-to-pay estimates

Applying the criteria expressed in previous sections, I divided the sample into two subgroups, “Basic Literacy” and “Advanced Literacy”. Based on this, I calculated the preference space models for the MNL and MMNL models available in tables C.1 and C.2, respectively.

Below we can see the corresponding models with willingness to pay estimates:

Table 5.7. MNL - WTP estimates per subgroup of privacy literacy

	BLMNL			ALMNL		
	Estimate	Std. Error		Estimate	Std. Error	
scalePar	0.27	0.03	***	0.21	0.03	***
recomyes	10.5	2.46	***	10.65	3.25	**
rating 4 to 5	-0.03	0.76		-1.59	1.28	
speedred mod to sev	-2.78	0.97	**	-4.18	1.72	*
speedred none to lig	2.38	0.76	**	2.8	1.27	*
logref	4.5	0.64	***	6.5	1.16	***
logref e audit	4.47	0.93	***	12	1.67	***
countryless50	-0.23	0.89		-0.1	1.49	
countryplus80	1.63	0.67	***	4.19	1.27	***
Log Likelihood	-270.77			-171.2		
null.logLik	-432.52			-266.2		
AIC	559.54			360.31		
BIC	599.47			395.87		
R^2	0.37			0.35		
Adj. R^2	0.35			0.32		
nobs	624			384		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5.8. MMNL - WTP estimates per subgroup of privacy literacy

	BLMMNL			ALMMNL		
	Estimate	Std. Error		Estimate	Std. Error	
scalePar	0.32	0.04	***	0.26	0.04	***
recomyes	8.85	5.68		13.39	16.10	
rating 4 to 5	-0.16	0.92		-1.46	1.51	
speedred mod to sev	-2.42	0.95	*	-3.37	1.68	*
speedred none to lig	2.12	0.81	**	2.63	1.37	.
logref	4.71	0.67	***	6.53	1.17	***
logref e audit	4.25	1.02	***	12.16	2.11	***
countryless50	-0.41	0.85		-1.44	1.47	
countryplus80	1.38	0.62	*	3.50	1.22	**
sd_recomyes	-0.52	43.46		-7.77	16.16	
sd_rating 4 to 5	3.75	1.12	***	-4.28	1.90	*
sd_speedred mod to sev	-0.42	1.31		0.02	1.51	
sd_speed red none to lig	0.27	2.08		0.39	1.83	
sd_logref	1.99	1.08		-2.44	1.71	
sd_logref e audit	-3.42	1.39	*	6.49	2.66	*
sd_countryless30	-0.14	1.43		0.16	2.78	
sd_countryplus80	0.12	1.08		0.05	1.25	

5.4. Model Comparison

5.4.1. Statistical hypothesis test

Hausman-McFadden

The Hausman-McFadden test, despite serving as a traditional method of testing for the IIA assumption, is not possible with the current data set. The reason is we forced respondents to answer always two choices, and the Haushman-McFadden tests requires at least three choices (since it evaluates changes in the choice behavior of the restricted choice set that is obtained by eliminating one of the alternatives).

5.4.2. Goodness-of-fit

The package logitr by Helveston (2022) naturally calculates for each model the associated pseudo R^2 as well its BIC and AIC. Pseudo R^2 is used traditionally as a rule of thumb to measure fit of a model, yet it can not be interpreted as directly as the R^2 used in linear regressions. Comparing pseudo R^2, which can be seen in the table below:

TABLE 5.9. Goodness-of-fit and Information Measures

	MNL	MMNL	BLMNL	ALMNL	BLMMNL	ALMMNL
McFadden R^2	0.35	0.35	0.37	0.36	0.38	0.36

I note the differences are relatively low between different models, with all displaying a good fit score. Unlike with standard linear regressions R^2, a score between 0.2 and 0.4 is traditionally considered a good fit.

5.4.3. Predicting probabilities

Once a model has been estimated, it can be used to predict probabilities, outcomes, or both for a set of alternatives Helveston (2022). In our case, it is useful to measure the expected probability of a VPN user choosing alternative A versus alternative B, taking into account the various attributes that compose it, which is weighted according to the models previously estimated.

This allows us to partially answer hypothesis one, regarding the expected probability of choosing a free alternative versus a paid one. Using both the MNL and MMLN model, I estimated the average probability of choosing a free alternative over a paid one when both choices are given, to the full set of data collected. We have the results in Table 5.10 that represent the output, with the corresponding confidence intervals and estimated probability:

TABLE 5.10. Estimated probability of choosing a free option over a paid option

	Average	Confidence interval
Model		0.95 %
MNL	0.7394	0.65 - 0.8
MMNL	0.7393	0.64 - 0.8

Therefore, given the probability of choosing a free alternative is higher than choosing a paid one, with the lower bound above 60 % for a 95 % confidence interval, I conclude:

Reject *Hypothesis 1: Probability of choosing a free VPN app over a paid one, all attributes constant, is equal to zero*

5.4.4. Evaluating prediction accuracy

A sensible and widely used way to assess the performance of multinomial models is through the prediction of correct results, estimated on the population under analysis. By taking into account correct guesses over incorrect ones, it is possible to assess how different models perform both against each other, and against the expected accuracy if the choices were random (in this case, 50 %). Table 5.11 displays the calculated accuracy percentage:

Table 5.11. Accuracy of main models estimated

Accuracy of main models estimated

Model	Preference - Space	WTP - Space	Sample (n)
MNL	0.71	0.728	84
MMNL	0.72	0.69	84
BLMNL	74.6	72.9	52
ALMNL	0.69	0.71	32
BLMMNL	0.73	0.72	52
ALMMNL	0.77	0.72	32

I note all models display above 50 % accuracy, with the highest displayed by the Basic Literacy Multinomial Model (BLMNL), followed by the base MNL and MMNL model with above 70 % accuracy.

5.5. Results

5.5.1. General results

Average monthly cost is significant across all models, which was expected. Price is often the first answer participants in the focus group and experts gave when asked about preferred characteristics in a VPN service. This is the only coefficient that I do not interpret directly - since it is used as a measure to estimate willingness to pay on all other relevant attributes. Nevertheless, its continuing significance allows me to:

Reject *Hypothesis 2: Average monthly cost has no impact on the likelihood of choosing a VPN service*

In regards to all multinomial logit models participants showed a clear disposition towards recommended alternatives. The variable recommendation, which took two values (yes and no) captured the highest coefficient of all attributes. This was followed by logging, then reduction on internet speed, and finally the number of countries with servers, with all displaying significance across all models estimated, both in the preference space and willingness to pay space (with p-value < 0.01).

The signs are the expected ones - with recommendation of friends/family, logging, and number of countries with servers having a positive marginal effect on expected utility, and with price and reduction on internet speed having a negative marginal effect on expected utility.

To extract meaningful insights and test the hypotheses I laid out in the research design, I calculated the marginal rate of substitution for both MNL and MMNL models, expressed in table 5.4 and table 5.6.

Recommendation of friends and family stands out as the most valued (but also with the greatest standard error) at around 10,72 € (MNL). However, for this attribute, the MMNL model predicts a value of 8.95 €, which is a lower and more conservative estimate of how much participants were willing to pay for a VPN service that was recommended by friends or family. This allows us to ascertain:

Reject *Hypothesis 3: Recommendations from friends/family have no impact on willingness to pay for a VPN service*

What was surprising, was the non-significance of rating at any significance level. It could be rating was interpreted as being solely rating given by rating firms, which was a feedback shared by one of the survey participants, who did not notice the variable included customer review rating. Perhaps rating does not mean much if consumers don't know the *origin* of the rating, which despite being enunciated in the DCE, could have been missed or not incorporated adequately in the context of the DCE. This is a topic that should be more explored in future work since recommendation websites are still a major force in the VPN ecosystem (Ramesh et al., 2022).

Don't reject *Hypothesis 4: Rating from specialized magazines and consumers has no impact on willingness to pay for a VPN service*

Followed by the attribute of recommendations of friends or family was logging, with an estimated willingness to pay 5,18 € (MNL) and 5,26 € (MMNL) for a VPN service that referenced the data it kept of its users in its website or privacy policy. Additionally, participants were willing to pay a extra 1,94 € in the MNL model and 1,13 € in the MMNL model for a VPN service that was audited in this regard by an independent firm. The dummy variables set behave as expected, therefore I:

Reject *Hypothesis 6: Logging has no impact on willingness to pay for VPN service*

When it comes to reduction on internet speed, a VPN service that has a chance to drastically reduce internet speed results in participants being willing to give less 3,3 € (MNL) and 2,6 € (MMNL) for it. On the other hand, a VPN service that has little impact on internet speed results in a added willingness to pay estimated in the MNL model of 2,47 € (2,46 in MMNL) . These levels should be interpreted in relation to the baseline level of little to moderate reduction on internet speed (giving us how much participants are more or less willing to pay in regards to VPN service in that speed level). Considering both the MNL and MMNL models gave very similar outputs, and all were significant, I:

Reject *Hypothesis 5: Reduction of internet speed associated with the VPN service has no impact on willingness to pay for a VPN service*

The final variable of interest is the number of countries with servers, which seems to only be significant to a very large set of countries (over 80). Participants were willing to pay in average more 2,6 € (MNL) and 2,14 € (MMNL) for such a level. Considering participants did not give much weight to number of countries with servers in a VPN service, it is likely the relevance of this attribute is lost on lower privacy literacy participants. After, I shall demonstrate how this attribute differs among different privacy literacy levels and what that could tell us. Taking this into account, based on the models estimated, I:

Partially reject *Hypothesis 7: Number of countries with servers the VPN service offers has no impact on willingness to pay for the VPN service*

5.5.2. Comparison with after DCE results

As mentioned earlier in chapter 4.4., at the end of the DCE participants had to answer three questions - 1) how difficult they found the choices, 2) if they took all attributes into account, and 3) if not, which attributes they did take into account (being able to select a maximum of 3).

This allows us to infer a few things from the model - first off, with question 1), if participants faced a cognitive burden. In this matter, provided the great majority of the participants (86%) answered they found the questions either very easy, easy, nor nor hard or easy, I conclude it is unlikely. Only 4 % found the questions very hard. However, 33 participants left during the choice making part of the DCE, which could indicate those that faced the hardest cognitive burden left before completing the survey. Since the data collected only included those who completed the experiment, leavers are excluded by default from the sample.

Question 2 and 3 allowed to make an additional judgment on what attributes participants perceived as less relevant, as well as what percentage of participants took all attributes into account (64 %). This means 36 % did not consider at least one attribute in their decision making.

Unsurprisingly, the most mentioned valued attribute was the average monthly cost of the VPN service, with 33 % of participants selecting it. Next was effect on internet speed (23 %), followed by logging (19 %). On the other hand, rating had a very small percentage of choices (the smallest amongst all alternatives, with 8 %), accompanied with number of countries with servers, with also 8 %.

An unexpected, puzzling scenario concerns recommendations (9 %), where the perceived effect versus the actual estimated effect is lower in magnitude. There seems to be a gap between the perceived weight participants give to recommendations of friends and family on their decision making and the actual estimated effects from the models I ran. From this early analysis, participants seem to undervalue their reliance on recommendations to make choices regarding VPN services. Investigating if this coefficient results from an error in model or experiment specification or if it truly represents a dichotomy in perceived influence versus actual influence of recommendations is out of scope of this work, but could be expanded in future research. The full table B.2 can be found in the appendix.

5.5.3. Privacy literacy results

After running both models, it is apparent valued attributes and levels are similar between each subgroup and across models, with country, logging, effect on internet speed and recommendations as significant (p-value < 0.01). Rating remains non significant for both groups, as well as levels of country below 80.

When accounting for privacy literacy differences, the willingness to pay in terms of logging and number of country differs. A VPN service that offers a greater amount of countries or a greater degree of transparency (measured through logging), is valued more by those with advanced privacy literacy than one who does not. This is expressed in willingness to pay estimates of 4,19 (MNL) € for a VPN with servers in over 80 countries for the Advanced Privacy Literacy subgroup versus 1,63 € (MNL) in the Basic Privacy Literacy subgroup. The same happens for logging, with the Advanced Privacy Literacy subgroup willing to pay 12 € (MNL) extra for a service that is audited, versus the Basic Privacy Literacy subgroup which is willing to pay around 4,47 (MNL) € for the same level.

It is possible, due to the small sample in the Advanced Privacy Literacy subgroup, effects are overestimated. Nevertheless, these differences indicate a degree of difference in preferences between both groups concerning these attributes.

This falls in hand with the qualitative research conducted, where experts brought up and gave greater importance to matters such as logging and location of servers. Since it seems, from this analysis, participants with a higher degree of privacy literacy are more likely to value independent audit opinions that ensure VPN firms are compliant with their privacy policies, we can estimate transparency, while relevant for all participants, is especially favored by those with a higher degree of privacy literacy.

Regarding number of countries with servers, it is also likely experts understand better the benefits and implications of a greater degree of countries with servers in the expected usefulness of a VPN service.

Based on these two differences in valuation of both logging and number of countries with servers, and seeming indifference in the rest of the attributes I can:

Partially reject *Hypothesis 8: Level of privacy literacy does not have an impact on the preferences for a VPN service.*

CHAPTER 6

Conclusion

6.1. Concluding remarks

I propose to contribute to an ever-expanding literature, in the realm of privacy economics in the form of a discrete choice experiment. Despite the multitude of discrete choice experiments conducted by past research, little attention has been given to the willingness to pay for privacy-enhancing tools, and in particular to how that willingness to pay could differ among individuals with a greater level of knowledge and awareness of online privacy risks. Another factor that often goes amiss in the literature is a proper discussion of attributes and levels, with frequent studies not spending enough effort and time on this crucial step for correct model specification. Additionally to the quantitative research, this work proposes to contribute to the advancement of current qualitative methodologies in discrete choice experiments by substantiating the design of the DCE with a series of expert interviews, focus group with non-experts, and literature review to justify the inclusion and exclusion of each attribute and level individually.

