CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

BRIDGING HEALTH DATA DONATION

The Impact of Private data organizations and Privacy Calculus on donation decisions

Master Thesis in Business Administration & Innovation in Healthcare

Pratyush Khanra (142886)

Date: 16th January 2023

Supervisor: Prof. Tawfiq Alashoor

Number of pages: 72
Character count: 139412

Abstract

As healthcare becomes increasingly digitized the promise to make use of health data is addressed by multiple players. Data donation is one such concept which provides individuals pathways to systematically donate their health data for health research. While such drives are increasingly beneficial for society, they also have privacy risks associated with them. A necessity also arises for individuals, government and private organizations to obtain a greater understanding and knowledge of privacy behaviour and decision making to maximize voluntary disclosures. To understand the underlying mechanism, we examine privacy calculus and explore it on a societal level. The study also distinguishes data donation organizations into government and private which very few studies have done and identifies individuals' attitudes towards each of them. To study their overall impact on data donation decisions, a 2 (government vs. private) x 2 (net negative privacy calculus outcome vs. net positive privacy calculus outcome) factorial experiment was conducted based on a mock data donation scenario. The main dependent outcome of the experiment was data donation score. The results of the experiments revealed that individuals are willing to donate more health data if they perceived a net positive privacy calculus, these individuals also had lesser privacy concerns. Individuals had lower and almost similar levels of data donation for both the private and government organizations if net privacy calculus was negative. When net privacy calculus was positive individuals preferred to significantly donate more data to a private organization than a government one. The study also found that private organizations have an interaction effect on the net privacy calculus which affects data donation scores. The results of the study have implications that societal benefits should be taken into consideration when encouraging people to donate health data. Additionally private organizations should work on negating risks associated with them while government organizations should work on creating assurances in the general population about the benefits associated with data donation. The study therefore contributes with new knowledge and understanding of privacy behavior furthermore, it also contributes to the challenges faced in healthcare by identifying decisions which would make data donation more desirable for the individual and improve data driven healthcare.

Acknowledgements

First of all, would I like to thank my supervisor Prof. Tawfiq Alashoor for his guidance and encouragement throughout the whole journey of writing the thesis. I am very grateful for his professional and academic knowledge that broadened my understanding of such a crucial topic in healthcare. Thank you for helping me manoeuvre through situations when I felt stuck on certain topics.

I would also like to mention the impact of the professors and supervisors at Copenhagen Business School had on me through their teaching and learnings. The thesis has been a culmination of every bit of knowledge gained through my studies.
Lastly, I want to thank my family and friends for patiently supporting us throughout our studies and especially this thesis.

Table of Contents

List of figures

List of tables

1. Introduction

Globally healthcare is being transformed significantly by a growing number of trends relating to medical research, patient experience, prediction, and prevention. These trends driving healthcare transformation forward are backed by one fundamental force: the power of data (Stanford Medicine Health Trends, 2017). In fact, "The world's most valuable resource is no longer oil, but data" (The Economist, 2017). Traditionally public systems like the World Health Organization (WHO), National Health Service (NHS), research centers and the medical industry have been old users of healthcare data, however recently, major corporations like Google, Meta (Facebook), Amazon, Apple, and Microsoft are doubling down on healthcare. Therefore, the question that arises is: *'What makes health data so attractive?'*

The global healthcare budget is expected to be an astounding $15 trillion by 2030 (RBC Capital Markets, n.d.) encasing electronic health records, clinical data, insurance, wearables and sensors. By 2025, the compound annual growth rate of healthcare data will reach 36% which is more than any other industry. (RBC Capital Markets, n.d.). Healthcare represents a large pie and data driven organizations can bite onto it thus benefiting from increased personalization, productivity and enhanced revenues (Marr B, 2016).

The information age also provides us opportunity to disclose health data through wearables or platforms. Society benefits significantly from big data as seen during the Covid-19 pandemic. Having such datasets is fundamental for real time research and policy responses to pandemics (Dhami et al., 2022). A novel way to make use of such valuable data is through Data donation which provides individuals pathway to systematically donate their health data for health research. While data donation is a subject of widespread debate the term is still largely unfamiliar with majority of the population. Health data donated can help identify patients at risk before they even present any symptoms, reduces scheduling time in hospitals and promotes research and development into new treatment and drugs. Most importantly data donation can help resolve some major problem within by promoting a standardized consensual approach where individuals have more control over what and how much they

want to share. Data donation organizations are far and few and include nonprofits like DataforGood [1], Open Humans[2] along with bigger companies like Pfizer[3]. Organizations can build trust with individuals by being transparent with the purpose of the data and its consequences (Acquisti et al., 2016). Further individuals also have inherent tendencies to donate data for social duty over personal self-serving motives (Skatova & Goulding, 2019). With lifesaving benefits also comes the potential risks and mishandling of such data. Most notable are the Tricare data breach in 2011 of 5 million patient records (Kost, 2022) and the Facebook breach in 2021 of 87 million users which called for tighter laws for privacy (Tech Republic, 2020). Healthcare, therefore, provides the perfect platform to study privacy concerns due to multiple reasons which includes its highly sensitive nature, different types of health data, the number of stakeholders involved and emotions involved with the information (Anderson & Agarwal, 2011).

The dilemma of disclosing data for the greater good or facing the consequences of disclosure gave rise to a highly researched topic in privacy research – The Privacy calculus. The privacy calculus states that people will disclose personal information on a tradeoff i.e., when the benefits exceed the cost (Culnan & Armstrong, 1999). Literature on privacy calculus reveals an existing gap focusing on the societal cost and benefits of health information disclosure as most of the literature is based on personal cost and benefit. As costs and benefits always coexist when making decisions, costs in this study context would imply a net negative outcome and benefits a net positive outcome. Addressing data donation challenges through the lens of privacy literature would highlight influencers and inhibitors of data sharing and help overcome barriers for such drives from both the individuals point of view and the data requesting organization. Prior research on influences relating to personal health data disclosure reveals complexity with its dependence on various situational factors and therefore also needs to be explored in more specific contexts. The factors explored in this study are the data requesting organization i.e., government and private. Thus, the study aims to examine privacy calculus from the societal cost and benefit aspect and how the requesting organization influences health data donation. The study aims to

[1] https://dataforgoodfoundation.com/en/
[2] https://www.openhumans.org/
[3] https://www.pfizer.com/news/articles/what_if_you_could_donate_your_data_for_research

contribute to the less researched domain of health data privacy and donation by examining the following research question:

How does the type of data requesting organization in combination with societal privacy calculus impact health data donation decisions?

The research question implies that an individual's willingness to donate health data is dependent on the type of data requesting organization (government vs. private). The requesting organization in combination with the privacy calculus (net negative privacy calculus vs. net positive privacy calculus) will reveal different levels of privacy behavior and data donation decision making. It also implies that not only do subjective factors influence decisions, but external factors also account for the overall decision-making process. To answer the research question a factorial randomized experiment was conducted having a 2 (societal cost vs. societal benefit) x 2 (government vs. private) design. The study expects people exposed to societal benefits to show more willingness to donate health data whereas individuals exposed to societal costs show less willingness to donate health data. The study also expects people to be more willing to donate data if the requesting organization is a government one. Further the data donation organization is also expected to interact with privacy calculus to influence donation decisions as each are perceived to have different privacy concern level.

To build the framework for the experiment a thorough literature review was conducted with the purpose of discussing privacy literature and health data donation as found in *Section 2* and *3*. *Section 4* contains the experimental framework based on the literature review and theoretical background. It also contains the hypotheses in the study. Section 5 shows the pathway of the experiment including the tools and manipulation checks. Section 6 discusses the ANOVA, t-test and regression statistical analysis methods used along with the tables and results of the analysis. Lastly, section 8 and 9 discuss the results in association with their theoretical and practical implications and ends with a conclusion.

2. Literature Review

2.1 Privacy

The term 'Privacy' has had considerable confusion attached to its meaning, value and scope. Although used frequently there seems to be no universal definition of it and it remains a topical question. Till date there are differing definitions and concepts of privacy as its meaning for every individual is different since its value is highly dependent on the context of it (Acquisti et al., 2018). Historically the appearance of "real" privacy appeared during the 19th century when people moved from the watchful eyes of people in small tribes into the bigger spacious cities. (Lukács, 2016). In 1890's Warren and Brandeis article *The Right to Privacy* defined privacy as the "right to be let alone" (Warren & Brandeis, 1890). The article influenced laws, especially in the US and became a groundwork for privacy as it ensured protection against unwanted disclosure of private thoughts, facts and emotions (Lukács, 2016). Despite privacy being a universal claim no universal definition of privacy could be created as its form differs according to the culture, economic environment and society (Witti & Konstantas, 2019). Solove (2002) explained that most definitions of privacy are too narrow highlighting some aspects of privacy or too broad. He created six categories of privacy definitions: 1) The right to be left alone, 2) limited access to self, 3) secrecy, 4) the control of personal information 5) Identity, and 6) Intimacy. One definition of privacy encasing all the above definitions was given by Szabo who states that "privacy is the right of an individual to decide for herself/himself" (Lukács, 2016). As privacy needs to be interpreted according to socio-economic structures the legal notion of privacy seems impossible according to Lukács (2016). International legislations acknowledge the right to privacy despite uncertainties, for e.g. The HIPAA[4] privacy context states that privacy pertains to the collection, storage, usage and use of personal information and addresses the question of who has access to personal information and under what condition (Sharyl J. Nass et al., 2019) whereas the GDPR[5] privacy context pertains to protection of personal data where personal data is 'any information that relates to an individual who can be directly or indirectly identified.' (gdpr.eu, n.d.). The question that arises is that:

[4] https://www.hhs.gov/hipaa/index.html
[5] https://gdpr.eu/

If the subject of the protection cannot be determined exactly how/whether an effective legal protection can be ensured? (Lukács, 2016).

The concept of privacy is multifaceted. While privacy is often defined with safeguarding and protection, privacy has value even in the absence of harm or embarrassment (Sharyl et al., 2019). Having control of who knows what about us can allow us to alter our behavior and control our social relationships (Rachels, 1975). Often terms like 'security,' 'confidentiality' and 'data protection' are used interchangeably with privacy; hence it is important to define them in the study: *Security* can be defined as "the procedural and technical measures required (a) to prevent unauthorized access, modification, use, and dissemination of data stored or processed in a computer system, (b) to prevent any deliberate denial of service, and (c) to protect the system in its entirety from physical harm. *Confidentiality* addresses the issue of keeping information exchanged within a relationship from being disclosed to external parties. It safeguards information that is gathered in the context of an intimate relationship. (Westin, 1976). On the other hand, *Data protection* is a set of strategies and processes used to secure the privacy, availability, and integrity of data (cloudian, n.d.). Privacy is a more abstract right while the right to data protection has a detailed regulation, with definitions, principles, dispositions (Lukács, 2016). Often data protection and privacy have overlapping contents. Data protection is wider, as it applies to all kinds of personal data processing, even when privacy is not infringed. It is also more specific as not all data processing is related to the privacy of an individual. Privacy is also wider and more specific, as it might apply to the processing of not personal data, but still influencing privacy; while it can apply to all data processing but does not interfere with the individual's privacy (Gellert & Gutwirth, 2013).

2.2 Health data

To understand health data, we need to understand what it contains. Any information that relates to the physical or mental health of an individual, or to the provision of health services to the individual can be categorized as Health/Medical data (Kitsos & Pappa, 2015). An individual's medical data can also include diagnosis from the information make-up of a health system. Therefore, health data implies all information pertaining to an individual's medical history, records, and other personal information

(Girdhari & Ndayizigamiye, 2022). On the surface, data from healthcare might appear similar to other privacy dilemmas in which individuals engage in decision making to assess the cost vs. benefit, (Anderson & Agarwal, 2011), however, healthcare is unique in two aspects: 1) emotions are linked to one's medical state and 2) the nature and variety of risks inherent in the compromise of sensitive health data (Trumbo et al., 2007). The digitization of the healthcare industry has moved it from paper records to disruptive innovations like electronic health records into a future where medicine and care are increasingly personalized (Glaser et al., 2008). The volume and scope of personal health data is increasing massively from hospitals, social media, devices and wearables (Anderson & Agarwal, 2011). Dron et al. (2022) found that during the first 600 days of the pandemic (September 24, 2021), there were 5951 peer-reviewed publications related to real-world data and by July 7, 2022, there had been 13395 publications of real-world evidence on COVID-19 highlighting the sheer increase of data. Healthcare can envision the same standards of how products and services are personalized for each individual as in the commercial industry (Awad & Krishnan, 2006). Access and analysis of large volumes of data that capture real-time patient records of routine clinical care have resulted in better disease surveillance and produced evidence to inform public-health decisions (Dron et al., 2022). As seen during the Covid-19 pandemic, data from electronic health records combined with clinical trial data was used for real time monitoring of treatment vaccines and strategies (Cake et al., 2022). With benefits also come concerns over its integrity being compromised. A nationwide poll conducted in the US in 2006 reveals over 50% of the participants have concerns over how they have lost control over their medical data and is distributed freely among government agencies, employers and insurance companies. (Harris, 2007). Anderson & Agarwal (2011) argue that given the rapid pace of digitalization in healthcare there is variability in the policies and legal systems to safeguard privacy; additionally, theory development regarding personal health data/information is lagging and there is a need for it to be explored in more specific contexts.

2.3 Data donation and Prosocial behavior

With the advent of the GDPR in the EU, HIPPA in the US and data breaches from Facebook and Tricare among others, public awareness on the issue of data sharing has never been higher (Shaw, 2020). Changes in data laws makes it possible for industry-collected data to be shared by individuals for

research benefiting public good as individuals are free to either store the data for personal use or to transmit it to another data controller (i.e., researcher) More specifically, the GDPR introduces Right to Data Portability, which allows individuals request to obtain data that a data controller holds and to reuse it for their own purposes (Skatova & Goulding, 2019). While we associate the concept of natural selection with selfish tendencies, evolution suggests this is far from the truth; sharing is in part of our nature and helped us survive by cooperation (Ash, 2012). In a modern digitized world, society has produced new pathways of sharing relationships. These range from social networking platforms where people engage among themselves to interact (Alashoor, Han, et al., 2017), digital fundraising (Vaidya, 2014), to crowdfunding (Cox et al., 2018) and food sharing (Harvey et al., 2019). In the same way we donate blood or organs, data donation takes on the process where individuals are encouraged to donate their digital information for medical research and academic research. Although data donation organizations existed previously such as the UK's NHS, non-profits like Data for Good and Open Human, the drive for such donations escalated during the Covid-19 pandemic (Fridman et al., 2022). The issue of systematically allowing individuals to donate their health data for research purposes has not yet been addressed in academic or popular literature, where emphasis has been placed mostly on data sharing between researchers and private corporations (Taddeo, 2016). The study therefore takes a theoretical and empirical approach to maximize data donation from the individual's perspective and contribute to research.

Sojka & Sojka (2008) investigated the motivators among 600 blood donors and found that altruism was the most common motivator for donating blood and continuing to be an active blood donor. Donations as we believe can be conceptualized as gifts and are tied to the donor's generosity as well as some form of obligation on the side of the recipient (Hummel et al., 2019). Skatova & Goulding (2019) in a similar theme found that prosocial factors were the biggest motivators for individuals donating personal data. While both altruism and prosocial behavior is guided by the will to help others it is important to make the distinction between them. *Altruism* is motivation to increase another person's welfare whereas *Prosocial behavior* covers the broad range of actions like sharing, helping and is often used as an umbrella term that describes activities undertaken to benefit other individuals or society (Schwartz & Bilsky, 1990). Batson & Powell (2003) state that prosocial behavior need not be motivated by altruism

and altruistic motivation need not produce prosocial behavior. Therefore, donating personal data could likely become a new act of digital economy prosocial behavior (Skatova & Goulding, 2019). An important driver for prosocial behaviour is social responsibility, or the feeling of duty. Steele et al. (2008) showed that social responsibility was higher in people who continued to donate blood, as compared to those who lapsed. Moreover, when feelings of duty were experimentally induced, it increased the frequency of actions to help others (Clark et al., 1986). Understanding differences in prosocial motivations to donate personal data therefore has implications for the efficacy of campaigns encouraging the sharing of personal data to benefit society (Skatova & Goulding, 2019). The subsequent section will shed light on the drivers and inhibitors affecting data donation.

2.4 Privacy behavior and Decision making

Smith et al. (1996) summarized all empirical assessments of privacy literature into what he described as the APCO (Antecedents > Privacy Concerns > Outcomes) macro model. Antecedents refer to individual traits or situational/contextual factors that influence one's privacy concerns. The model widely used is centered around situational factors such as privacy experiences, demographics, personality, privacy awareness and cultural differences influencing privacy. Privacy calculus, regulations and trust were other variables included in the model and it was concluded privacy is affected at multiple levels by several factors. In the quest to fully understand the complexity of the interactions researchers added other factors influencing privacy decision making; Trust in other parties, knowledge of risk and protection, faith in the ability to protect information and monetary benefits was added by Acquisti & Grossklags (2005). Dinev et al. (2015) critiqued the APCO model and expanded on it by implementing psychology (such as the elaboration likelihood model) and behavioral economics (biases and bounded rationality). Existing research also points to other antecedents that commonly influence privacy: *Privacy experiences*- previous negative experiences will increase privacy concerns (Ozdemir et al., 2018; Smith et al., 1996b). *Demographics*- Demographics is another parameter that can affect privacy concerns (Culnan & Armstrong, 1999). Older users and Females have higher privacy concerns than their counterparts (Hoy & Milne, 2010; Youn & Hall, 2008). *Personality differences*- Being trusting, sympathetic, straightforward, and selfless is shown to increase privacy concerns (Terracciano et al., 2003). *Privacy awareness* **-** knowledge about organizational privacy practices is referred to as privacy

awareness (Malhotra et al., 2004). Awareness particularly increases concerns when users learn that the company used their personal data without their consent (Cespedes & Smith, 1993). *Culture*- high masculinity cultures who prioritize material success over caring relationships show higher concerns for unauthorized usages than low masculinity cultures e.g., Sweden (Bellman et al., 2004). There is also considerable evidence attention depends of familiarity (Mather, 2013) thus being more familiar with an entity could reduce attention to privacy details Thus there certainly are multiple factors which need to be considered to understand data donation decisions. F. Xu et al. (2013) in her study revealed when making a final decision, the perceived benefits of disclosing information outweigh the risk of privacy concerns. The most used theory explaining overall disclosure behavior is the Privacy calculus which states that individuals evaluate anticipated benefits vs. perceived cost in order to disclose their personal data (Princi & Krämer, 2020). If benefits (cumulation of control and trust) outweigh costs (cumulation of risk and privacy concerns) then individuals have higher motivations for disclosure (Dinev et al., 2016). On the contrary individuals also deviate from the norm and disclose their personal information despite there being risks and concerns for potential misuse. This is a privacy paradox. The following will be discussed broadly in section 4. Although the perspectives and decisions on self-disclosure are not mutually exclusive they most likely interact with each other. In another study on context-specific privacy concerns the perceived control over one's personal information was a key factor for information disclosure (Xu et al., 2012).