Experts' interviews and the focus group with non experts were fundamental since part of the research question was concerned with how different privacy literacy levels could impact the valuation of attributes within a VPN service.

In this phase, I uncovered central differences among groups in how they perceived their privacy online, as well as what could be done to mitigate online privacy risks. Expert's emphasis on the importance of logging, location and number of servers as well as factors influencing a VPN services effect on reduction of internet speed were illuminating, and allowed to discern probable key attributes in higher privacy literacy participants. In the focus group, participants indicated the strength of personal recommendations, effect on internet speed, user interface, offering of the VPN service as part of a Anti-virus or internet browser. Price and price promotions were noted by both groups as significant, and in particular participants in the focus group showed a very strong preference for free alternatives.

I juxtaposed the relevant attributes identified in the qualitative research to a literature review as well as an independent search of the top 20 VPN services by webpage visits, in order to construct proposed attributes and levels.

This allowed me to design an online experiment, where twelve choice sets were presented to 84 participants, for a combined number of 2016 observations and 1008 choices. Besides a demographic questionnaire, a privacy literacy questionnaire was included after the choices based on the dimensions identified by Trepte et al. (2015), which were used as basis to design questions for individuals to self-assess their online privacy literacy. This self assessment was subsequently validated based on the score of a true and false questionnaire, to categorize participants into either a basic or advanced privacy literacy category.

With the use of a standard multinomial model as well as a mixed multinomial model, using novel R packages developed by Helveston (2022) and Croissant (2020), I estimated marginal utility estimates and corresponding willingness to pay for several attributes within a VPN service in a Portuguese sample. Models were generated for the base sample, and for the subgroups corresponding to the basic and advanced literacy categories, in order to arrive at meaningful differences between both.

The models were then used to predict probabilities and outcomes, and the accuracy of them was tested against real choices taken to compare success rates among different models. Besides this, standard statistical information such as McFaddens pseudo-R^2, BIC and AIC were also used as a means to distinguish goodness-of-fit among the various models.

Results across models were similar. Recommendation by friends and family is noted as the attribute with the highest willingness to pay, subject however to a high standard deviation. An unexpected finding concerns the difference between the high coefficient of this attribute and the relatively low importance participants gave to the attribute, when asked directly for the most valued attributes. This discrepancy did not happen for other attributes, where for instances, rating was both unpopular as answer and insignificant as a estimated willingness to pay coefficient.

The willingness to pay of logging and reduction of internet speed were also significant, and behaved as expected. The attribute with the lowest willingness to pay estimates was number of countries with servers, where participants seemingly only valuated VPNs with over 80 countries. Differences between subgroups when taking into account different levels of privacy literacy demonstrated for both models differences, both in logging and number of countries with servers. There is a limitation in the accuracy of these sub models, since they take into account smaller samples and as such coefficients and estimates could be overvalued. This is particularly true for the advanced privacy literacy subgroup, where a bigger sample could provide more accurate estimates.

6.2. Future Research

I identify a couple of important steps future research could take. The first is replaying the experiment, conducting the attribute and level selection separately among different populations, and expanding on interviewing a wider and more diverse population. Time and budget constraints made this task difficult, and greater insight into how different populations view rating, in particular the many forms it takes (customer reviews, VPN services comparison websites, expert magazines, consumer defense magazines) could provide greater insight into how this attribute can or should be incorporated in future DCEs, taking it was non significant in all models and subgroups analyzed in the scope of this thesis. The effect of recommendations that was observed, where experiment results greatly differed from the perceived effects participants gave to the attribute could also be expanded on. Secondly, experiment design is a topic with extensive literature. I postulate that more efficient designs could present more efficient ways of presenting choices to participants, which would lead to the generation of more accurate results with the same sample size. During the course of this DCE, participants were presented with two choice sets with 2 dominant choices, where one alternative choice clearly dominated the other in all levels and attributes. The reasoning was the design was conducted on the basis of orthogonality, which despite being useful as a mathematical propriety, also acts as a constraint. Generating efficient designs taking into account a no-dominant alternative scenario is possible, however, besides such specification not being available in all programs (including the one I used, JMP), doing so usually also requires a prior estimate of the utility participants derive from one level over another. This means usually such designs are estimated by running a pilot experiment in a percentage of the population (say 20 %), then adjusted taking into account estimated coefficients, and deployed again to capture the data that will be used for the model. Another part where this could be expanded is in power analysis, which is achieved by running Monte-Carlo simulations on different designs, and then comparing estimated efficiency of results as criteria to choose the most efficient design. This was left out of the analysis, since it presents a computationally intensive effort that could not be expended in the scope of this work.

Finally the ever evolving privacy landscape shows this topic is meant to stay at the forefront of public opinion and discourse. As new tools are ushered in to often assuage fears of data loss, privacy breaches and mass surveillance, individuals care about transparency from all sides, not only from the firms that collect their data but also of those who prevent it (e.g. VPN services). Future work should keep evaluating the decision-making behind choosing these tools, and bring additional insight into the privacy debate.

References

Acquisti, A., Brandimarte, L., and Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221):509–514.

Acquisti, A., Brandimarte, L., and Loewenstein, G. (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*, 30(4):736–758.

Acquisti, A. and Grossklags, J. (2003). Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behavior. In *2nd Annual Workshop on Economics and Information Security-WEIS*, volume 3, pages 1–27. Citeseer.

Acquisti, A. and Grossklags, J. (2004). Privacy attitudes and privacy behavior. In *Economics of information security*, pages 165–178. Springer.

Acquisti, A., John, L. K., and Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274.

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of economic Literature*, 54(2):442–92.

Akgul, O., Roberts, R., Namara, M., Levin, D., and Mazurek, M. L. (2022). Investigating Influencer VPN Ads on YouTube. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 876–892. ISSN: 2375-1207.

Altman, I. (1976). A conceptual analysis. *Environment and behavior*, 8(1):7–29.

Bhageshpur, K. (2019). *Data Is The New Oil – And That's A Good Thing.* https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=ca85d8d73045 [Accessed: 20-03-2021].

Bhageshpur, K. (2020). *What is GDPR, the EU's new data protection law?* https://gdpr.eu/what-is-gdpr/ [Accessed: 20-03-2021].

Brown, M. and Muchira, R. (2004). Investigating the relationship between internet privacy concerns and online purchase behavior. *Journal of Electronic Commerce Research*, 5(1):62–70.

Brush, A. B., Krumm, J., and Scott, J. (2010). Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 95–104.

Center for Data Innovation (2020). *GDPR in numbers.* https://euagenda.eu/publications/gdpr-in-numbers [Accessed: 20-03-2021].

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., and Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Economics*, 21(6):730–741. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.1739.

Cohen, P., Cohen, J., Aiken, L. S., and West, S. G. (1999). The Problem of Units and the Circumstance for POMP. *Multivariate Behavioral Research*, 34(3):315–346. Publisher: Routledge _eprint: https://doi.org/10.1207/S15327906MBR3403_2.

Croissant, Y. (2020). Estimation of random utility models in r: The mlogit package. *J. Stat. Softw.*, 95(11).

De Capitani Di Vimercati, S., Foresti, S., Livraga, G., and Samarati, P. (2012). Data privacy: definitions and techniques. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20(06):793–817.

Draper, N. A. and Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, 21(8):1824–1839.

Edelman, B. G. and Luca, M. (2014). Digital discrimination: The case of airbnb. com. *Harvard Business School NOM Unit Working Paper*, (14-054).

European Comission (2020). *GDPR checklist for data controllers.* https://gdpr.eu/checklist/ [Accessed: 20-03-2021].

Helveston, J. (2022). *logitr: Fast Estimation of Multinomial and Mixed Logit Models with Preference Space and Willingness to Pay Space Utility Parameterizations.*

Hermstruwer, Y. (2017). Contracting around privacy: the (behavioral) law and economics of consent and big data. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, 8:9.

Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments. *Ecological Economics*, 69(8):1595–1603.

Johnston, M. (2021). *What Your Data Is Really Worth to Facebook.* https://www.investopedia.com/ask/answers/120114/how-does-facebook-fb-make-money.asp [Accessed: 20-03-2021].

Khan, M. T., DeBlasio, J., Voelker, G. M., Snoeren, A. C., Kanich, C., and Vallina-Rodriguez, N. (2018). An Empirical Analysis of the Commercial VPN Ecosystem. In *Proceedings of the Internet Measurement Conference 2018*, IMC '18, pages 443–456, New York, NY, USA. Association for Computing Machinery.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*, 64:122–134.

Kumaraguru, P. and Cranor, L. F. (2005). *Privacy indexes: a survey of Westin's studies.* Carnegie Mellon University, School of Computer Science, Institute for . . . .

Lancsar, E. and Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A User's Guide. *PharmacoEconomics*, 26:661–77.

Liu, B., Andersen, M. S., Schaub, F., Almuhimedi, H., Zhang, S. A., Sadeh, N., Agarwal, Y., and Acquisti, A. (2016). Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, pages 27–41.

Löfgren, Å., Martinsson, P., Hennlock, M., and Sterner, T. (2012). Are experienced people affected by a pre-set default option—results from a field experiment. *Journal of Environmental Economics and management*, 63(1):66–72.

Louviere, J. J., Flynn, T. N., and Carson, R. T. (2010). Discrete Choice Experiments Are Not Conjoint Analysis. *Journal of Choice Modelling*, 3(3):57–72.

Mangham, L. J., Hanson, K., and McPake, B. (2008). How to do (or not to do) ... Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning*, 24(2):151–158. _eprint: https://academic.oup.com/heapol/article-pdf/24/2/151/40836144/heapol_24_2_151.pdf.

McDonald, A. M. and Cranor, L. F. (2008). The cost of reading privacy policies. *Isjlp*, 4:543.

Namara, M., Wilkinson, D., Caine, K., and Knijnenburg, B. P. (2020). Emotional and Practical Considerations Towards the Adoption and Abandonment of VPNs as a Privacy-Enhancing Technology. Accepted: 2022-05-11T17:05:01Z Publisher: Proceedings on Privacy Enhancing Technologies.

Oh, S. and Syn, S. Y. (2015). Motivations for sharing information and social support in social media: A comparative analysis of f acebook, t witter, d elicious, y ou t ube, and f lickr. *Journal of the Association for Information Science and Technology*, 66(10):2045–2060.

Potoglou, D., Dunkerley, F., Patil, S., and Robinson, N. (2017). Public preferences for internet surveillance, data retention and privacy enhancing services: Evidence from a pan-european study. *Computers in Human Behavior*, 75:811–825.

Pérez-Troncoso, D. (2020). A step-by-step guide to design, implement, and analyze a discrete choice experiment. arXiv:2009.11235 [econ].

Ramesh, R., Vyas, A., and Ensafi, R. (2022). "All of them claim to be the best": Multi-perspective study of VPN users and VPN providers. arXiv:2208.03505 [cs].

Reed Johnson, F., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., and Bridges, J. F. P. (2013). Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health*, 16(1):3–13.

Rodriguez, K. and Alimonti, V. (2020). *A Look-Back and Ahead on Data Protection in Latin America and Spain.* https://www.eff.org/deeplinks/2020/09/look-back-and-ahead-data-protection-latin-america-and-spain [Accessed: 20-03-2021].

Shapiro, R. J. (2019). *What Your Data Is Really Worth to Facebook.* https://washingtonmonthly.com/magazine/july-august-2019/what-your-data-is-really-worth-to-facebook/ [Accessed: 20-03-2021].

Skatova, A., McDonald, R. L., Ma, S., and Maple, C. (2019). Unpacking privacy: Willingness to pay to protect personal data.

Sobolewski, M., Paliński, M., et al. (2017). How much consumers value on-line privacy? welfare assessment of new data protection regulation (gdpr). Technical report.

Solove, D. J. (2020). The myth of the privacy paradox. *Available at SSRN*.

Sombatruang, N., Omiya, T., Miyamoto, D., Sasse, M. A., Kadobayashi, Y., and Baddeley, M. (2020). Attributes Affecting User Decision to Adopt a Virtual Private Network (VPN) App. In Meng, W., Gollmann, D., Jensen, C. D., and Zhou, J., editors, *Information and Communications Security*, Lecture Notes in Computer Science, pages 223–242, Cham. Springer International Publishing.

Stewart, E. (2018). *Lawmakers seem confused about what Facebook does — and how to fix it.* https://www.vox.com/policy-and-politics/2018/4/10/17222062/mark-zuckerberg-testimony-graham-facebook-regulations [Accessed: 20-03-2021].

Tamir, D. I. and Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, 109(21):8038–8043.

Trepte, S., Teutsch, D., Masur, P. K., Eicher, C., Fischer, M., Hennhöfer, A., and Lind, F. (2015). Do People Know About Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS). In Gutwirth, S., Leenes, R., and de Hert, P., editors, *Reforming European Data Protection Law*, Law, Governance and Technology Series, pages 333–365. Springer Netherlands, Dordrecht.

van Ooijen, I. and Vrabec, H. U. (2019). Does the gdpr enhance consumers' control over personal data? an analysis from a behavioural perspective. *Journal of consumer policy*, 42(1):91–107.

Wagner, A. and Mesbah, N. (2019). Too confident to care: Investigating overconfidence in privacy decision making.

Wagner, A., Wessels, N., Buxmann, P., and Krasnova, H. (2018). Putting a price tag on personal information-a literature review. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Wigand, C., Mercier, G., and Kolanko, K. (2020). *Two years of the GDPR: Questions and answers.* https://ec.europa.eu/commission/presscorner/detail/es/qanda$_2$0$_1$166[$Accessed : 20 - 03 - 2021$].

Winegar, A. G. and Sunstein, C. R. (2019). How much is data privacy worth? a preliminary investigation. *Journal of Consumer Policy*, 42(3):425–440.

APPENDIX A

R Code

A.1. R Code - MNL

Below, we can see the relevant R code for the base multinomial logit models estimated:

```r


#I highly recommend the work by Yves Croissant (2020) in terms of theoretical basis for the mlogit package. This package is also one of the older
#R packages for estimation of multinomial models.
#Besides this, the work by John Helveston from Washinghton University is useful because of the logitr package, for Logit Models w/Preference & WTP Space Utility Parameterizations as well as estimation of MLN and mixed MLN models.
#the mlogit functions works with data in wide/long format, while logitr only works with data in long. Conjointly exports by default
#the data to long format, so it's easier to work on the logitr packages, but mlogit also works.
#on the other hand, mlogit is easier to print (stargazer package prints) and has more support since it is older and more popular
#both models can be exported to code that can be directly put in LaTeX
#This R script has just the basic MNL model, aka has no no-choice alternatives,
#or mixed/nested MNL models. See other R scripts I wrote for estimation of those.
#all models include also prediction of results and estimated probabilities

#                               Index
# Part 1.1: Estimation of the base model utilities and WTP with DCE long data - mlogit
# Part 1.2: Estimation of the base model utilities and WTP with basic literacy data - mlogit
# Part 1.3: Estimation of the base model utilities and WTP with advanced literacy data - mlogit


# Part 2.1: Estimation of base MNL model with logitr package and estimation in the preference/WTP space - logitr
# Part 2.2: Estimation of base MNL model with basic literacy data, logitr package and estimation in the preference/WTP space - logitr
# Part 2.3: Estimation of base MNL model with advanced literacy data, logitr package and estimation in the preference/WTP space - logitr

#Each part should include printing of models - mlogit and logitr

#tests to test assumptions of models and compare models

#predicting probabilities with generated models - logitr

#                              Start


#Part 1: Estimation of the base model with DCE long data

#install relevant packages mlogit and logitr (for estimations), stargazer and texreg (for output in LaTeX of tables)
#install.packages(”mlogit”)
#install.packages(”logitr”)
#install.packages(”texreg”)
#install.packages(”stargazer”)


#loading package mlogit and logitr
library (”mlogit”)
library(”logitr”)


#DCE sample data set -> loading the data set, already in long format from path set
dcedatalong <- read.csv(”C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalongmlogit62.csv”)
attach(dcedatalong)

#top rows of dataset
head(dcedatalong)

#run base model
mlogitdata <- mlogit.data(dcedatalong, choice = ”choice”, shape = 'long',
                          alt.var = ”mode”, chid.var = ”individual”,
                          drop.index = FALSE)
mbl <- mlogit(choice ~ price + recom + rating + speed + log + country | -1, mlogitdata)
summary(mbl)

#calculate WTP coefficients by computing marginal rates of substitution
#Analysis questions here: Are the coefficients significant? Are the signs the predicted ones? Warning: this is not directly interpretable
```

```r
#but it's possible to interpret in terms of positive/negative marginal utility, and on which coefficients are higher.
wtp_mlogit <- coef(m)[-1]/coef(m)[1]
wtp_mlogit


#Print the models

#we have it's easy to print to LaTex using stargazer, for the mlogit package (since stargazer supports it)
#but the logitr package can't be printed with stargazer, ergo the data needs to be filled manually which is very time consuming, so I recommend the package texreg
#intall.packages("texreg")

#install.packages("stargazer")
library(stargazer)

#print basic MNL model with stargazer, allows to export directly to laTeX if you remove "text"
stargazer(mbl)

#prints the wtp, the data is sent to a dataframe and it' prints it (same logic to LaTex as above)
stargazer(wtp_mlogit)

##

#Part 1.2: Estimation of the base model utilities and WTP with basic literacy data - mlogit


#DCE sample data set -> loading the data subset, already in long format from path set.

dcedatalong_bl <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalongmlogit62.csv")

attach(dcedatalong_bl)

#DCE sample data set - Basic Literacy subset
summary(dcedatalong_bl)


#putting the data into mlogit data format (in this case data is in long format)
```

```
mlogitdata_bl <- mlogit.data(dcedatalong, choice = "choice", shape = "long",
                             alt.var = "mode", chid.var = "individual",
                             drop.index = FALSE)

#First n rows
head(mlogitdata_bl)

#Estimation of the model and summary
m_bl <- mlogit(choice ~ price + recom + rating + speed + log + country | -1, mlogitdata_bl)
summary(m_bl)

#Analysis questions here: Are the coefficients significant? Are the signs the predicted ones?
#Part 2: Calculation of WTP of each attribute based on sample DCE long data
wtp_mlogit_bl <- coef(m_bl)[-1]/coef(m_bl)[1]
wtp_mlogit_bl

##

#Part 1.3: Estimation of the base model utilities and WTP with advanced literacy data - mlogit

dcedatalong_al <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalong62.csv")

attach(dcedatalong_al)

#DCE sample data set - Basic Literacy subset
summary(dcedatalong_al)


#putting the data into mlogit data format (in this case data is in long format)
mlogitdata_al <- mlogit.data(dcedatalong_al, choice = "choice", shape = "long",
                             alt.var = "mode", chid.var = "individual",
                             drop.index = FALSE)

#First n rows
head(mlogitdata_al)

#Estimation of the model and summary
mbl_al <- mlogit(choice ~ price + recom + rating + speed + log + country | -1, mlogitdata_al)
summary(mbl_al)
```

```

#Analysis questions here: Are the coefficients significant? Are the signs
    the predicted ones?


#Calculation of WTP of each attribute based on sample DCE long data
wtp_mlogit_al <- coef(mbl_al)[-1]/coef(mbl_al)[1]
wtp_mlogit


##

#  Part 2.1: Estimation of base MNL model with logitr package and
    estimation in the preference/WTP space - logitr

#We can also (see work by Helveston) directly model WTPs through the WTP
    space. There are useful
#advantages to this that shouldn't be undermined, the most obvious one
    being you can interpret this model directly in the WTP space.
#besides this, it gives a higher degree of granuality on valuable
    attributes, which is really useful

#Part 3: Estimation of base MNL model with logitr package and estimation in
     the WTP space

#load logitr if not loaded already
library(logitr)

# Estimate a preference space model (the data here needs to be in a long
    format)

#load data in long format

dcedatalong_lr <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis
    Documents/Programming and Modelling/R/Data/dcedatalong_84.csv")

attach(dcedatalong_lr)

#top rows of the data imported
head(dcedatalong_lr)

#This is going to do the model in the preference space
mnl_pref <- logitr(
  data = dcedatalong_lr,
  outcome = "choice",
  obsID = "obsID",
  pars = c("price", "recom", "rating", "speed", "log", "country")
```

```
)

summary(mnl_pref)

mnl_pref3 <- wtp(mnl_pref, scalePar =  "price")
mnl_pref3

stargazer(wtp_pref2)
stargazer(wtp_pref3)

wtp_pref2 <- coef(mnl_pref)[-1]/coef(mnl_pref)[1]
wtp_pref2

wtpCompare(mnl_pref, mnl_wtp, scalePar = 'price')


texreg(wtpCompare())

#for some reason, this package doesn't allow to estimate the WTP space directly anymore - it needs to be
#done indirectly. See below:

#estimate WTP from base MNL model
wtp_mnl_pref <- wtp(mnl_pref, scalePar = "price")

mnl_wtp <- logitr(
  data = dcedatalong_lr,
  outcome = "choice",
  obsID = "obsID",
  pars = c("recom", "rating", "speed", "log", "country"),
  scalePar = "price",
  startVals = wtp_mnl_pref$Estimate
)

summary(mnl_wtp)

#prediction of accuracy

outcomes_pref <- predict(
  mnl_pref,
  type = "outcome",
  returnData = TRUE
)

outcomes_wtp <- predict(
  mnl_wtp,
  type = "outcome",
```

```r
  returnData = TRUE
)

chosen_pref <- subset(outcomes_pref, choice == 1)
chosen_pref$correct <- chosen_pref$choice == chosen_pref$predicted_outcome
accuracy_pref <- sum(chosen_pref$correct) / nrow(chosen_pref)
accuracy_pref

chosen_wtp <- subset(outcomes_wtp, choice == 1)
chosen_wtp$correct <- chosen_wtp$choice == chosen_wtp$predicted_outcome
accuracy_wtp <- sum(chosen_wtp$correct) / nrow(chosen_wtp)
accuracy_wtp

#printing results of models to LaTeX

texreg(mnl_pref)

##

#  Part 2.2: Estimation of base MNL model with logitr package and estimation in the preference/WTP space - logitr

#Estimation of base MNL model with logitr package and estimation in the WTP space

#load logitr if not loaded already
library(logitr)

# Estimate a preference space model (the data here needs to be in a long format)

#load data in long format

dcedatalong_lrbl <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalong_84_bl.csv")

attach(dcedatalong_lrbl)

head(dcedatalong_lrbl)

#This is going to do the model in the preference space, and subsequently give the summary

mnl_pref_lrbl <- logitr(
  data = dcedatalong_lrbl,
  outcome = "choice",
  obsID = "obsID",
  pars = c("price", "recom", "rating", "speed", "log", "country")
)
summary(mnl_pref_lrbl)

#for some reason, this package doesn't allow to estimate the WTP space directly anymore - it needs to be
#done indirectly. See below:

#estimate WTP from base MNL model
wtp_mnl_pref_bl <- wtp(mnl_pref_lrbl, scalePar = "price")

#compute model in the WTP space
mnl_wtp_lrbl <- logitr(
  data = dcedatalong_lrbl,
  outcome = "choice",
  obsID = "obsID",
  pars = c("recom", "rating", "speed", "log", "country"),
  scalePar = "price",
  startVals = wtp_mnl_pref$Estimate
)
summary(mnl_wtp_lrbl)


#prediction of accuracy and estimation of probabilities

outcomes_pref_lrbl <- predict(
  mnl_pref_lrbl,
  type = "outcome",
  returnData = TRUE
)

outcomes_wtp_lrbl <- predict(
  mnl_wtp_lrbl,
  newdata = data,
  obsID    = "obsID",
  ci       = 0.95
)

chosen_pref_lrbl <- subset(outcomes_pref_lrbl, choice == 1)
chosen_pref_lrbl$correct <- chosen_pref_lrbl$choice == chosen_pref_lrbl$predicted_outcome
accuracy_pref_lrbl <- sum(chosen_pref_lrbl$correct) / nrow(chosen_pref_lrbl)
accuracy_pref_lrbl

chosen_wtp_lrbl <- subset(outcomes_wtp_lrbl, choice == 1)
```

```
chosen_wtp_lrbl$correct <- chosen_wtp_lrbl$choice == chosen_wtp_lrbl$predicted_outcome
accuracy_wtp_lrbl <- sum(chosen_wtp_lrbl$correct) / nrow(chosen_wtp_lrbl)
accuracy_wtp_lrbl

##

# Part 2.3: Estimation of base MNL model with advanced literacy data, logitr package and estimation in the preference/WTP space - logitr

#Estimation of base MNL model with logitr package and estimation in the WTP space

#load logitr if not loaded already
library(logitr)

# Estimate a preference space model (the data here needs to be in a long format)

#load data in long format

dcedatalong_lral <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalong_84_al.csv")

attach(dcedatalong_lral)

head(dcedatalong_lral)

#you put here the attributes - this is going to do the model in the preference space

mnl_pref_lral <- logitr(
  data = dcedatalong_lral,
  outcome = "choice",
  obsID = "obsID",
  pars = c("price", "recom", "rating", "speed", "log", "country")
)

summary(mnl_pref_lral)

#estimate WTP from base MNL model
wtp_mnl_pref_lral <- wtp(mnl_pref_lral, scalePar = "price")

mnl_wtp_lral <- logitr(
  data = dcedatalong_lral,
  outcome = "choice",
  obsID = "obsID",
```

```r
  pars = c("recom", "rating", "speed", "log", "country"),
  scalePar = "price",
  startVals = wtp_mnl_pref$Estimate
)
summary(mnl_wtp_lral)

#prediction of accuracy

outcomes_pref_lral <- predict(
  mnl_pref_lral,
  type = "outcome",
  returnData = TRUE
)

outcomes_wtp_lral <- predict(
  mnl_wtp_lral,
  type = "outcome",
  returnData = TRUE
)

chosen_pref_lral <- subset(outcomes_pref_lral, choice == 1)
chosen_pref_lral$correct <- chosen_pref_lral$choice == chosen_pref_lral$predicted_outcome
accuracy_pref_lral <- sum(chosen_pref_lral$correct) / nrow(chosen_pref_lral)
accuracy_pref_lral

chosen_wtp_lral <- subset(outcomes_wtp_lral, choice == 1)
chosen_wtp_lral$correct <- chosen_wtp_lral$choice == chosen_wtp_lral$predicted_outcome
accuracy_wtp_lral <- sum(chosen_wtp_lral$correct) / nrow(chosen_wtp_lral)
accuracy_wtp_lral


##

# Part 3: Printing of models - mlogit and logitr

#texreg package, works with both - here is a example, calculate output as necessary for body of work
library(texreg)

texreg(mnl_wtp, stars = c(0.01, 0.05, 0.1))

texreg(
  list(
    mnl_wtp_lrbl,
    mnl_wtp_lral
  ),
  stars = c(0.01, 0.05, 0.1),
  custom.model.names = c("Basic Privacy Literacy", "Advanced Privacy Literacy")
)
texreg(model_al, stars = c(0.01, 0.05, 0.1))
model_bl <- wtp_mnl_pref_bl
library(texreg)
Sys.setenv("R_REMOTES_NO_ERRORS_FROM_WARNINGS" = "true")
texreg(mnl_wtp)

#see relevance of results and interpretation


#####
```