(Anderson & Agarwal, 2011) further expanded on the contextual factors influencing disclosure of information within healthcare domain and revealed that privacy behavior within healthcare, and its risks and benefits differ in three contexts: 1) healthcare is built up multiple types of health information with varying degrees of protection e.g., Mental health, drug history (Beckerman et al., 2008). 2) the purpose for which the information is used, and 3) The healthcare value chain has multiple players with a need to access and use health information, and the individual has varied levels and modes of interaction with these players (Rohm & Milne, 2004). Anderson & Agarwal (2011) in their study about personal health disclosure found that emotions and the data requesting organization are important influencers with hospitals having more levels of trust whereas pharmaceutical companies and government organizations have almost similar levels of trust. Skatova & Goulding (2019) in a similar

study on health data donation found conflicting results showing individuals prefer to donate more to a generic health organization than to a hospital. In privacy literature trust in the data requesting entity is a key factor affecting disclosure of sensitive information (Acquisti & Grossklags, 2005; Dinev & Hart, 2006); Trust is further influenced by familiarity with the organization, decision context and moral relevance (Anderson & Agarwal, 2011; McKnight et al., 2002). There however exists to be an ambiguity and lack of clear distinction regarding an individuals' trust in government vs. private organizations for health data donation. The sudden increase of government support in times of crisis is well documented in the literature and labeled as 'rally-round-the flag' effect (Radu, 2022). Presently trust and distrust are evenly split despite the pandemic; A survey found that on average across European countries 41.4% of respondents say they trust their national government while 41.1% say they do not (oecd.org, 2022). Crisis also drives us towards increased selfishness (Rodrigues et al., 2009) or increased altruism and generosity (Glynn et al., 2003) with proof of both during the pandemic (Fridman et al., 2022). To contribute to the existing gap in privacy literature the study will look at the government and private organizations to understand privacy attitudes towards them.

2.5 Privacy protection

Worldwide privacy concerns among populations are on the rise and the reason why the topic of privacy protection has been widely discussed by researchers. Despite widespread privacy regulations many studies point that individuals could still experience privacy infringement and harm (Lin et al., 2021). Early economists believed that protecting personal information would harm the individual and society by creating inefficiency in the market (Posner, 1978, 1981; Stigler, 1980). Since people only disclose favorable information, the market would bear increased costs protecting and concealing negative traits (Stigler, 1980). Hermalin & Katz (2006) disagreed with the early research stating data protection would lead to a positive effect on welfare as it can support insurance products. Varian (2002) states although consumers may not have the full knowledge or control how their data is being used disclosing too little information to third parties could lead to limited offers. Alternately sharing too much information could cause harm for the individuals as it can impact deals between the individual and the external entity. When sharing information with an external entity there is always a chance of individuals bearing costs (Acquisti et al., 2016). Laudon (1996) believed that inefficiency in the market is also affected by privacy

invasion. Privacy invasion refers to the use and obtaining personal information without the consent of the individual. Privacy legislation should be adapted according to technology and the best way to structure an information market would be regulations that require an individual's consent before utilizing their data.

Healthcare is unique in a way that it has multiple stakeholders at each level (individuals/patients, service providers, governmental bodies and private organizations). Therefore, privacy solutions should be in a way that respects the individual's preference, the service provider's privacy policy, while complying with the set privacy laws and regulations (el Majdoubi et al., 2022). To gather insight from academia, government, and industry on medical data donation two workshops were conducted in 2018 and 2019 by Krutzinna & Floridi (2019). Key challenges were identified, and it was revealed that trust, data quality, social values affect the willingness to share data while impediments to corporate data sharing include concerns around justice and inclusion. Suggestions were made to make sharing more tangible, by giving concrete examples of benefits for the stakeholders involved and practical information about the use and re-use of donated data. Such would help remove barriers to data donation by fostering a greater understanding of the process, including the risks involved. Additionally, inclusion was mentioned as a key theme for further investigation (Krutzinna & Floridi, 2019). Prainsack (2019) suggestions for designing a strong regulatory framework needed an understanding of the characteristics of data donation. These were: 1) *Relationality*- a characteristic of donation that tells us to be attentive to the relationships of both the giver and receiver of donations to their human, natural and artefactual environments, and to the needs and capabilities that emerge out of these relations. 2) *Indirect reciprocity*- to ensure that the relationship between givers and receivers is not starkly unbalanced in terms of the overall distribution of costs and benefits, duties and entitlements and, 3) multiplicity of data–i.e., the fact that data can be, and often are, in different places at the same time. means that we can, and arguably need to, ask the question under what circumstances data donation should entail a transfer of rights to exclusive use, if at all (Prainsack, 2019).

Data donation and privacy in the same sentence exists as a conundrum. Once data is donated whose data is it anyway? "Information 'about me' does not cease to be connected to my privacy when I give

(or sell) it to others" (Montgomery, 2017, p. 82). Montgomery (2017) further argues that when individuals donate data there may not be a transfer involved instead donations can be understood as a suspension of privacy claims. Although this is a debatable take, research agrees costs of donations incur to the donor alone (Hummel et al., 2019) thereby making the data his claim. Despite many efforts trying to balance between privacy management and health data secondary use from both the legislation side and the technology side a perfect balance is difficult to achieve; instead, a certain tradeoff or compromise must always be made (Xiang & Cai, 2021). The conundrums of data ownership and availability was illustrated in the Covid-19 pandemic as many countries mandated use public of health data with few privacy controls (Mcgraw & Mandl, n.d.). While numerous data and privacy protection measures are being researched to protect individuals; El Majdoubi et al (2022) conducted a systematic review of Privacy preserving solutions in the smart healthcare environment and found that cryptography is the most dominant architecture proposed as a solution (Blockchain-based category accounted for 37% of the solutions). The most addressed architectures in the papers were centralized followed by decentralized architectures. Although discussion of such mechanism are outside the scope of the study it was important to shed light on it as they help protect the disclosed data and reduce privacy concerns prior to data donation decision.

3. Theoretical framework

3.1 Privacy calculus

An individual always needs to weigh up the benefits expected from his information disclosure against the risks of its breach. This balancing process is provided by the Privacy calculus. First described by Laufer & Wolfe (1977) the model has been highly discussed in privacy research and built upon. The theory assumes that individuals evaluate anticipated benefits and perceived risks to make a rational decision regarding the disclosure of their personal data (Princi & Krämer, 2020).

(Culnan & Armstrong, 1999) expanded on the privacy calculus model and added the decision making process concluding that the more information individuals have about the vendor and perceived fairness the greater the willingness to disclose. (Dinev & Hart, 2006) extended the privacy calculus to E-commerce and assumed if benefits (cumulation of control and trust) were higher than costs

(cumulation of risk and privacy concerns) then the individual would have higher motivations for disclosure. (Dinev et al., 2016) further extended it to healthcare domain and found that individuals perceived benefits, convenience and internet experience can reduce privacy concerns. Culnan & Bies (2003) described the privacy model in another angle as a cost-benefit analysis where each set of beliefs can outweigh the other; An individual decides to disclose his information or not based on the individual's probability calculation of how likely each risk or benefit will occur. "*A net positive outcome should mean people are more likely to accept the loss of privacy that accompanies the disclosure of personal information as long as an acceptable risk accompanies the benefit*" (Culnan & Bies, 2003 p. 327). It is important to note that prior research focused only on the personal cost and benefit level, however, privacy calculus extends to the societal level as well and therefore the need for the study focusing on societal privacy calculus. This also aligns with the context of data donation in healthcare where disclosure is often if not always associated with societal benefit instead of personal.

Privacy calculus has been criticized by scholars stating behavioral biases can restrict decision making (Acquisti & Grossklags, 2005). Privacy decisions are usually based on incomplete information (Marreiros et al., 2017) and the ability of individual's cognitive processing might be limited even if detailed information is available resulting in an imperfect decision (Princi & Krämer, 2020). Gerber et al. (2018) argues that decisions may be irrational to an external observer, but at the same time rational to the one making the decision. Several studies in various contexts additionally provide evidence for the privacy calculus model explaining privacy-related behavioral intention and actual behavior of individuals (Dienlin & Metzger, 2016; Kim et al., 2019; Krasnova et al., 2012). Acquisti et al. (2013) implied individuals have limited rationality by having asymmetric information (knowledge regarding the disclosure) and bounded rationality (inability to process cost/benefits accurately) thereby engaging in disclosing information even if the rewards are minute and privacy concerns exist. Barth & de Jong (2017) supported this stating users claim to have privacy concerns but disclose private information, nonetheless. These deviations in normalcy are described as a privacy paradox. A privacy paradox occurs when individuals disclose their personal information despite there being risks and concerns for potential misuse. Such deviations regarding the privacy calculus need to be further investigated,

moreover measuring the actual behavior instead of relying on a behavioral intention questionnaire seems necessary since they often can be contradictory (Buck et al., 2022)

3.2 Elaboration likelihood model and priming

The Elaboration Likelihood Model (ELM) seeks to explore how individuals process stimuli differently and how the outcomes of these stimuli result in changing attitudes and behavior (Nickerson, 2022). Petty & Cacioppo (1986) first created the model as a general theory of attitude change- a framework for organizing, categorizing, and understanding the basic processes underlying the effectiveness of persuasive communications (Nickerson, 2022). The ELM states that when individuals are presented with a message, they take a central high elaboration route or the peripheral route to assess decisions. The central route is logic driven and uses data and facts whereas the peripheral route uses cues and heuristics (Petty & Cacioppo, 1986). In privacy literature Angst & Agarwal (2009) revealed how positively framed messages influence personal health information disclosure. Till date a limited number of studies examine the effect of ELM of privacy and privacy concerns (Alashoor, et al., 2017; Angst, 2009; Gu et al., 2017; Zhu et al., 2021). Alashoor et al. (2017) examined ELM through the lens of privacy concern priming and studied the effect of it on disclosure. When privacy concerns are provided prior to disclosure the assumption is privacy concepts would be easily accessible to the individual's mind resulting in lower disclosure. Other studies acknowledge that privacy priming causes individuals to take more safety into account (Chong et al., 2018). A priming effect takes place when an individual's exposure to a certain stimulus influences his or her response. Privacy concern scales in this study are regarded as a persuasive stimulus on the central ELM decision making in alignment with Alashoor, et al. (2017) argument that privacy concern scales are inherently worded in a negative manner causing individuals to elaborate more on their privacy attitude.