MNL Code R

A.2. R Code - MMNLM

```r
#Running the mixed logit model using logitr


library("logitr")

dcemxl62 <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalong_84.csv")

attach(dcemxl62)

install.packages("stringi")


remotes::update_packages()


head(dcemxl62)

set.seed(456)

mxl_pref <- logitr(
  data      = dcemxl62,
  outcome   = 'choice',
  obsID     = 'obsID',
  panelID   = 'id',
  pars      = c('price', 'recom', 'rating', 'speed', 'log', 'country'),
  randPars = c(recom = 'n', rating = 'n', speed = 'n', log = 'n', country = 'n'),
  numMultiStarts = 10,
  numDraws = 200,
  drawType = 'sobol',
)

summary(mxl_pref)

wtp_mxl_pref <- wtp(mxl_pref, scalePar = "price")
wtp_mxl_pref


#estimating in WTP space

set.seed(6789)

mxl_wtp <- logitr(
  data      = dcemxl62,
  outcome   = 'choice',
  obsID     = 'obsID',
  panelID   = 'id',
  pars      = c('price', 'recom', 'rating', 'speed', 'log', 'country'),
  randPars = c(recom = 'n', rating = 'n', speed = 'n', log = 'n', country = 'n'),
  numMultiStarts = 10,
  drawType = 'sobol',
  startVals = wtp_mxl_pref$Estimate,
)

summary(mxl_wtp)


#prediction of probabilities

probs_mxl_pref <- predict(
  mxl_pref, returnData = TRUE,
  ci = 0.95
  )

probs_mxl_wtp <- predict(
  mxl_pref, returnData = TRUE,
  ci = 0.95
)



```

```
head(probs_mxl_pref)

write.csv(probs_mxl_pref,"C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/probs_mxnl_prefci.csv", row.names = FALSE)


#prediction of accuracy


#prediction of accuracy

outcomes_pref_mxl <- predict(
  mxl_pref,
  type = "outcome",
  returnData = TRUE
)

outcomes_wtp_mxl <- predict(
  mxl_wtp,
  type = "outcome",
  returnData = TRUE
)

chosen_pref_mxl <- subset(outcomes_pref_mxl, choice == 1)
chosen_pref_mxl$correct <- chosen_pref_mxl$choice == chosen_pref_mxl$predicted_outcome
accuracy_pref_mxl <- sum(chosen_pref_mxl$correct) / nrow(chosen_pref_mxl)
accuracy_pref_mxl

chosen_wtp_mxl <- subset(outcomes_wtp_mxl, choice == 1)
chosen_wtp_mxl$correct <- chosen_wtp_mxl$choice == chosen_wtp_mxl$predicted_outcome
accuracy_wtp_mxl <- sum(chosen_wtp_mxl$correct) / nrow(chosen_wtp_mxl)
accuracy_wtp_mxl





#basic literacy mixed logit model
library(logitr)

dcemxl84_bl <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalong_84_bl.csv")

attach(dcemxl84_bl)
```

```r

head(dcemxl84_bl)

set.seed(456)

mxl_pref84_bl <- logitr(
  data      = dcemxl84_bl,
  outcome   = 'choice',
  obsID     = 'obsID',
  panelID   = 'id',
  pars      = c('price', 'recom', 'rating', 'speed', 'log', 'country'),
  randPars = c(recom = 'n', rating = 'n', speed = 'n', log = 'n', country = 'n'),
  numMultiStarts = 10,
  drawType = 'sobol',
  numDraws = 200
)

summary(mxl_pref84_bl)

wtp_mxl_pref84_bl <- wtp(mxl_pref84_bl, scalePar =  "price")
wtp_mxl_pref84_bl


#estimating in WTP space

set.seed(6789)

mxl_wtp84_bl <- logitr(
  data      = dcemxl84_bl,
  outcome   = 'choice',
  obsID     = 'obsID',
  panelID   = 'id',
  pars      = c('price', 'recom', 'rating', 'speed', 'log', 'country'),
  randPars = c(recom = 'n', rating = 'n', speed = 'n', log = 'n', country = 'n'),
  numMultiStarts = 10,
  drawType = 'sobol',
  startVals = wtp_mxl_pref84_bl$Estimate
)

summary(mxl_wtp84_bl)

#accuracy of model

outcomes_pref_mxl84_bl <- predict(
  mxl_pref84_bl,
  type = "outcome",
  returnData = TRUE
)

outcomes_wtp_mxl84_bl <- predict(
  mxl_wtp84_bl,
  type = "outcome",
  returnData = TRUE
)

chosen_pref_mxl84_bl <- subset(outcomes_pref_mxl84_bl, choice == 1)
chosen_pref_mxl84_bl$correct <- chosen_pref_mxl84_bl$choice == chosen_pref_mxl84_bl$predicted_outcome
accuracy_pref_mxl84_bl <- sum(chosen_pref_mxl84_bl$correct) / nrow(chosen_pref_mxl84_bl)
accuracy_pref_mxl84_bl

chosen_wtp_mxl84_bl <- subset(outcomes_wtp_mxl84_bl, choice == 1)
chosen_wtp_mxl84_bl$correct <- chosen_wtp_mxl84_bl$choice == chosen_wtp_mxl84_bl$predicted_outcome
accuracy_wtp_mxl84_bl <- sum(chosen_wtp_mxl84_bl$correct) / nrow(chosen_wtp_mxl84_bl)
accuracy_wtp_mxl84_bl




#advanced literacy mixed multinomial mode




dcemxl84_al <- read.csv("C:/Users/Eduardo Jardine/Desktop/Life/Thesis Documents/Programming and Modelling/R/Data/dcedatalong_84_al.csv")

attach(dcemxl84_al)

head(dcemxl84_al)

set.seed(456)

mxl_pref84_al <- logitr(
  data      = dcemxl84_al,
  outcome   = 'choice',
  obsID     = 'obsID',
  panelID   = 'id',
  pars      = c('price', 'recom', 'rating', 'speed', 'log', 'country'),
  randPars = c(recom = 'n', rating = 'n', speed = 'n', log = 'n', country = 'n'),
  numMultiStarts = 10,
  drawType = 'sobol',
  numDraws = 200
)

summary(mxl_pref84_al)

wtp_mxl_pref84_al <- wtp(mxl_pref84_al, scalePar =  "price")
wtp_mxl_pref84_al


#estimating in WTP space

set.seed(6789)

mxl_wtp84_al <- logitr(
  data      = dcemxl84_bl,
  outcome   = 'choice',
  obsID     = 'obsID',
  panelID   = 'id',
  pars      = c('price', 'recom', 'rating', 'speed', 'log', 'country'),
  randPars = c(recom = 'n', rating = 'n', speed = 'n', log = 'n', country = 'n'),
  numMultiStarts = 10,
  drawType = 'sobol',
  startVals = wtp_mxl_pref84_al$Estimate
)

summary(mxl_wtp84_al)


#accuracy of model

outcomes_pref_mxl84_al <- predict(
  mxl_pref84_al,
  type = "outcome",
  returnData = TRUE
)

outcomes_wtp_mxl84_al <- predict(
  mxl_wtp84_al,
  type = "outcome",
  returnData = TRUE
)

```

```r
chosen_pref_mxl84_al <- subset(outcomes_pref_mxl84_al, choice == 1)
chosen_pref_mxl84_al$correct <- chosen_pref_mxl84_al$choice == chosen_pref_mxl84_al$predicted_outcome
accuracy_pref_mxl84_al <- sum(chosen_pref_mxl84_al$correct) / nrow(chosen_pref_mxl84_al)
accuracy_pref_mxl84_al

chosen_wtp_mxl84_al <- subset(outcomes_wtp_mxl84_al, choice == 1)
chosen_wtp_mxl84_al$correct <- chosen_wtp_mxl84_al$choice == chosen_wtp_mxl84_al$predicted_outcome
accuracy_wtp_mxl84_al <- sum(chosen_wtp_mxl84_al$correct) / nrow(chosen_wtp_mxl84_al)
accuracy_wtp_mxl84_al
```

MMNLM R Code

APPENDIX B

Characteristics of Sample Questionnaires

B.1. Demographic Characteristics

Table B.1. Valued Attributes

Question Nº	Demographic Characteristic	Number of participants	% of total
Q1: Gender	Male	42	50%
	Female	41	49%
	Prefer not to answer	1	1%
Q2: Age	0-17	1	1%
	18-24	33	39%
	25-34	19	23%
	35-44	8	10%
	45-54	12	14%
	55-64	7	8%
	65-74	4	5%
Q3: Income	Until 500 €	22	26%
	Between 500 € and 1000 €	10	12%
	Between 1000 € and 1500 €	19	23%
	Between 1500 € and 2000 €	11	13%
	Over 2000 €	22	26%
Q4: Education	Primary education (PT = 4th grade)	0	0%
	Middle school (PT = 9th grade)	1	1%
	Secundary School (PT = 12th grade)	11	13%
	Bachelors degree (PT = Licenciatura)	39	46%
	Masters degree (PT = Mestrado)	21	25%
	PhD degree (PT = Doutoramento)	12	14%

B.2. Valued Attributes

Table B.2. Valued Attributes

Question Nº	Choices	Number of participants	% of total
Q8	Very easy	7	8%
	Easy	32	38%
	Neither hard nor easy	33	39%
	Hard	9	11%
	Very hard	3	4%
Q9	Yes	54	64%
	No	30	36%
Q10	Average Monthly Cost	44	33%
	Recommendation	12	9%
	Rating	11	8%
	Effect on speed reduction	31	23%
	Logging	25	19%
	Number of countries with servers	11	8%

Q8: Difficulty of choices

Q9: In the previous choices, were all attributes taken into account

Q10: If no, which attributes did you value the most (by number of mentions)

B.3. Privacy Literacy Questions

How would you self-evaluate your knowledge regarding the following topics?
Categories of answers – (Extensive, Plenty, Some, Little, None)

(1) What data and internet traffic my internet service provider (e.g. NOS, Vodafone) keeps of me **(D1.1.)**
(2) How my internet service provider treats my data and internet traffic **(D1.1.)**
(3) What data and internet traffic my internet service provider (e.g. NOS, Vodafone) keeps of me **(D1.1.)**
(4) Tools and software to protect privacy online (e.g. TOR, VPN) **(D2.1.)**
(5) Infrastructure and functionality of the internet (e.g., HTML, IP addresses, Cloud services) **(D2.2.)**
(6) Data protection legislation in Europe**(D2.2.)**
(7) The rights I have towards firms or institutions that treat my person data**(D2.2.)**

Have you ever taken any of these data protection and/or privacy strategies?
Categories of answers – (Always, Often, Sometimes, Rarely, Never)

(1) Opting for not divulging personal information when acquiring a service **(D4)**
(2) Use of private navigation in the internet browser **(D4)**
(3) Use of different passwords for different services **(D4)**
(4) Use of a VPN service **(D4)**

(5) Use of the Onion internet browser (TOR) or Brave internet browser **(D4)**

Please answer with the option you consider the most adequate: Categories of answers: (True, False, I don't know)

(1) Individuals have the right of knowing what personal data firms keep of them (e.g. Facebook) **(D3)** - True

(2) Individuals have the right of requesting a firm (e.g. Facebook) to delete their personal data. **(D3)** - True

(3) Firms collect user data through several websites for creation of a profile **(D1.2)** - True

(4) In the browsing history, dangerous websites are recorded differently, according to the browser **(D2.2.)** - False

(5) It's harder to track internet use if a person deletes their browser information (e.g. *cookies*) **(D4)** - True

(6) The use of public WiFi's instead of mobile data provides greater data protection and privacy **((D42.2)** - False

(7) The use of private browsing hides my online activity from my internet service provider **(D1 and D4)** - False

(8) My internet service provider can block my access to certain websites **(D41.1)** - False

APPENDIX C

Additional Models

Below are models calculated, that while relevant for analysis, are not directly analyzed in the body of text:

Table C.1. MNL - Preference estimates per subgroup of privacy literacy

	BLMNL			ALMNL		
	Estimate	Std. Error		Estimate	Std. Error	
price	-0.27	0.03	***	-0.21	0.031	***
recomyes	2.8	0.6	**	2.27	0.61	***
rating 4 to 5	0.01	0.2		-0.34	0.26	
speedred mod to sev	-0.74	0.24	*	-0.89	0.33	**
speedred none to lig	0.63	0.21	*	0.59	0.29	*
logref	1.19	0.19	***	1.38	0.23	***
logref e audit	1.19	0.26	***	2.55	0.38	***
countryless50	-0.06	0.23		-0.21	0.32	
countryplus80	0.43	0.17	***	0.89	0.26	***
Log Likelihood	-270.8			-171.2		
null.logLik	-432.5			-266.2		
AIC	559.5			360.3		
BIC	599.5			395.9		
R$\hat{2}$	0.37			0.36		
Adj. R$\hat{2}$	0.35			0.32		
nobs	624			384		

***p ¡ 0.001; **p ¡ 0.01; *p ¡ 0.05

Table C.2. MMNL - Preference estimates per subgroup of privacy literacy

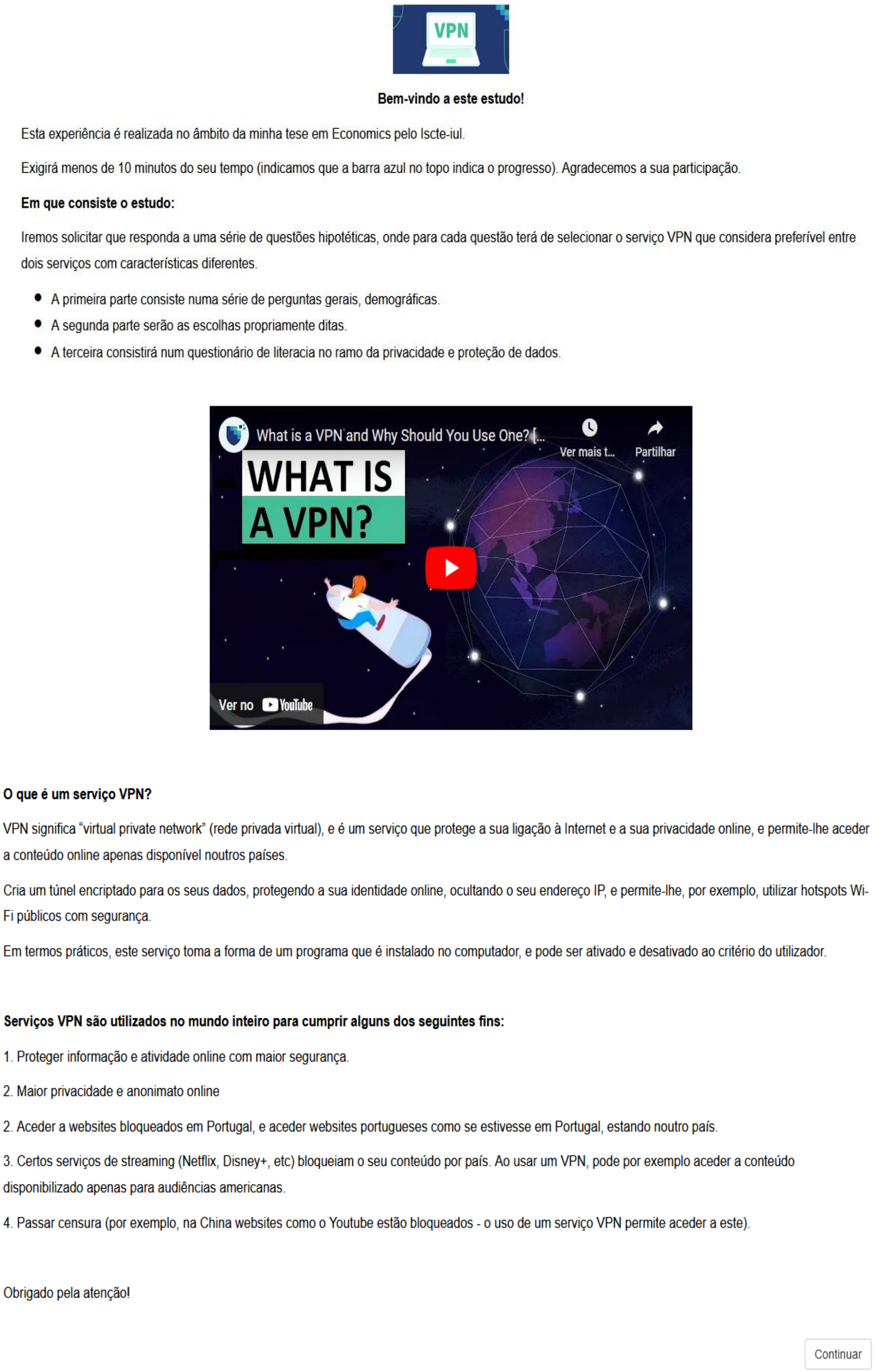
	BLMMNL	ALMMNL
price	-0.32^{***}	-0.26^{***}
	(0.04)	(0.04)
recomyes	2.8	3.48
	(1.76)	(3.92)
rating4 a 5	-0.04	-0.38
	(0.28)	(0.35)
speedred mod to sev	-0.76^{***}	-0.87
	(0.27)	(0.37)
speedred none to lig	-0.67^{**}	0.68
	(0.26)	(0.35)
logRef	1.49^{***}	1.69
	(0.26)	(0.33)
logRef e audit	1.35^{***}	3.16
	(0.26)	(0.63)
countryplus80	0.44^{*}	0.91
	(0.18)	(0.27)
countryless50	-0.12	-0.37
	(0.26)	(0.36)
sd_recomyes	-0.16	-2.0
	(13.3)	(3.97)
sd_rating4 to 5	1.19^{***}	-1.11
	(0.38)	(0.49)
sd_speedred mod a sev	-0.13	(0.01)
	(0.41)	(0.37)
sd_speedred none to lig	0.08	0.10
	(0.64)	(0.46)
sd_logRef	0.63^{***}	-0.63
	(0.37)	(0.45)
sd_logRef e audit	-1.08	1.68
	(0.47)	(0.74)
sd_countryplus80	-0.04	0.014
	(0.36)	(0.31)
sd_countryless50	-0.04	0.04
	(0.44)	(0.69)
Log Likelihood	-266.73	-168.03
null.logLik	-432.52	-266.17
AIC	567.47	-370.06
BIC	642.88	-437.22
R^2	0.38	0.36
Adj. R^2	0.34	0.30
nobs	624.00	-432.52

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

APPENDIX D

DCE Prints

Below we can see the prints of the survey delivered to participants, in Portuguese.



Bem-vindo a este estudo!

Esta experiência é realizada no âmbito da minha tese em Economics pelo Iscte-iul.

Exigirá menos de 10 minutos do seu tempo (indicamos que a barra azul no topo indica o progresso). Agradecemos a sua participação.

Em que consiste o estudo:

Iremos solicitar que responda a uma série de questões hipotéticas, onde para cada questão terá de selecionar o serviço VPN que considera preferível entre dois serviços com características diferentes.

- A primeira parte consiste numa série de perguntas gerais, demográficas.
- A segunda parte serão as escolhas propriamente ditas.
- A terceira consistirá num questionário de literacia no ramo da privacidade e proteção de dados.

O que é um serviço VPN?

VPN significa "virtual private network" (rede privada virtual), e é um serviço que protege a sua ligação à Internet e a sua privacidade online, e permite-lhe aceder a conteúdo online apenas disponível noutros países.

Cria um túnel encriptado para os seus dados, protegendo a sua identidade online, ocultando o seu endereço IP, e permite-lhe, por exemplo, utilizar hotspots Wi-Fi públicos com segurança.

Em termos práticos, este serviço toma a forma de um programa que é instalado no computador, e pode ser ativado e desativado ao critério do utilizador.

Serviços VPN são utilizados no mundo inteiro para cumprir alguns dos seguintes fins:

1. Proteger informação e atividade online com maior segurança.

2. Maior privacidade e anonimato online

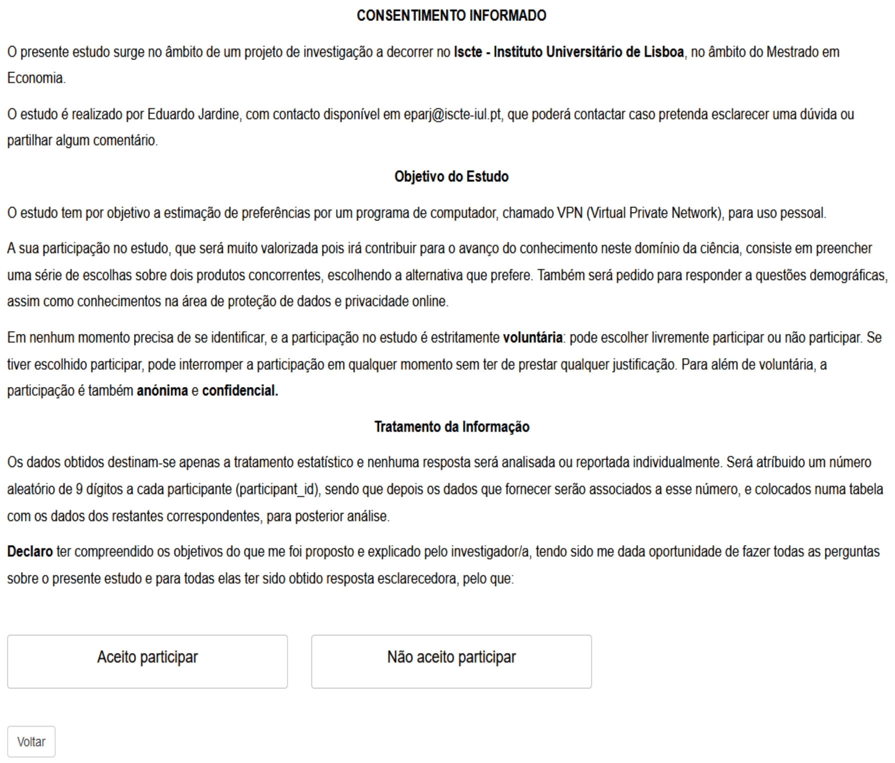
2. Aceder a websites bloqueados em Portugal, e aceder websites portugueses como se estivesse em Portugal, estando noutro país.

3. Certos serviços de streaming (Netflix, Disney+, etc) bloqueiam o seu conteúdo por país. Ao usar um VPN, pode por exemplo aceder a conteúdo disponibilizado apenas para audiências americanas.

4. Passar censura (por exemplo, na China websites como o Youtube estão bloqueados - o uso de um serviço VPN permite aceder a este).

Obrigado pela atenção!

Continuar



CONSENTIMENTO INFORMADO

O presente estudo surge no âmbito de um projeto de investigação a decorrer no **Iscte - Instituto Universitário de Lisboa**, no âmbito do Mestrado em Economia.

O estudo é realizado por Eduardo Jardine, com contacto disponível em eparj@iscte-iul.pt, que poderá contactar caso pretenda esclarecer uma dúvida ou partilhar algum comentário.

Objetivo do Estudo

O estudo tem por objetivo a estimação de preferências por um programa de computador, chamado VPN (Virtual Private Network), para uso pessoal.

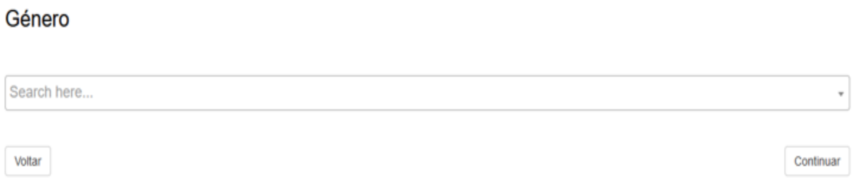
A sua participação no estudo, que será muito valorizada pois irá contribuir para o avanço do conhecimento neste domínio da ciência, consiste em preencher uma série de escolhas sobre dois produtos concorrentes, escolhendo a alternativa que prefere. Também será pedido para responder a questões demográficas, assim como conhecimentos na área de proteção de dados e privacidade online.

Em nenhum momento precisa de se identificar, e a participação no estudo é estritamente **voluntária**: pode escolher livremente participar ou não participar. Se tiver escolhido participar, pode interromper a participação em qualquer momento sem ter de prestar qualquer justificação. Para além de voluntária, a participação é também **anónima** e **confidencial.**

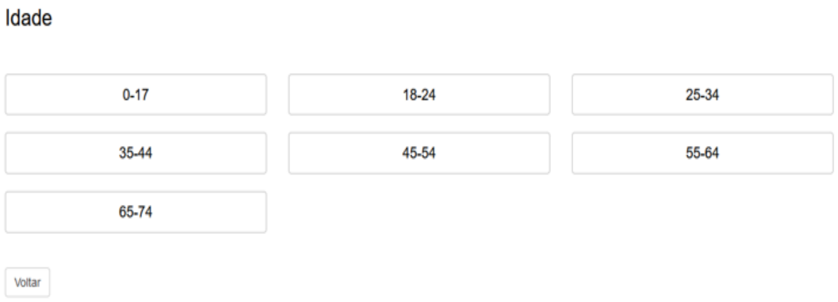
Tratamento da Informação

Os dados obtidos destinam-se apenas a tratamento estatístico e nenhuma resposta será analisada ou reportada individualmente. Será atríbuido um número aleatório de 9 dígitos a cada participante (participant_id), sendo que depois os dados que fornecer serão associados a esse número, e colocados numa tabela com os dados dos restantes correspondentes, para posterior análise.