3.3 Privacy concerns, Perceived benefits, and costs

Privacy calculus is a complex process where multiple factors are taken into consideration prior to information disclosure. Most privacy literature reveals privacy concerns have a negative impact on willingness to reveal personal information. Privacy concerns are an individual's fear about disclosure of personal data and information privacy (Li et al., 2017). Thus, the more privacy concerns are the less an

individual can be expected to disclose personal information. This however is not the case due to interactions of benefits and costs with an individual's privacy concerns. If an individual derives greater value from the perceived benefits, he/she discloses more information despite presence of privacy concerns (Wilson & Valacich, 2012). If perceived benefits are more immediate then benefits again often outweigh concerns (Acquisti & Grossklags, 2005). Three major factors of promoting benefit are: personalization, financial reward, convenience and social adjustments (Smith et al., 2011). In the healthcare scenario data donations often do not come with any personal benefits except data exchange scenarios like through a fitness app where we there is expected returns. When an individual shares data, he/she tries to exert control over one's data: where it goes, who has access to it, and what is being done with it. Hummel et al. (2019) terms this as data sovereignty and states that a data sovereign individual does not just close off his/her data but shares it with others by natural complex relations. Donations as such can be considered as a gift. Hénaff et al. (2010) states that gifts transcend beyond the economic circle and what is deemed profitable; It figures in establishment and fostering social bonds through relations of recognition and esteem while also prompting attitudes of generosity, benevolence, and gratefulness. Skatova & Gouldings, (2019) study on personal data donation found that the strongest predictor of the decision to donate was the desire to serve society, while the strongest predictor of not to donate personal data was the need to gain direct benefits because of data donation. Other studies in different domains also reveal individuals desire to serve society over any personal benefits (Evans & Ferguson, 2014; Luccasen & Grossman, 2017). Hummel et al. (2019) states that the novel concept of health data donations can be set in motion if individuals are ready to engage in this domain which opens opportunities, but also where frustrations and harms can never be ruled out

As benefits are accounted for in the decisions process, an individual also assesses severity of the consequences of disclosure. Perceived costs are the degree to which an individual believes there is a high potential for loss following disclosure of personal information; an individual perceived to be at risk therefore shows less willingness to disclose (Smith et al., 2011). Perceived costs stem from the belief how likely an organization shares personal information with unauthorized entities (Dinev & Hart, 2006). Privacy concerns are also produced and enhanced by perceived costs (Li et al., 2017). Privacy costs

include identity theft, financial fraud, misuse/abuse of information among others and can lead to physical, material and emotional consequences (Smith et al., 2011).

4. Hypotheses and Framework

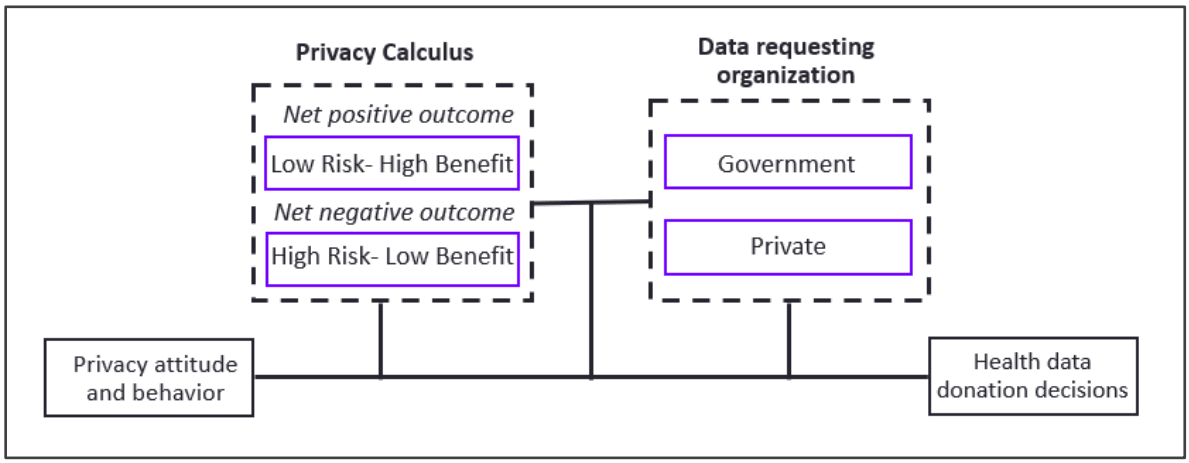

Figure 1: Experimental model

Figure 1 depicts the research model. The model predicts that an individual's subjective privacy attitude and behavior will impact his/her health data donation decisions. The model also predicts that this relationship will be altered by societal privacy calculus (net positive outcome vs. net negative outcome) and the data requesting organization (government vs private). These factors will account for the final data donation decision-making process of the participant. As mentioned in the previous sections, privacy literature shows that there has been limited research conducted on societal privacy calculus and effects of the data requesting organization on data donation decisions therefore the study aims to explore these topics.

Consistent to privacy literature, benefits and risks tend to always coexist in the privacy calculus. Culnan & Bies (2003) point out that in a net positive outcome setting people would be more willing to disclose their personal information. As such for this study context risks imply a net negative privacy calculus outcome (High Risk & Low Benefit) and benefits imply a net positive outcome (Low Risk & High Benefit). Furthermore, Wilson & Valacich (2012) acknowledge the assumption that individuals disclose more information when they derive greater value from the benefits. In the study participants are informed

that the purpose of their data donation will be for health research The purpose of data donation is seemingly important as Varian (2002) revealed can increase the willingness to disclose information. Moreover, Skatova & Goulding (2019) found that the biggest motivators to donate data for individuals was the desire to serve society as in the case of health research. Literature also reveals that just as benefits are associate with disclosure, risks and privacy concerns make individuals less willing to disclose personal information. Li et al. (2017) points that when individuals believe there are risk/ cost associated with the disclosure privacy concerns tend to be enhanced, therefore the study expects that by presenting participants with net negative outcomes associated with greater risks they would have less incentive to donate health data. This leads to the first hypothesis.

H1: Perceived societal Privacy calculus will influence data donation decisions, such that individuals will donate more(less) data when they perceive net positive(negative) outcomes.

An ambiguity exists in privacy literature due to limited research focusing on distinctions between the data requesting organization and their effect on health data donation. Some studies point to government and public institutions having more level of trust (Anderson & Agarwal, 2011; Radu, 2022), other studies point to generic organizations having more trust than the government (Skatova & Goulding, 2019) while some reveal similar levels of trust between them (oecd.org, 2022). Privacy literature shows that trust is a highly important factor affecting sensitive information disclosure and privacy concerns (Acquisti & Grossklags, 2005; Dinev & Hart, 2006). Familiarity further influences the level of trust and concerns (Anderson & Agarwal, 2011; McKnight et al., 2002), therefore the assumption is that as individuals are more familiar with the government than private organizations, they are assumed to trust the government more which leads to the hypothesis.

H2: The data requesting organization will influence data donation decision, such that if the data requesting organization is a government one individuals will donate more data.

In alignment with the previous assumption how organizations have different familiarity and trust levels, they also tend to have different privacy practices which individuals are aware of (Malhotra et al., 2004).

These privacy awareness from individuals help decide which organizations are trustworthy. Among other contextual factors relating to an organization is also the demographics and previous privacy experiences. To concur, no two organization have the same level of risks and benefit therefore they are likely to have an impact on the net privacy calculus. As in our study we experiment on the combined conditions of organization and privacy calculus we are likely to experience an interaction effect.

H3: The type of data requesting organization will have an interaction effect on the privacy calculus, such that it will influence data donation decisions.

5. Method

The study method used is a randomized experiment where privacy behavior and decision making of the participants were manipulated through the two data requesting organizations and net negative privacy calculus and net positive privacy calculus to see data donation scores. The experiments were inspired by approaches done by Anderson & Agarwal (2011), Dinev & Hart (2006), Alashoor et al. (2019) and Culnan & Bies (2003) methods of risk-benefits in the privacy calculus. The experiment participants were led to believe that they would be asked to donate their health data. Additionally, they were given the liberty of choosing the type of health data they would like to donate. A vignette technique was used to present a set of hypothetical yet realistic scenarios for health data donation. A vignette technique provides actionable insights into the judgements, activities, and behaviors especially when situations under scrutiny are rare, occur in complex settings, or raise difficult ethical questions (Sheringham et al., 2021). The hypothetical scenario also served to reveal actual donation decision than simply stating intentions. The experiment design is 2 (data requesting organization: government vs. private) x 2 (societal privacy calculus: net positive [LRHB] vs. net negative [HRLB]) factorial design. The experiment provides a test for hypothesis H1, H2 and H3.

The participants were asked two sets of questions inspired from Datadonor projects and Alashoor et al. (2019): The first set comprising basic demographic data questions (age, gender, nationality, education, employment) and next set comprising of 21 health related questions that they could select to donate (height, weight, blood type, sleep activity, diet, allergies, diet, allergies, Covid-19 vaccination status,

medications, surgical history, alcohol activity, smoking activity, marijuana activity, drug/narcotic activity, exercise habit, social media usage, acute disease, chronic disease, family health history, mental conditions, health checkup activity, insurance). While the general notion is that some health information is more sensitive than other, Anderson & Agarwal (2011) in their study found that there were no significant differences between the type of health information, therefore it was important to formulate the donation questions with different sensitivity.

Additionally, privacy concern was measured by four items based on the works of Dinev & Hart (2006) built on scales by Smith et al. (1996a) and Culnan & Armstrong (1999). A 7-point Likert scale was used to provide detailed answers (agree-disagree) and measure correlations. Participants were randomized in the order they received the privacy concern scale and data donation questions; those who received data donation question first received the privacy concern question later and vice versa. This was important to the experiment as privacy concerns scale is a priming method used in the study and would be expected to affect data disclosure. The assumption is privacy concerns scale when provided to an individual produces a priming effect due to the negative phrasing of the statements ("I am concerned..."). Primed individuals are also shown to have different disclosure outcomes (Alashoor, et al., 2017). Further this would also help reveal whether the privacy paradox holds true and whether decisions are consistent with attitude. Thereafter, participants were given manipulation checks and then attention checks were provided to screen for improved data quality (Appendix 1).

The platform Qualtrics was used for constructing the survey and collecting data which was published to Prolific. Prolific is a popular survey platform with a reliable and valid user base. To further improve the quality of the selection criteria of the participants were: 1) narrowed down to USA 2) survey published in English, and 3) only users who completed 400 and more surveys could participate in the survey.

5.1 Procedure

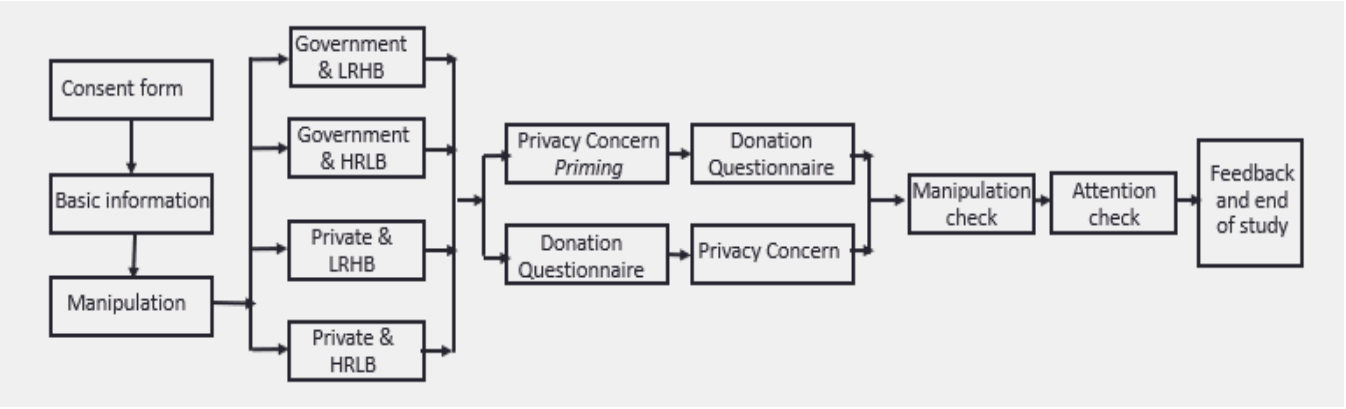

Figure 2: Flowchart of the experiment

Selected participants were present with a consent form prior to starting the experiment. This was required as the study asked for disclosure of personal information. The consent included the purpose of the study where the process would include answering demographic questions, knowledge about the simulation scenario and its project, data donation options and attitude towards health data donation. The consent section ended with an option of agreeing to the terms and conditions of the consent form and the participants' Prolific Id.

Participants that agreed to take part were randomly distributed with one of the four manipulation conditions. The combinations were: government and low-risk high-benefit (LRHB), government and high-risk low-benefit (HRLB), private and low-risk high-benefit (LRHB) and private and high-risk low-benefit (HRLB). Each were then provided an introductory passage to the simulation scenario (The ART foundation and their objective of improving arthritis treatment) combined with the random manipulation conditions. The introduction of the data requesting organization (ART foundation) helps in engagement and prevents ambiguity of information. (Appendix 1)

The introduction was followed by demographic questions. Hereafter, participants were allocated both the privacy concern scale and the health data donation questionnaire but with one preceding the other in an evenly randomized manner. Each one of the 21 health data donation questions had the option of not donating certain data (*'I prefer not to donate this information')* for checking the sensitivity of certain

questions. Scores were recorded for each donation by 1 point to understand their disclosure behavior. The privacy concerns questions had a 7-point Likert scale implying the concerns (ranging from strongly disagree to strongly agree). Manipulation checks of the given privacy conditions were hereafter given.

Attention checks were then provided asking participants to identify the two conditions they were distributed to. This facilitates better data quality for analysis. Finally, the experiment ends with the description of the actual study and its purpose and a feedback question. The feedback question would provide qualitative information and aid future studies.

5.2 Manipulation

The manipulation was constructed to identify implications of combined data requesting organization and cost and benefit on privacy behavior and decision-making of the participants. Manipulations help establish the framework of an experimental study. The manipulation is the first step presented to the participants. Based on the 2x2 factorial design four manipulation conditions were constructed. These were evenly and randomly distributed among the participants. The combined manipulation conditions are illustrated in Figure 3 below.

		Privacy calculus	
		Net positive outcome	Net negative outcome
Data requesting organization	Government	**Condition 1:** Gov. LRHB	**Condition 2:** Gov. HRLB
	Private	**Condition 3:** Priv. LRHB	**Condition 4:** Priv. HRLB

Figure 3: Manipulation conditions

The scenarios allocated for each condition used a similar sentence structure to prevent any deviations other than the manipulations. For the net positive outcomes of the privacy calculus participants were informed that the research would have low likelihood of privacy breach due while having significant benefits for people suffering from arthritis. For net negative outcomes participants were warned about

the high likelihood of breaches due to lack of proper data protection while having only slight improvements in arthritis treatment. The manipulated conditions were critical to examine the participants privacy behavior and decision making. They acted as the independent variable of this study and help test our hypothesis. Furthermore, the manipulation conditions also served as attention checks. The detailed manipulation scenarios are found in Figure 4 below.

Condition 1	**Condition 2**
The ART Foundation, a government-funded project, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals. Experts believe that this data donation governmental research is expected to have: 1. **Low Privacy Risks to Society**. The reason is that this research has a low likelihood of information breaches because it is subjected to data protection laws. 2. **High Benefits to Society.** The reason is that this research could significantly reduce the prevalence of arthritis in the population.	The ART Foundation**,** a government-funded project, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals. Experts believe that this data donation governmental research is expected to have: 1. **High Privacy Risks to Society**. The reason is that this research has a high likelihood of information breaches because it is not subjected to data protection laws. 2. **Low Benefits to Society.** The reason is that this research could overestimate projections and only have slight improvement in arthritis treatment.
Condition 3	**Condition 4**
The ART Foundation, a private-funded project, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals. Experts believe that this data donation private research is expected to have: 1. **Low Privacy Risks to Society**. The reason is that this research has a low likelihood of information breaches because it is subjected to data protection laws. 2. **High Benefits to Society.** The reason is that this research could significantly	The ART Foundation, a private-funded project, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals. Experts believe that this data donation private research is expected to have: 1. **High Privacy Risks to Society**. The reason is that this research has a high likelihood of information breaches because it is not subjected to data protection laws.

reduce the prevalence of arthritis in the population.	2. **Low Benefits to Society.** The reason is that this research could overestimate projections and only have slight improvement in arthritis treatment.

Figure 4: Manipulation scenarios for the conditions

5.3 Manipulation checks

Participants in the experiment were presented with manipulation checks before the attention check. They were essential to the experiment and had the purpose of 1) ensuring the functionality of the manipulation and 2) exclusion criteria. Manipulation checks were provided in two items using a 7-point Likert scale. The scale ranged from 1 to 7 (1- Strongly disagree, 7- Strongly agree). The first item checked for the risk condition ("I *believe the ART foundation will be protected against data breaches")* while the second item checked for benefit condition (*"I believe that the ART foundation will provide a lot of benefits to society")* (Appendix 1). The core idea of the manipulation was that participants would have scores that were congruent to their given conditions. If for example participants were given LRHB (Low Risk -High Benefit) condition, they would be expected to score high on the manipulation check scales. The manipulation checks reveal which participants understood the given manipulation conditions and who did not. In the study this was also used as an exclusion criterion.

6 Analysis and results

Attention check

A total of 402 participants completed the survey on Prolific. Prior to stepping into analysis and interpretation of the data, participants who failed the attention check questions were filtered out (Appendix 1). Participants who explicitly got either of the conditions wrong or were unsure of both the conditions they were given ("I don't remember") were excluded. Attention checks are crucial to the experiment as they prevent careless responders from directly affecting the quality of data (Kung et al., 2018). After applying the attention check, a total of 276 participants were included. When tabulated by the groups it was important to take note of the sample size of each group as represented in table 1.

Table 1: Inclusions and exclusions after Attention check

	attention	check	
condition	not failed	failed	Total
Gov.LRHB	59	42	101
Gov.HRLB	50	52	102
Priv.LRHB	94	7	101
Priv.HRLB	73	25	98
Total	276	126	402

As seen in the table we can see a huge difference in attention check failure for participants who were given the government as data requesting organization (*Gov*) vs. private data requesting organization (*Priv*). There is also a difference in exclusion numbers between participants who were given a net positive privacy calculus (*HRLB*) vs. ones with a net positive calculus (*LRHB*). It was therefore important to check whether the distribution took place by chance or whether there was a variable that accounted for such distribution. To understand this, a Chi2 test of the failed attention check group was done against variables that could possibly cause these distributions. Testing was done for age, gender, education, employment, conditions for data requesting organization and net privacy calculus and the combined 2X2 condition. Age ($p=0.836$), gender ($p=0.626$), education ($p=0.310$) and employment (0.502) showed no statistical significance. However, from Table 2 below we see that net privacy calculus scenario had a statistically significant association ($0.002<0.05$) with participants who failed the attention check. The same was also seen for the data requesting organization (*Priv*) which also showed a statistically significant association ($0.000<0.05$) between the attention check. Further as it was a 2x2 table we measured the *phi coefficient* which measures the strength of the association. Looking at *phi values* the table shows that the strength of association on attention check failures is higher with the data requesting organization (phi=0.325) than with net privacy calculus (phi =0.153). Thus, these factors help explain the difference in failure rates between different conditions and their combinations. Furthermore, analyzing the results without exclusion of participants (Appendix 2) also show the same results regarding donation and privacy concerns as we will see in subsequent sections.