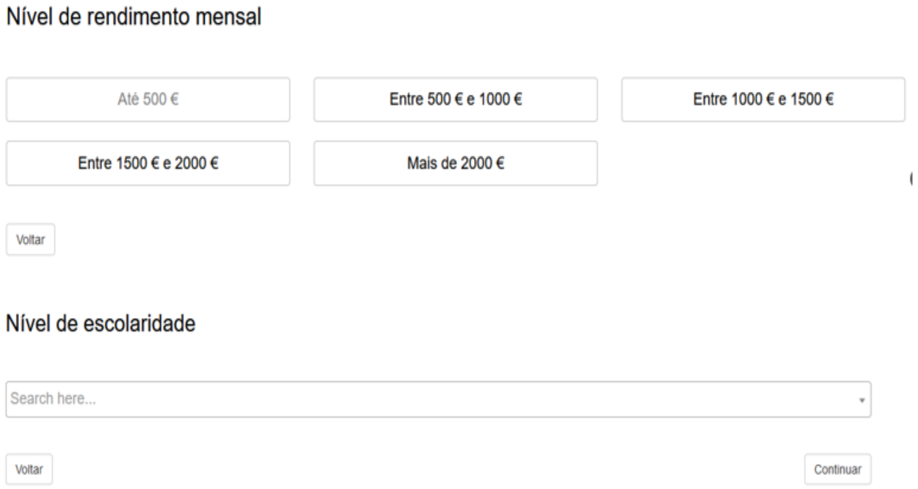
Declaro ter compreendido os objetivos do que me foi proposto e explicado pelo investigador/a, tendo sido me dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter sido obtido resposta esclarecedora, pelo que:



Género



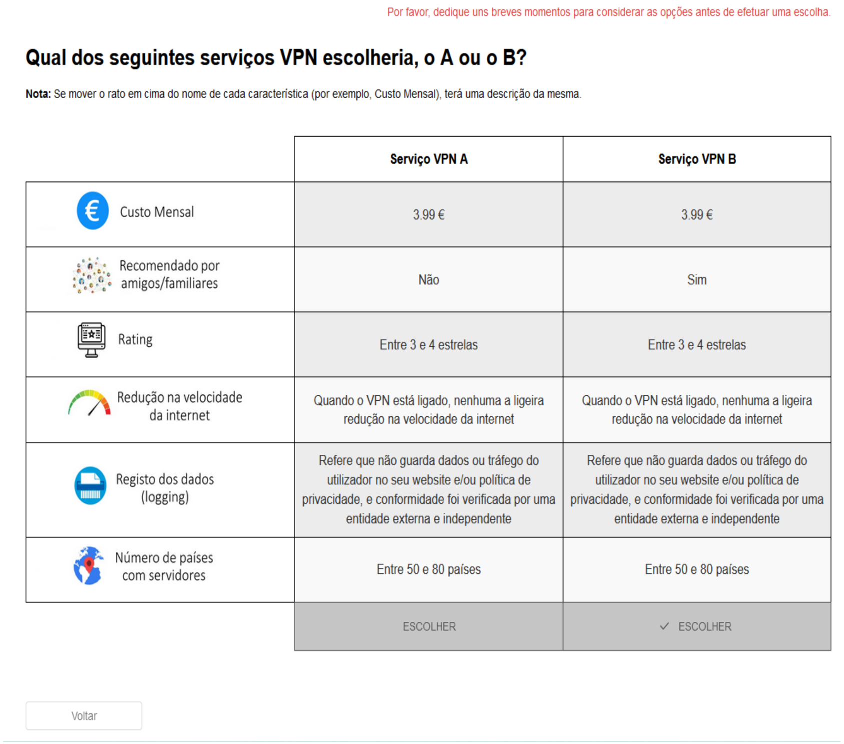
Idade



Nível de rendimento mensal

Nível de escolaridade

~



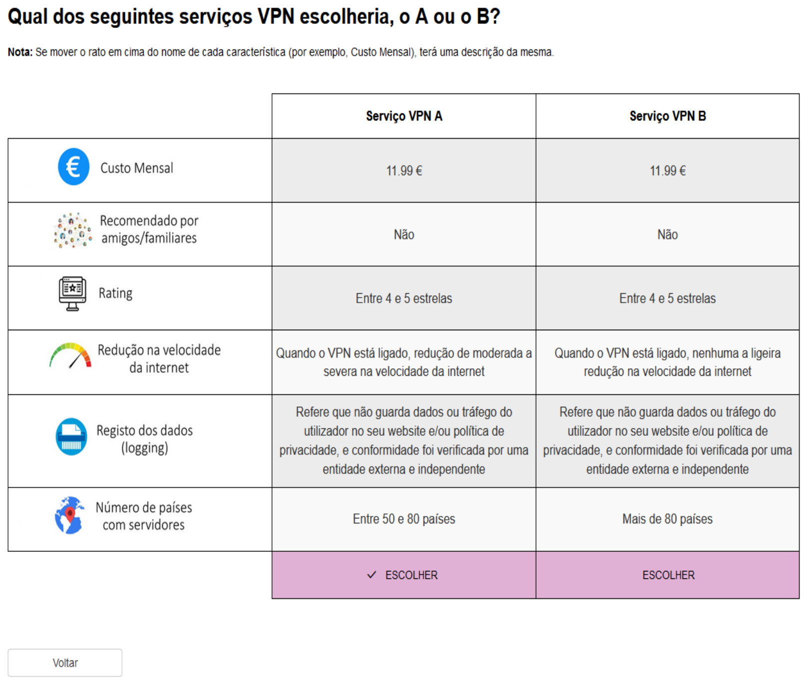
Por favor, dedique uns breves momentos para considerar as opções antes de efetuar uma escolha.

Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
Custo Mensal	3.99 €	3.99 €
Recomendado por amigos/familiares	Não	Sim
Rating	Entre 3 e 4 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente
Número de países com servidores	Entre 50 e 80 países	Entre 50 e 80 países
	ESCOLHER	✓ ESCOLHER

Voltar

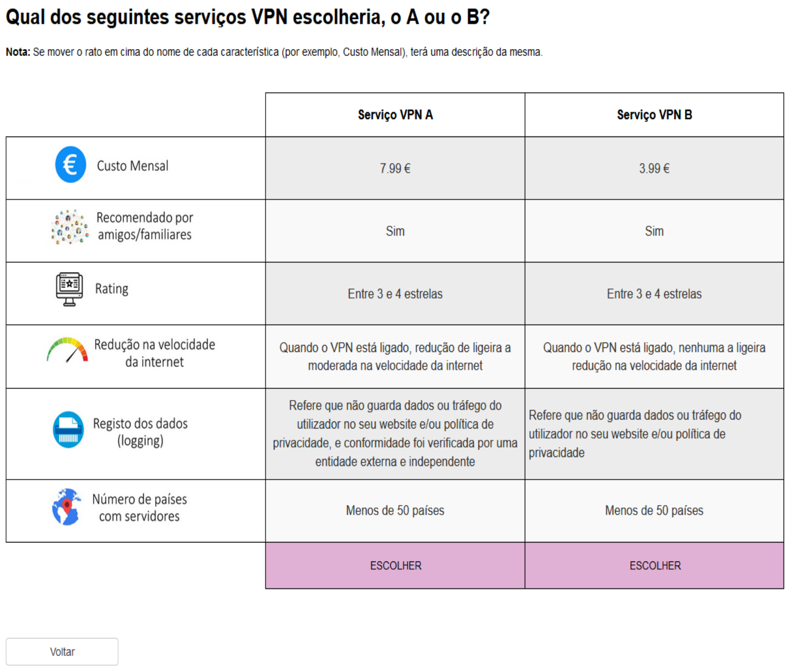


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
Custo Mensal	11.99 €	11.99 €
Recomendado por amigos/familiares	Não	Não
Rating	Entre 4 e 5 estrelas	Entre 4 e 5 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, redução de moderada a severa na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente
Número de países com servidores	Entre 50 e 80 países	Mais de 80 países
	✓ ESCOLHER	ESCOLHER

Voltar

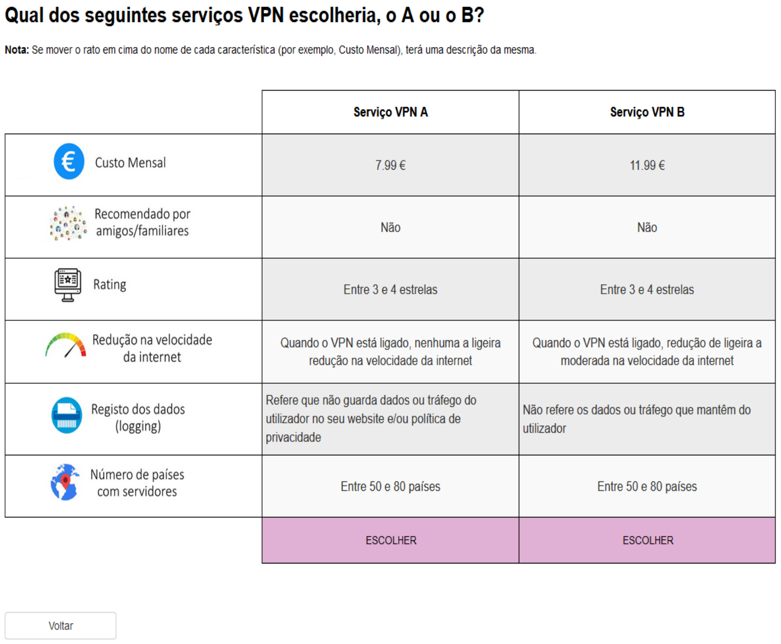


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
Custo Mensal	7.99 €	3.99 €
Recomendado por amigos/familiares	Sim	Sim
Rating	Entre 3 e 4 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade
Número de países com servidores	Menos de 50 países	Menos de 50 países
	ESCOLHER	ESCOLHER

Voltar

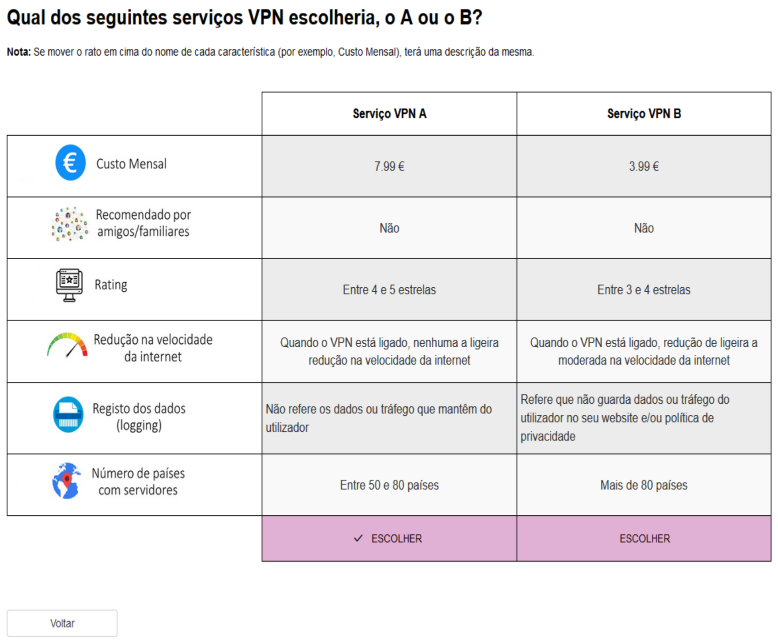


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
Custo Mensal	7.99 €	11.99 €
Recomendado por amigos/familiares	Não	Não
Rating	Entre 3 e 4 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade	Não refere os dados ou tráfego que mantêm do utilizador
Número de países com servidores	Entre 50 e 80 países	Entre 50 e 80 países
	ESCOLHER	ESCOLHER

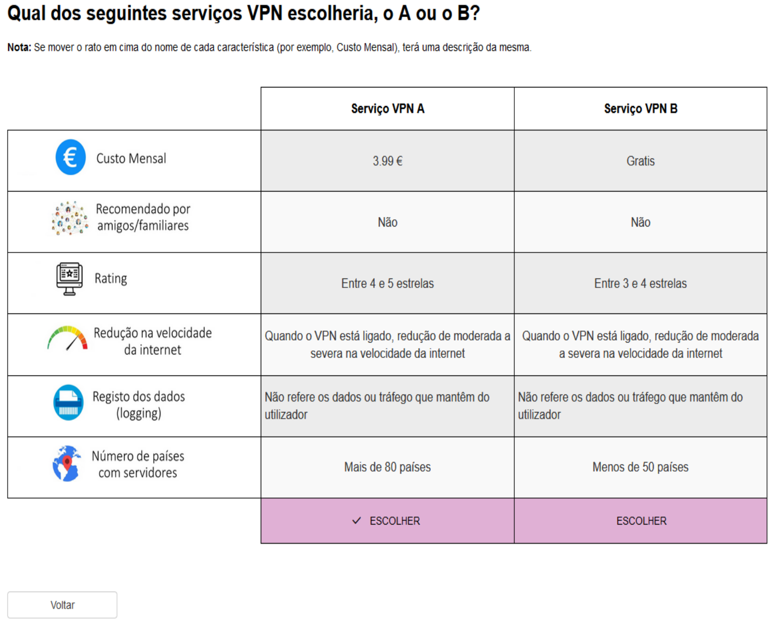
Voltar



Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

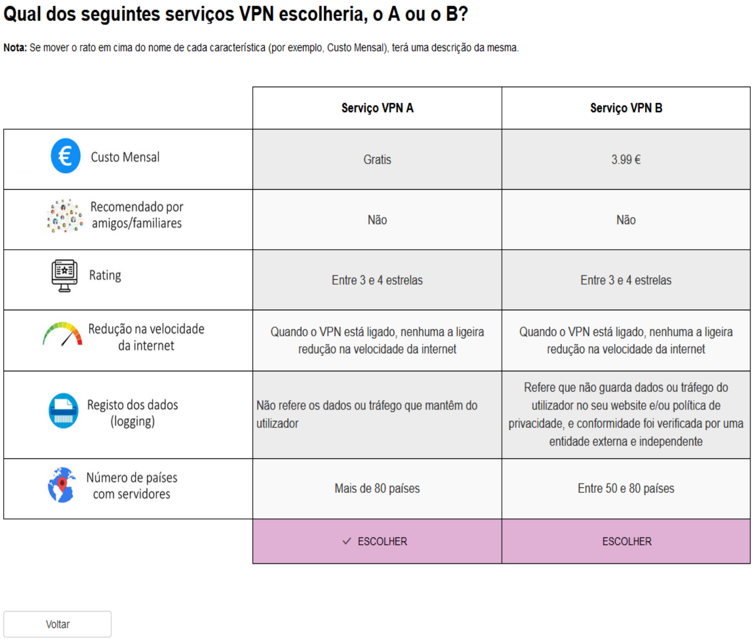
	Serviço VPN A	Serviço VPN B
Custo Mensal	7.99 €	3.99 €
Recomendado por amigos/familiares	Não	Não
Rating	Entre 4 e 5 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet
Registo dos dados (logging)	Não refere os dados ou tráfego que mantêm do utilizador	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade
Número de países com servidores	Entre 50 e 80 países	Mais de 80 países
	✓ ESCOLHER	ESCOLHER



Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
Custo Mensal	3.99 €	Gratis
Recomendado por amigos/familiares	Não	Não
Rating	Entre 4 e 5 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, redução de moderada a severa na velocidade da internet	Quando o VPN está ligado, redução de moderada a severa na velocidade da internet
Registo dos dados (logging)	Não refere os dados ou tráfego que mantêm do utilizador	Não refere os dados ou tráfego que mantêm do utilizador
Número de países com servidores	Mais de 80 países	Menos de 50 países
	✓ ESCOLHER	ESCOLHER

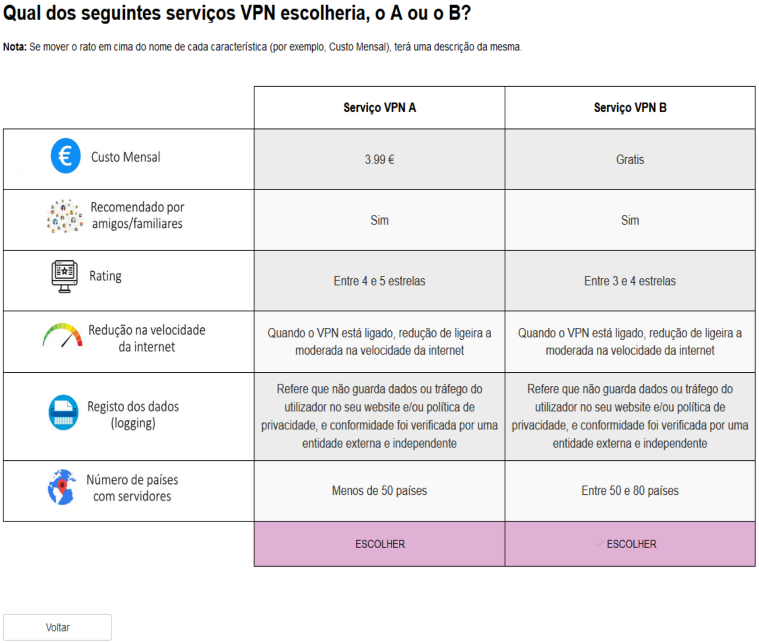


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	**Serviço VPN B**
Custo Mensal	Gratis	3.99 €
Recomendado por amigos/familiares	Não	Não
Rating	Entre 3 e 4 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet
Registo dos dados (logging)	Não refere os dados ou tráfego que mantêm do utilizador	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente
Número de países com servidores	Mais de 80 países	Entre 50 e 80 países
	✓ ESCOLHER	ESCOLHER

Voltar

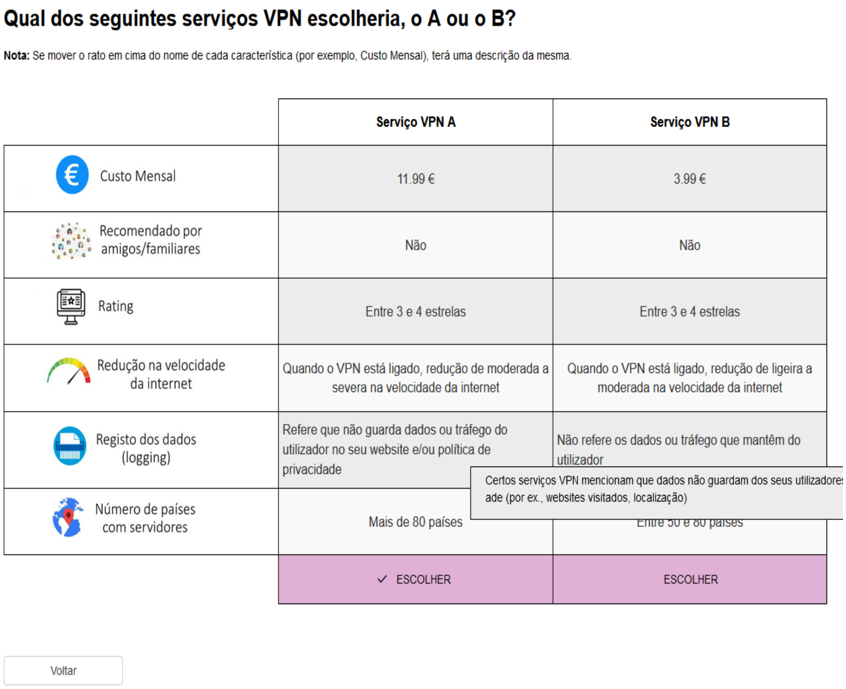


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	**Serviço VPN B**
Custo Mensal	3.99 €	Gratis
Recomendado por amigos/familiares	Sim	Sim
Rating	Entre 4 e 5 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente
Número de países com servidores	Menos de 50 países	Entre 50 e 80 países
	ESCOLHER	✓ ESCOLHER

Voltar

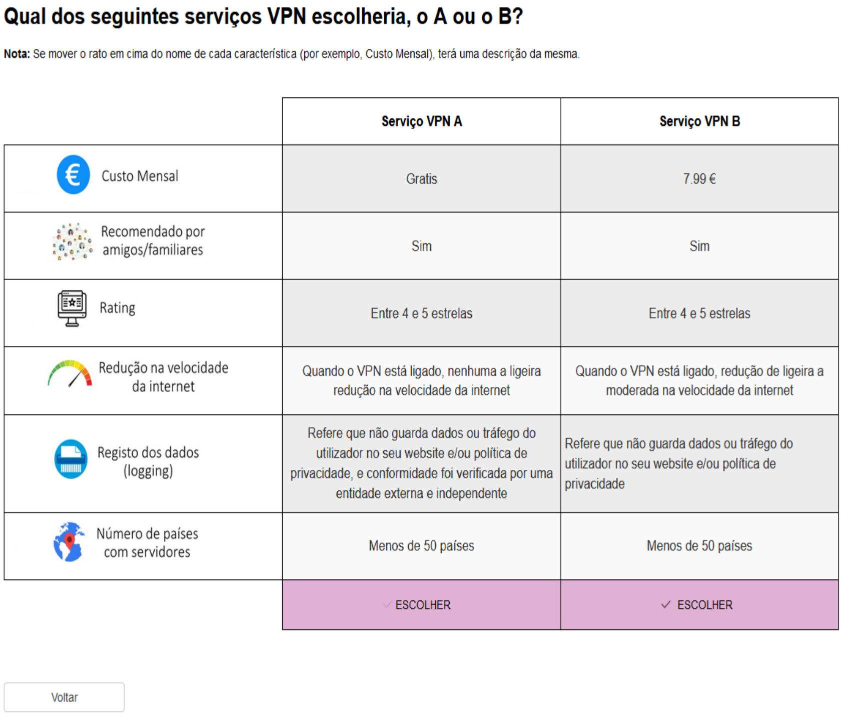


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
€ Custo Mensal	11.99 €	3.99 €
Recomendado por amigos/familiares	Não	Não
Rating	Entre 3 e 4 estrelas	Entre 3 e 4 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, redução de moderada a severa na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade	Não refere os dados ou tráfego que mantêm do utilizador
Número de países com servidores	Mais de 80 países	Entre 50 e 80 países
	✓ ESCOLHER	ESCOLHER

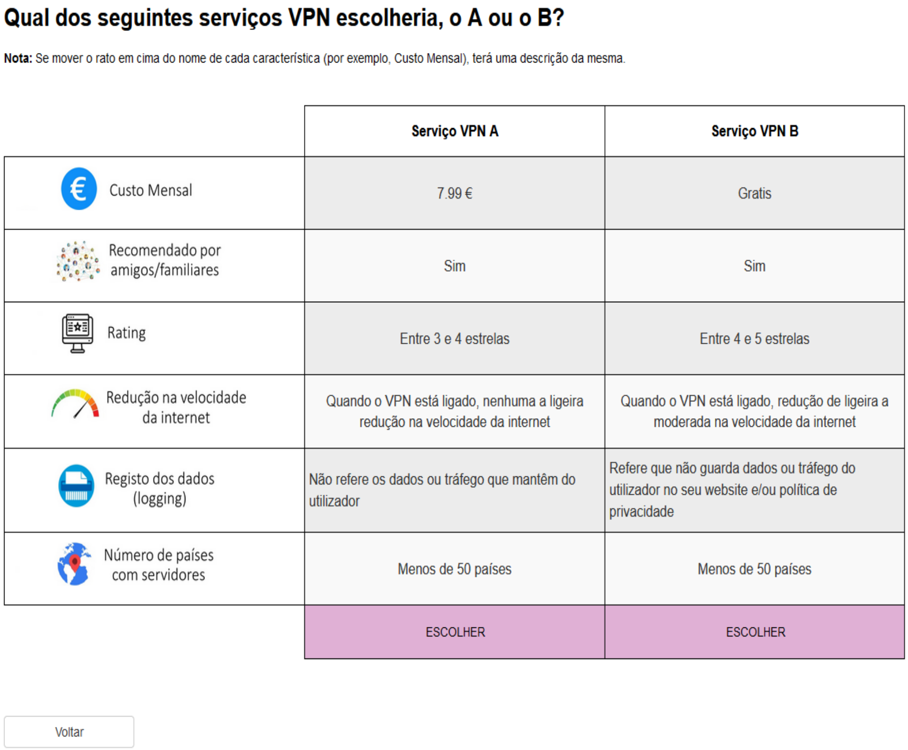
Certos serviços VPN mencionam que dados não guardam dos seus utilizadores
ade (por ex., websites visitados, localização)



Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	Serviço VPN B
€ Custo Mensal	Gratis	7.99 €
Recomendado por amigos/familiares	Sim	Sim
Rating	Entre 4 e 5 estrelas	Entre 4 e 5 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade, e conformidade foi verificada por uma entidade externa e independente	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade
Número de países com servidores	Menos de 50 países	Menos de 50 países
	✓ ESCOLHER	✓ ESCOLHER

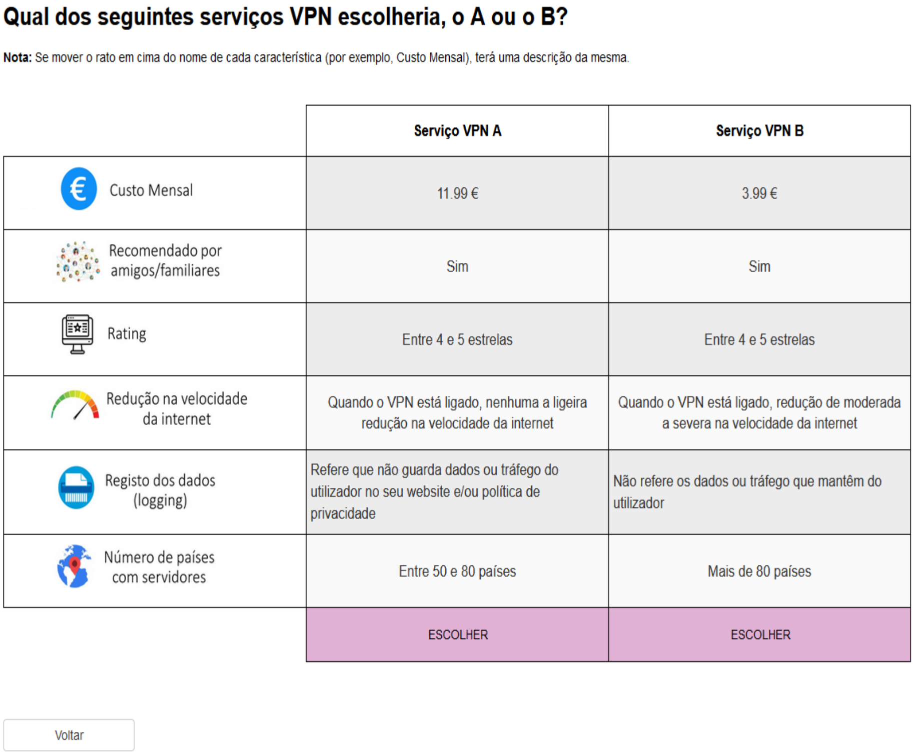


Qual dos seguintes serviços VPN escolheria, o A ou o B?

Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	**Serviço VPN B**
Custo Mensal	7.99 €	Gratis
Recomendado por amigos/familiares	Sim	Sim
Rating	Entre 3 e 4 estrelas	Entre 4 e 5 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, redução de ligeira a moderada na velocidade da internet
Registo dos dados (logging)	Não refere os dados ou tráfego que mantêm do utilizador	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade
Número de países com servidores	Menos de 50 países	Menos de 50 países
	ESCOLHER	ESCOLHER

Voltar

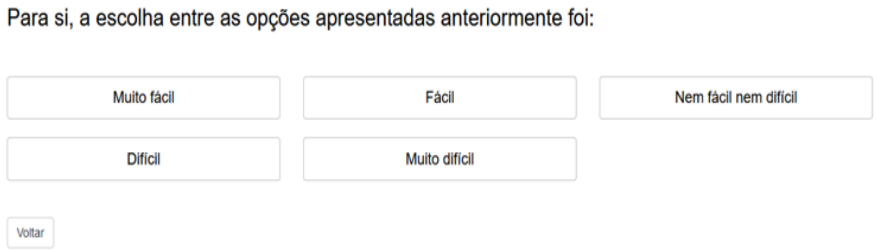


Qual dos seguintes serviços VPN escolheria, o A ou o B?