Table 2: Chi2 tests for Private and HRLB conditions

attention check	HRLB LRHB	HRLB	Total
not failed	153	123	276
failed	49	77	126
Total	202	200	402

Pearson chi2(1) = 9.4734 Pr = 0.002
phi = Cohen's w = fourfold point correlation = 0.1535

attention check	private Governmen	Private	Total
not failed	109	167	276
failed	94	32	126
Total	203	199	402

Pearson chi2(1) = 42.6608 Pr = 0.000
phi = Cohen's w = fourfold point correlation = 0.3258

Manipulation check

After completing the attention check participants were checked if they passed the manipulation check for privacy concerns. A variable *PC_MC* was generated by the average score of the two statements for the manipulation check as it was necessary to combine the scores of two statements to obtain one variable. A histogram was created to indicate cut-off value for the manipulation check criterion. It was first used due to the limitations of the box plot chart that identifies outliers based on the median of the numeric variables (Figure 5). Participants having values significantly above/below the cut-off value would be excluded. A box plot was then created from the cut off in the histogram to identify any remaining outliers. *Figure 5* shows the box plot chart after the exclusions. It is important to note that some of the outliers in the box plot were not excluded since it aligned with the given manipulation.

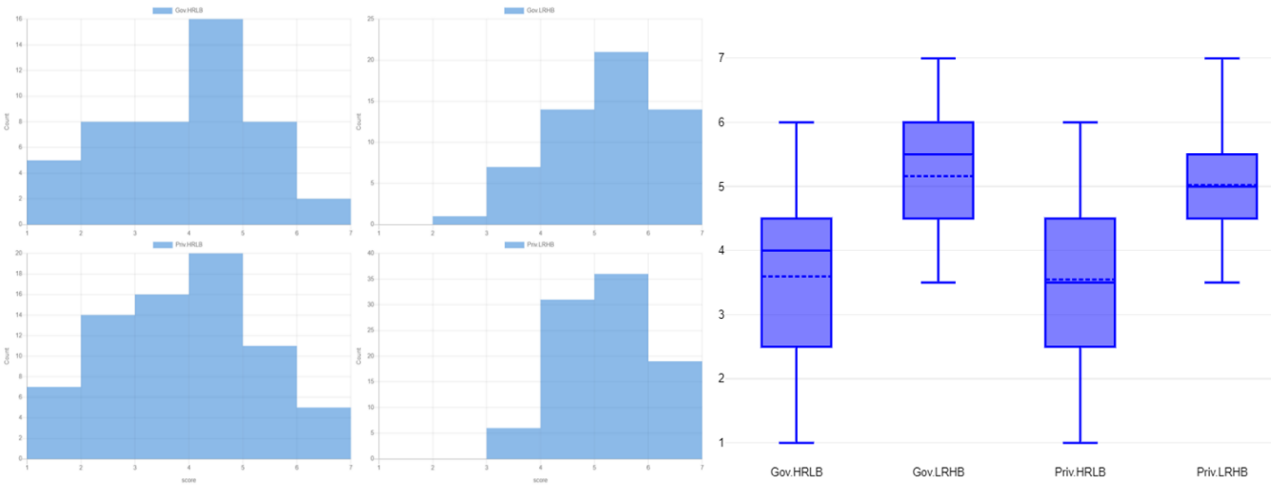

Figure 5: The histograms and box plot of privacy manipulation check on the conditions

Data organization

In total 268 participants passed the attention and manipulation checks. Data collected from Prolific platform was cleaned and exclusion methods applied. The data of these participants was imported to STATA for analysis. Except for the qualitative answers to the data donation questions all other variables were transformed into numeric or binary values. Further, independent variables were created for privacy concern, priming and conditions participants they were provided. As the privacy concern had four items of measure a variable *PC_avg* (average of the scales) was created to measure the overall concerns. The priming variable (*Prim*) was dummy coded as 1=primed, 0=not primed. The data requesting organization variable (*Priv*) was coded as 1=private, 0=government and finally net outcomes of the societal privacy calculus (*HRLB*) was coded as 1= net negative/High Risk-Low Benefit and 0= net positive/Low Risk-High Benefit. After the collection and organization of the data, the calculus statistical analysis was constructed. Three statistical analyses, the one-way ANOVA, a t-test, and a regression analysis was done to reveal any findings on the combined effects of data requesting organization and net privacy calculus outcomes on data donation behavior and decisions.

Anova

An ANOVA test is used for comparing and identifying differences in the mean among three or more independent groups. Prior to carrying out a one-way ANOVA, six assumptions must be satisfied, else a multivariate ANOVA or a two-way ANOVA is required. Laerd statistics, (n.d.) states that the following six assumptions are needed for a one-way ANOVA:

1) "Dependent variable should be measured at the interval or ratio level (i.e., they are continuous)." 2) "Independent variable should consist of two or more categorical, independent groups." 3) "Independence of observations, which means that there is no relationship between the observations in each group or between the group themselves." 4) "There should be no significant outliers." 5) "Dependent variable should be approximately normal distributed for each category of the independent variable." and 6) There needs to be homogeneity of variances."

In the study the dependent variable *Score* is measured on an interval of [1−21]. The independent variables *Priv, HRLB, Prim, Cond* all contain two or more categorical groups. Furthermore, the participants were randomly assigned to one of the four combined manipulation conditions stated in the manipulation section, thereby the independent variables had no relationship between or within the categorical groups. Outliers are also eliminated as mentioned in the exclusion section. Thus, the first four assumptions are fulfilled. As the design of the study is an experimental one, assumption 5 is not applicable. The dependent variable *Score* (donation score) would be an upward or downward sloping distribution instead of a bell-shaped distribution as it depends on the amount of answered questions therefore the assumption of homogeneity of variance was not required. The p-value (significance less than 0.05) and the f-value (amount of difference in mean between the groups) will be discussed in relation to the hypothesis and post hoc tests. A post hoc test helps in determining which specific groups differ from each other. The Bonferroni method is used when the group comparison is decided before the experiment making it the most applicable to the experimental design in the study.

T-test

A t-test is also used to complement the ANOVA test and check between the group means for any statistical significance. A t-test is, however, a hypothesis test and only compares the means of two conditions only. As the same variable will be compared for two independent groups an independent t-test is used instead of a paired t-test. A paired t-test compares different variables of the same group. The t-value and f-value will help to support the hypothesis.

Regression

A regression analysis was necessary to examine the estimated beta coefficients and examine the effects of the main independent variables on the dependent variable (*Score*). It was also necessary for examining the interaction variable. A simple linear regression was conducted, and three different regression models were created to identify the predictable powers of the independent variables on the dependent variable. The dependent variable in the analysis was *Score* while the independent variables for the models were *Priv, HRLB, PC_avg and Cond.*

Test of manipulation check.

To check whether the manipulations were successful within the experimental study a one-way ANOVA test was used on it. This indicates robustness of the experimental data. The test showed that there is a significance difference in mean for the privacy concern manipulation check (f=117.97, significance=0.000<0.05) between participants receiving net negative privacy calculus outcome vs. those with a net positive outcome. This supports that evidence that the privacy concern manipulation check worked as intended. Participants who received higher privacy concerns received lower scores compared to those who did not, as seen in *Table 3.*

Table 3: Anova results of manipulation check

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	150.630752	1	150.630752	117.97	0.0000
Within groups	339.649099	266	1.27687631		
Total	490.279851	267	1.83625412		

Bartlett's test for equal variances: chi2(1) = 27.1566 Prob>chi2 = 0.000

6.1 Descriptive statistics

The statistics *Table.4* shows the descriptive statistics table of the separate manipulation conditions- data requesting organization (Gov. & Priv.) and net privacy calculus outcome (LRHB & HRLB). Each conditions mean, standard deviation, standard error, minimum, maximum and coefficient of variance are provided. When looking at the effect on the dependent variable – donation score (*Score*), we see participants receiving the net positive privacy calculus outcome were more willing to donate health data (n=148; mean=16.614; Sd=5.792) than participants receiving net negative privacy calculus (n=120; mean=12.725; Sd=8.261). The difference in mean between the two categories shows preliminary support for H1. Further, participants who were provided the private data requesting organization category had higher donation scores (n=165; mean=15.175; Sd=7.22) than ones that had the government category (n=103; mean=14.388; Sd=7.318032). This also shows preliminary support for H2. Further analysis needs to be done to gain greater evidence. The table also shows measures of the

privacy concerns (*PC_avg*) for the different conditions. As assumed privacy concerns among participants shows increase when they were provided with net negative privacy calculus outcome (n=120; mean=4.66; Sd=1.626) than with net positive outcome (n=148; mean=3.721; Sd=1.678). Looking at the effect of priming (*Prim*) we see that priming has more effect on data donation disclosure than it has on the privacy concerns of participants. Primed participants for the data requesting organization condition had overall lower scores (n=131; mean=14.694; Sd=7.114) than non-primed participants (n=137; mean=15.043; Sd=7.411). It was also worth noting that the mean donation scores across all conditions were above the mean value (21 questions = 10.5). The distribution of the scores for each donation question was also above average as see in Appendix 4.

In *Table 5* when looking at the descriptive statistics of the four combined manipulations we can see a difference in the mean of participants who had net negative outcomes in combination across all the conditions. Participants who received net positive outcomes had more donation scores (*Gov.LRHB and Priv.LRHB*) than those who received net negative outcomes (*Gov.HRLB and Priv.HRLB*). This further directs evidence in support of H1. Participants receiving the private category with net positive outcome and net negative outcome exhibited the highest as well as the lowest donation scores (n=92; mean=17.282; Sd=5.187) and (n=73; mean=12.520; Sd=8.485); this familiar pattern was also seen in their privacy concern score. On the other hand, participants with the government category had average levels of data disclosure and privacy concerns. This lights the way to provide initial support of H3. The levels of donation scores were higher when privacy concerns were lower which also indicates individuals acted according to their privacy attitudes. The table also shows that primed participants had slightly lower donation scores (n=131; mean=14.694; Sd=7.114) than non-primed participants (n=137; mean=15.043; Sd=7.411).

Table 4: Descriptive statistics of the manipulation conditions (Separated)

	condition	N	mean	sd	se(mean)	min	max	cv
Donation score	Government	103	14.38835	7.318032	0.721067	0	21	0.509
	Private	165	15.17576	7.22289	0.562301	0	21	0.476
	Total	268	14.87313	7.256067	0.443235	0	21	0.488
	Low risk-High benfit	148	16.61486	5.792939	0.476177	0	21	0.349
	High risk- Low benefit	120	12.725	8.261445	0.754163	0	21	0.649
	Total	268	14.87313	7.256067	0.443235	0	21	0.488
Privacy concern	Government	103	4.099515	1.817447	0.179078	1	7	0.443
	Private	165	4.168182	1.657201	0.129013	1	7	0.398
	Total	268	4.141791	1.717515	0.104914	1	7	0.415
	Low risk-High benfit	148	3.721284	1.678831	0.137999	1	7	0.451
	High risk- Low benefit	120	4.660417	1.62659	0.148487	1	7	0.349
	Total	268	4.141791	1.717515	0.104914	1	7	0.415
Donation score	Government	51	14.92157	6.647836	0.930883	0	21	0.446
Primed	Private	80	14.55	7.433928	0.831139	0	21	0.511
	Total	131	14.69466	7.114333	0.621582	0	21	0.484
	Low risk-High benfit	73	16	5.894913	0.689947	0	21	0.368
	High risk- Low benefit	58	13.05172	8.161575	1.071667	0	21	0.625
	Total	131	14.69466	7.114333	0.621582	0	21	0.484
not primed	Government	52	13.86538	7.9509	1.102592	0	21	0.573
	Private	85	15.76471	7.011295	0.760482	0	21	0.445
	Total	137	15.0438	7.411109	0.633174	0	21	0.493
	Low risk-High benfit	75	17.21333	5.667101	0.654381	0	21	0.329
	High risk- Low benefit	62	12.41935	8.408716	1.067908	0	21	0.677
	Total	137	15.0438	7.411109	0.633174	0	21	0.493
Privacy concern	Government	51	4.122549	1.795392	0.251405	1	7	0.436
Primed	Private	80	4.46875	1.580326	0.176686	1	7	0.354
	Total	131	4.333969	1.669181	0.145837	1	7	0.385
	Low risk-High benfit	73	3.839041	1.663732	0.194725	1	7	0.433
	High risk- Low benefit	58	4.956897	1.466092	0.192507	1	7	0.296
	Total	131	4.333969	1.669181	0.145837	1	7	0.385
not primed	Government	52	4.076923	1.856057	0.257389	1	7	0.455
	Private	85	3.885294	1.687027	0.182984	1	7	0.434
	Total	137	3.958029	1.748836	0.149413	1	7	0.442
	Low risk-High benfit	75	3.606667	1.696665	0.195914	1	7	0.47
	High risk- Low benefit	62	4.383065	1.729516	0.219649	1.5	7	0.395
	Total	137	3.958029	1.748836	0.149413	1	7	0.442

Table 5: Descriptive statistics of the manipulation conditions (combined)

	condition	N	mean	sd	se(mean)	min	max	cv
Donation score	Gov.LRHB	56	15.51786	6.572646	0.878307	0	21	0.424
	Gov.HRLB	47	13.04255	7.98084	1.164125	0	21	0.612
	Priv.LRHB	92	17.28261	5.187318	0.540815	0	21	0.3
	Priv.HRLB	73	12.52055	8.485461	0.993148	0	21	0.678
	Total	268	14.87313	7.256067	0.443235	0	21	0.488
privacy concern	Gov.LRHB	56	3.745536	1.739574	0.23246	1	7	0.464
	Gov.HRLB	47	4.521277	1.836252	0.267845	1	7	0.406
	Priv.LRHB	92	3.706522	1.650262	0.172052	1	7	0.445
	Priv.HRLB	73	4.75	1.482537	0.173518	2	7	0.312
	Total	268	4.141791	1.717515	0.104914	1	7	0.415
Donation score primed	Gov.LRHB	31	15.25806	5.966393	1.071596	1	21	0.391
	Gov.HRLB	20	14.4	7.721467	1.726573	0	21	0.536
	Priv.LRHB	42	16.54762	5.852672	0.903087	0	21	0.354
	Priv.HRLB	38	12.34211	8.396513	1.362094	0	21	0.68
	Total	131	14.69466	7.114333	0.621582	0	21	0.484
not primed	Gov.LRHB	25	15.84	7.369306	1.473861	0	21	0.465
	Gov.HRLB	27	12.03704	8.164093	1.571181	0	21	0.678
	Priv.LRHB	50	17.9	4.523183	0.639675	0	21	0.253
	Priv.HRLB	35	12.71429	8.699464	1.470478	0	21	0.684
	Total	137	15.0438	7.411109	0.633174	0	21	0.493

6.2 Anova

The differences in mean of the descriptive statistics show evidence in support of H1, H2 and H3. To further support the hypothesis, we look at the significant levels between the means of the groups in the one-way ANOVA table. The ANOVA (Table 6) based on the separated conditions show a significant difference for the dependent variable- donation score (f-value=20.43; Sig.=0.000<0.05) between participants who received net negative privacy calculus outcome (*HRLB*) and participants who received net positive privacy calculus (*LRHB*). The significant mean helps support H1 where participants will donate more data if they receive net positive outcomes from the donation. There was no significant

difference detected on donation scores between individuals who received a private data requesting organization (*Private*) vs. those who received a government one (f-value=0.75; Sig.=0.389>0.05). As expected, privacy concerns (*PC_avg*) had a significant effect on data donation scores (f-value=4.44; Sig.=0.000<0.05). Interestingly, the effect of priming did not have any significant effect on data donation scores of participants (f-value=0.75; Sig.=0.389).

		Source	SS	df	MS	F	Sig.
Donation score	Private	Between groups	39.317517	1	39.31752	0.75	0.389
		Within groups	14018.369	266	52.70064		
		Total	14057.687	267	52.65051		
	HRLB	Between groups	1002.7143	1	1002.714	20.43	0.000
		Within groups	13054.972	266	49.07884		
		Total	14057.687	267	52.65051		
	PC_avg	Between groups	4287.7491	24	178.6562	4.44	0.000
		Within groups	9769.9375	243	40.2055		
		Total	14057.687	267	52.65051		
	primed	Between groups	39.317517	1	39.31752	0.75	0.389
		Within groups	14018.369	266	52.70064		
		Total	14057.687	267	52.65051		

Table 6: One way ANOVA of the conditions

Looking at the effect of combined manipulation conditions on the data donation score (*Gov.HRLB, Gov.LRHB, Priv.HRLB and Priv.LRHB*), Table 7 shows a significant difference in mean between them (f-value=7.61; Sig.=0.000<0.05). Looking at effect on privacy concern by the combined condition we also see a significant difference (f-value=7.26; Sig.=0.000<0.05) in privacy attitudes among them. While significance can be seen for the combined manipulation conditions, they do not illustrate which of these conditions are actually significant from one another, therefore, a Bonferroni post hoc test comparing the different combined conditions was done. *Table 8* of the post hoc tests show a significant difference in means of the donation scores comes from the fact that Priv.HRLB, Priv.LRHB and Gov.HRLB are significantly different from all the other manipulations: *Priv.LRHB vs. Gov.HRLB (Sig.=0.005); Priv.HRLB vs. Priv.LRHB (Sig.=0.000).* Furthermore, these factors also account for a significant difference in means of privacy concerns (*Priv.LRHB vs. Gov.HRLB (Sig.=0.040); Priv.HRLB vs. Priv.LRHB (Sig.=0.000); Priv.HRLB vs. Gov.LRHB (Sig.=0.005).* While the net outcome of the privacy calculus was significantly associated with data donation scores, the result of the post hoc analysis shows that the data requesting

organization (Gov. and Priv.) may interact with the privacy calculus outcomes and thereby influence willingness to disclose data. This also points to evidence for H3 which will be discussed further in the t-test and regression sections.

		Source	SS	df	MS	F	Sig.
Donation score	combined	Between groups	1118.918	3	372.9727	7.61	0.000
		Within groups	12938.77	264	49.01049		
		Total	14057.69	267	52.65051		
Privacy concern		Between groups	59.99575	3	19.99858	7.26	0.000
		Within groups	727.6162	264	2.756122		
		Total	787.6119	267	2.949857		

Table 7: One-way Anova of combined conditions

Mean & Significance for Donation scores				**Mean & Significance for Privacy concern**			
	Gov.LRHB	Gov.HRLB	Priv.LRHB		Gov.LRHB	Gov.HRLB	Priv.LRHB
Gov.HRLB	-2.475			Gov.HRLB	0.776		
	0.450				0.113		
Priv.LRHB	1.765	4.240		Priv.LRHB	-0.039	-0.815	
	0.829	0.005			1.000	0.040	
Priv.HRLB	-2.997	-0.522	-4.762	Priv.HRLB	1.004	0.229	1.043
	0.100	1.000	0.000		0.005	1.000	0.000

Table 8: Bonferroni post hoc test between the combined conditions

6.3 T-test

The t-tests are used to gain greater evidence in support of the hypothesis. The table below shows the independent sample t-test on the data requesting organization (*Priv*). The table revealed that data donation score (*Score*) is not statistically significant (t-value=-0.863; Sig.= 0.389>0.005) between participants receiving private organization and those receiving a government organization. As the interactions are not significant H2 is not supported.