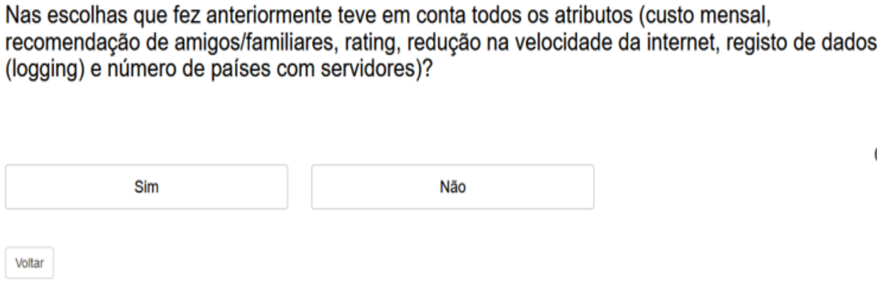
Nota: Se mover o rato em cima do nome de cada característica (por exemplo, Custo Mensal), terá uma descrição da mesma.

	Serviço VPN A	**Serviço VPN B**
Custo Mensal	11.99 €	3.99 €
Recomendado por amigos/familiares	Sim	Sim
Rating	Entre 4 e 5 estrelas	Entre 4 e 5 estrelas
Redução na velocidade da internet	Quando o VPN está ligado, nenhuma a ligeira redução na velocidade da internet	Quando o VPN está ligado, redução de moderada a severa na velocidade da internet
Registo dos dados (logging)	Refere que não guarda dados ou tráfego do utilizador no seu website e/ou política de privacidade	Não refere os dados ou tráfego que mantêm do utilizador
Número de países com servidores	Entre 50 e 80 países	Mais de 80 países
	ESCOLHER	ESCOLHER

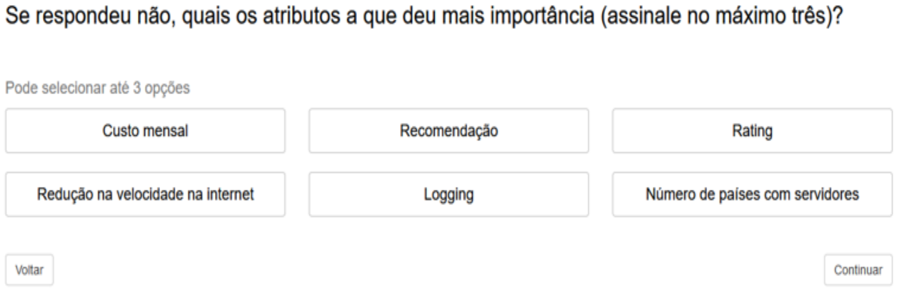
Voltar



Para si, a escolha entre as opções apresentadas anteriormente foi:



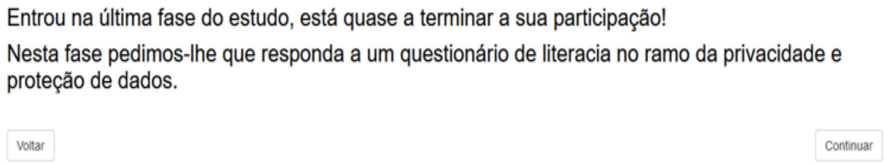
Nas escolhas que fez anteriormente teve em conta todos os atributos (custo mensal, recomendação de amigos/familiares, rating, redução na velocidade da internet, registo de dados (logging) e número de países com servidores)?



Se respondeu não, quais os atributos a que deu mais importância (assinale no máximo três)?

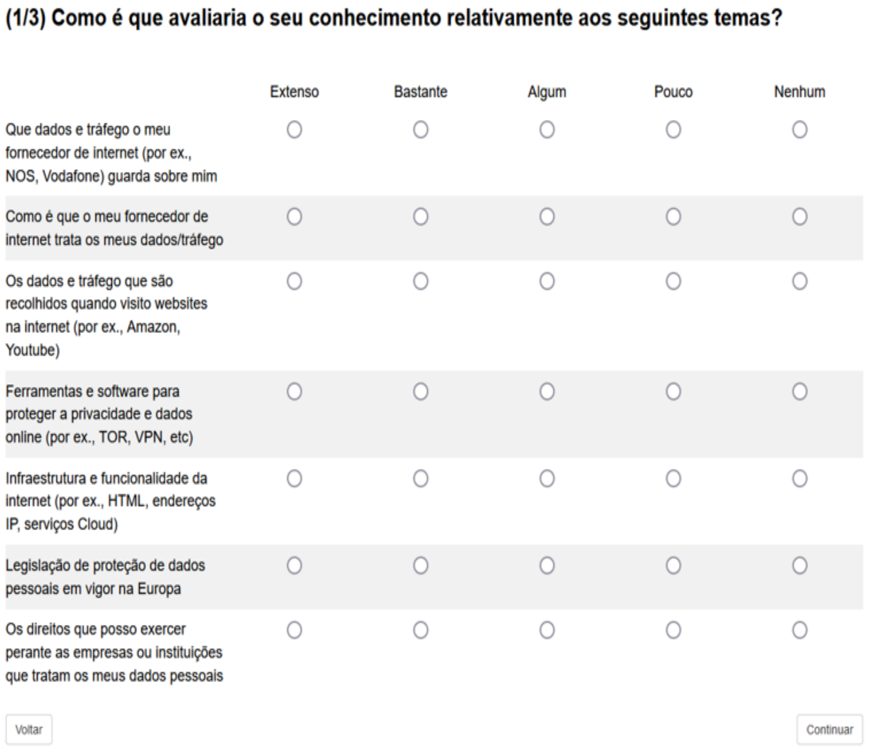
Pode selecionar até 3 opções

Custo mensal
Recomendação
Rating
Redução na velocidade na internet
Logging
Número de países com servidores
Voltar
Continuar



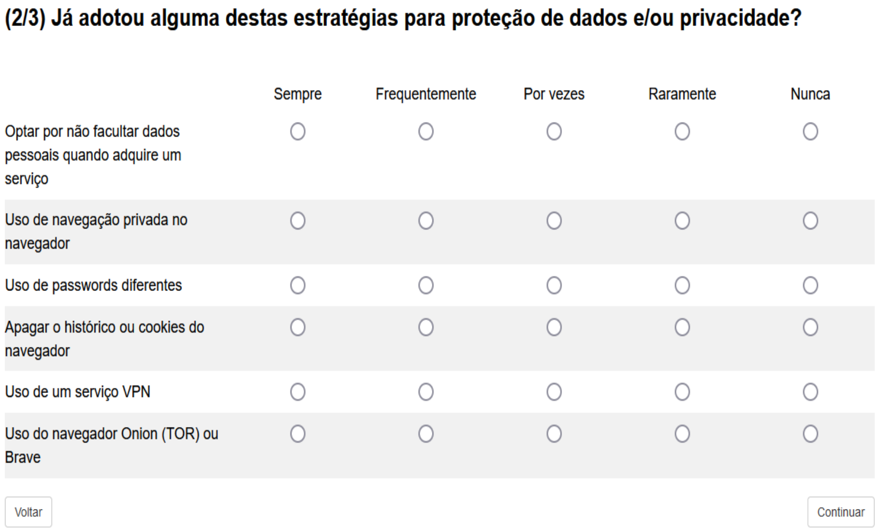
Entrou na última fase do estudo, está quase a terminar a sua participação!

Nesta fase pedimos-lhe que responda a um questionário de literacia no ramo da privacidade e proteção de dados.



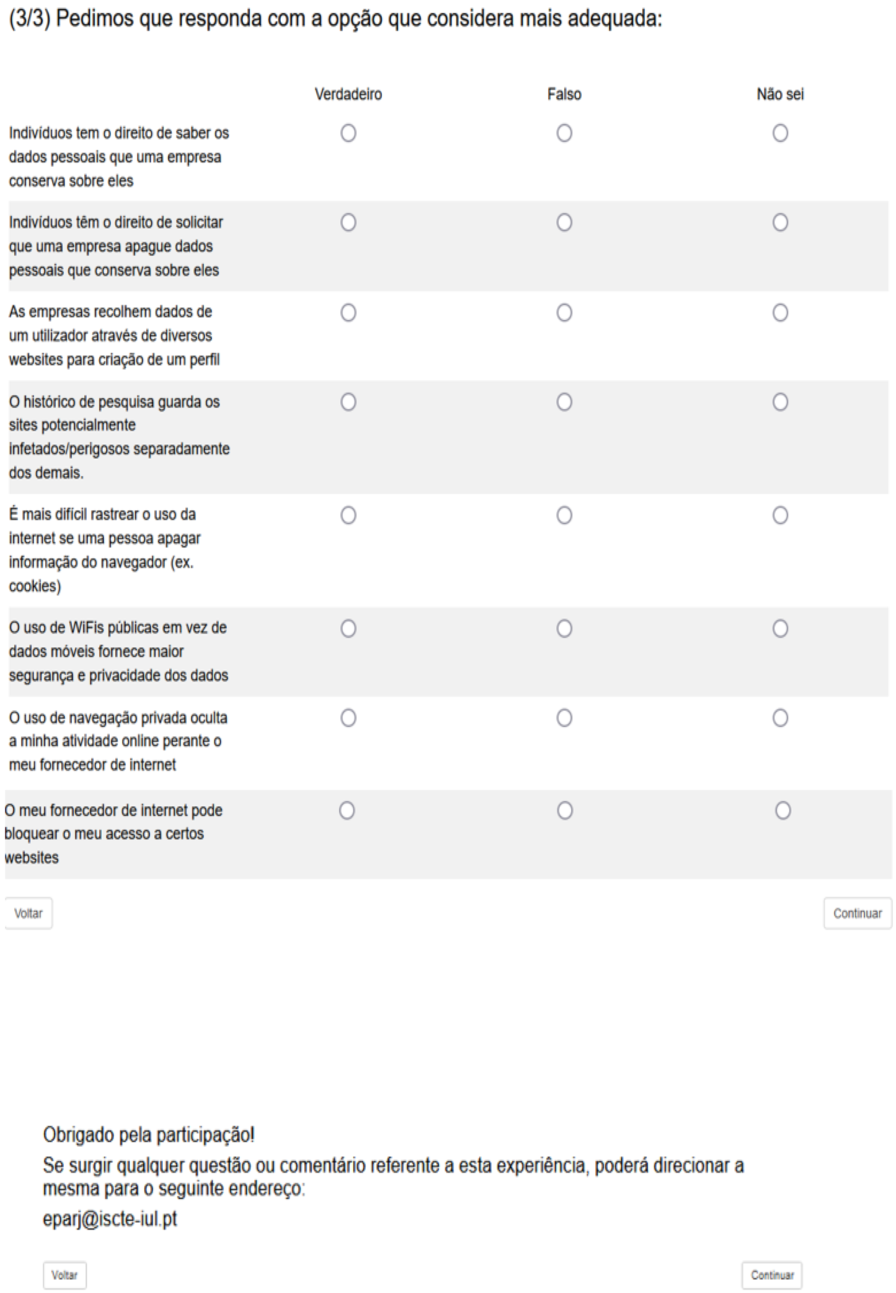
(1/3) Como é que avaliaria o seu conhecimento relativamente aos seguintes temas?

	Extenso	Bastante	Algum	Pouco	Nenhum
Que dados e tráfego o meu fornecedor de internet (por ex., NOS, Vodafone) guarda sobre mim	○	○	○	○	○
Como é que o meu fornecedor de internet trata os meus dados/tráfego	○	○	○	○	○
Os dados e tráfego que são recolhidos quando visito websites na internet (por ex., Amazon, Youtube)	○	○	○	○	○
Ferramentas e software para proteger a privacidade e dados online (por ex., TOR, VPN, etc)	○	○	○	○	○
Infraestrutura e funcionalidade da internet (por ex., HTML, endereços IP, serviços Cloud)	○	○	○	○	○
Legislação de proteção de dados pessoais em vigor na Europa	○	○	○	○	○
Os direitos que posso exercer perante as empresas ou instituições que tratam os meus dados pessoais	○	○	○	○	○



(2/3) Já adotou alguma destas estratégias para proteção de dados e/ou privacidade?

	Sempre	Frequentemente	Por vezes	Raramente	Nunca
Optar por não facultar dados pessoais quando adquire um serviço	○	○	○	○	○
Uso de navegação privada no navegador	○	○	○	○	○
Uso de passwords diferentes	○	○	○	○	○
Apagar o histórico ou cookies do navegador	○	○	○	○	○
Uso de um serviço VPN	○	○	○	○	○
Uso do navegador Onion (TOR) ou Brave	○	○	○	○	○



(3/3) Pedimos que responda com a opção que considera mais adequada:

	Verdadeiro	Falso	Não sei
Indivíduos tem o direito de saber os dados pessoais que uma empresa conserva sobre eles	○	○	○
Indivíduos têm o direito de solicitar que uma empresa apague dados pessoais que conserva sobre eles	○	○	○
As empresas recolhem dados de um utilizador através de diversos websites para criação de um perfil	○	○	○
O histórico de pesquisa guarda os sites potencialmente infetados/perigosos separadamente dos demais.	○	○	○
É mais difícil rastrear o uso da internet se uma pessoa apagar informação do navegador (ex. cookies)	○	○	○
O uso de WiFis públicas em vez de dados móveis fornece maior segurança e privacidade dos dados	○	○	○
O uso de navegação privada oculta a minha atividade online perante o meu fornecedor de internet	○	○	○
O meu fornecedor de internet pode bloquear o meu acesso a certos websites	○	○	○

Obrigado pela participação!

Se surgir qualquer questão ou comentário referente a esta experiência, poderá direcionar a mesma para o seguinte endereço:

eparj@iscte-iul.pt

Figure D.1. Prints of Deployed DCE Experiment in Conjointly