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.389	-0.7874081	0.9116219	266	-2.582321 1.007505	-0.8637
	Equal variance not assumed	0.390	-0.7874081	0.9143962	216.358	-2.589673 1.014857	-0.8611

Table 9: T test of data requesting organization

An independent t-test was also created for the net Privacy calculus outcome (*HRLB*) on donation scores in *Table 10*. The test reveals that data donation scores is highly significant between the group receiving a net positive outcome and the group with net negative outcome (t-value=4.52; Sig.=0.000<0.05). Individuals who receive a net positive privacy calculus are more likely to donate their health data than those who receive a net negative privacy calculus. Thus, the levels of significance support H1.

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.000	3.889865	0.8605836	266	2.195443 5.584287	4.52
	Equal variance not assumed	0.000	3.889865	0.8919117	207.649	2.131502 5.648228	4.3613

Table 10: T test of net privacy calculus outcome

To gather evidence for H3 it was necessary to make an independent sample t-test for every manipulation conditions provided to participants. This was necessary to see the interaction between the data requesting organization interaction (*Gov.& Priv*.) with the net privacy calculus outcome (*HRLB & LRHB*) on data donation scores. The t-tests on donation scores indicate that a significant difference between *Priv.HRLB* and the other manipulation conditions *Priv.LRHB* (t-value=4.439; Sig=0.000<0.05), *Gov.LRHB* (t-value=2.186; Sig.=0.031<0.05). Comparing *Priv.LRHB* with *Gov.HRLB* (Table 14) we also see a statistical significance (t-value= -3.7743; Sig.=0.000<0.05) between the groups. The results indicate more evidence in support of H3 as they illustrate how the data requesting organization especially private interact with privacy calculus to influence data donation.

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.000	4.762061	1.072598	163	14.06547 16.28604	4.4397
	Equal variance not assumed	0.000	4.762061	1.130851	113.159	2.521675 7.002447	4.211

Table 11: T test between Priv.HRLB vs Priv.LRHB

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.031	2.997309	1.370579	127	.2851797 5.709439	2.1869
	Equal variance not assumed	0.026	2.997309	1.325808	126.983	.3737717 5.620847	2.2607

Table 12: T test between Priv.HRLB vs Gov.LRHB

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.0722	-1.764752	0.9743135	146	-3.690332 .1608286	-1.8113
	Equal variance not assumed	0.0903	-1.764752	1.031457	96.2498	-3.81211 .282607	-1.7109

Table 13: T test between PrivLRHB vs Gov.LRHB

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.000	-4.240056	1.123408	137	-6.461518 -2.018593	-3.7743
	Equal variance not assumed	0.002	-4.240056	1.283615	66.4342	-6.802564 -1.677547	-3.3032

Table 14: T test between Priv.LRHB vs Gov.HRLB

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.087	2.475304	1.433819	101	-.3690077 5.319616	1.7264
	Equal variance not assumed	0.093	2.475304	1.45829	89.1224	-.4222326 5.372841	1.6974

Table 15: T test between Gov.HRLB vs Gov.LRHB

		sig.	mean difference	Std. error difference	df	95% CI difference - Lower and Upper	t
Donation score	Equal variance assumed	0.737	0.5220052	1.550816	118	-2.549033 3.593044	0.3366
	Equal variance not assumed	0.7337	0.5220052	1.530206	102.603	-2.512937 3.556948	0.3411

Table 16: T test between Gov.HRLB vs Priv.HRLB

6.4 Regression

A regression analysis was done to examine the effects of private organization (*Priv*) and net negative privacy calculus (*HRLB*) in relation to their dependent variable– data donation score (*Score*). A regression was necessary to help reveal the estimated beta-coefficients as they would predict the health data donation disclosure in relation to the manipulation condition provided. Three models were created for regression analysis. Model 1 is the baseline model where only the main independent variables are regressed. Privacy concerns (*PC_avg*) is added to the baseline model in Models 2. Lastly in Model 3 regression is done with the combined manipulation conditions (*Cond*). The models and the results are given below (Table 17-19).

1) Data donation Score $=\beta 0 + \beta 1\ Priv + \beta 2\ HRLB + \varepsilon i$
2) Data donation Score $=\beta 0 + \beta 1\ Priv + \beta 2\ HRLB + \beta 3\ PC\_avg + \varepsilon i$
3) Data donation Score $=\beta 0 + \beta 1\ Priv + \beta 2\ HRLB + \beta 3\ PC\_avg + \beta 4\ Combined\ conditions + \varepsilon i$

Source	SS	df	MS	Number of obs	=	268
				F(2, 265)	=	10.55
Model	1036.82873	2	518.414365	Prob > F	=	0.0000
Residual	13020.8578	265	49.1353126	R-squared	=	0.0738
				Adj R-squared	=	0.0668
Total	14057.6866	267	52.6505115	Root MSE	=	7.0097

Score	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Priv	.7335271	.8803264	0.83	0.405	-.9997972	2.466851
HRLB	-3.880118	.861158	-4.51	0.000	-5.5757	-2.184535
_cons	16.15889	.794642	20.33	0.000	14.59427	17.7235

Table 17; Regression for Model 1

Source	SS	df	MS	Number of obs	=	268
				F(3, 264)	=	24.87
Model	3097.3537	3	1032.45123	Prob > F	=	0.0000
Residual	10960.3329	264	41.5164124	R-squared	=	0.2203
				Adj R-squared	=	0.2115
Total	14057.6866	267	52.6505115	Root MSE	=	6.4433

Score	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Priv	.8709469	.8094368	1.08	0.283	-.7228265	2.46472
HRLB	-2.29912	.8227783	-2.79	0.006	-3.919163	-.6790775
PC_avg	-1.681521	.2386839	-7.04	0.000	-2.151487	-1.211554
_cons	22.33088	1.140644	19.58	0.000	20.08496	24.5768

Table 18: Regression for Model 2

Source	SS	df	MS	Number of obs	=	268
				F(4, 263)	=	18.99
Model	3150.38404	4	787.596011	Prob > F	=	0.0000
Residual	10907.3025	263	41.4726332	R-squared	=	0.2241
				Adj R-squared	=	0.2123
Total	14057.6866	267	52.6505115	Root MSE	=	6.4399

Score	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Priv	-1.366089	2.137325	-0.64	0.523	-5.574536	2.842357
HRLB	-3.631631	1.43696	-2.53	0.012	-6.461041	-.8022203
PC_avg	-1.670912	.2387424	-7.00	0.000	-2.141002	-1.200822
Cond	.6131304	.5422149	1.13	0.259	-.4545042	1.680765
_cons	-19.30342	36.83646	-0.52	0.601	-91.83534	53.2285

Table 19: Regression for Model 3

The results of Model 1 (Table 17) show a positive association of *Priv with Score* and a negative association of *HRLB with Score* following their beta-coefficients (β0= 16.159, β1= 0.733, β2=-3.880). Individuals who receive net negative privacy calculus along with the private organization would on average have donation scores of 16.159 (mean). The results of the model therefore provide support for H1.

In Model 2 where Privacy concerns were added to the regression model (Table 18) shows that privacy concerns had a negative association with donation scores (β3=-1.681). This also aligns with privacy literature as the greater the privacy concerns are the less the data disclosure The second model accounted for 22% of variability of the observed data (R-squared= 0.220).

For Model 3 the results (Table 19) show the effect of the combined manipulation conditions (*Cond*) on the regression model (β4=0.613; Sig=0.259>0.05). Prior results of the ANOVA test (Table 8) and t-test (Table 11,12,14) between each of the four combined manipulation conditions show that they point the path to an interaction effect of the data requesting organization on the net privacy calculus; the effects being more significant if it is a private organization. Therefore a post hoc test was done on the well assumed interaction effect. Firstly, probing the overall marginal effect estimations of organization on the net privacy calculus, the results (Table 20) show that the difference is highly significant for the private organization (p=0.006<0.05) with a decrease in -3.028 points on the response scale. Next, we test out the predictive margins at each level of the privacy calculus. From the results in *Table 21* we see that at the negative privacy calculus level the margin difference is small for both the organizations (government=13.676; private=13.536; *difference:13.676-13.536*= 0.14). Interestingly at the level of the positive privacy calculus we see a large margin difference between the organizations (government= 14.855; private=16.555; *difference:16.555-14.855 =* 1.7). Thus, we can conclude from the post hoc analysis that the data requesting specifically private organization has an interaction effect with privacy calculus which influences data donation scores. The Private data requesting organization accentuates the effects of net privacy calculus to affect data donation decisions. The evidence therefore supports H3 of the study. The relations are displayed in *Figure 6.*

```
Average marginal effects                        Number of obs     =        268
Model VCE    : Robust

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 1.HRLB

------------------------------------------------------------------------------
             |                 Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0.HRLB       |  (base outcome)
-------------+----------------------------------------------------------------
1.HRLB       |
        Priv |
           0 |   -1.179109   1.329548    -0.89   0.376    -3.797022    1.438804
           1 |     -3.0185   1.083784    -2.79   0.006    -5.152499   -.8845017
------------------------------------------------------------------------------
```

Table 20: Marginal Effects of the data organization with privacy calculus

```
Predictive margins                              Number of obs     =        268
Model VCE    : Robust

Expression   : Linear prediction, predict()

1._at        : HRLB            =           0

2._at        : HRLB            =           1

------------------------------------------------------------------------------
             |                 Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _at#Priv |
        1 0  |   14.85575   .8496729    17.48   0.000     13.18272    16.52878
        1 1  |   16.55531   .5655766    29.27   0.000     15.44168    17.66895
        2 0  |   13.67664   1.009532    13.55   0.000     11.68885    15.66443
        2 1  |   13.53681   .8971397    15.09   0.000     11.77032     15.3033
------------------------------------------------------------------------------
```

Table 21: Predictive margins of the data organization with privacy calculus

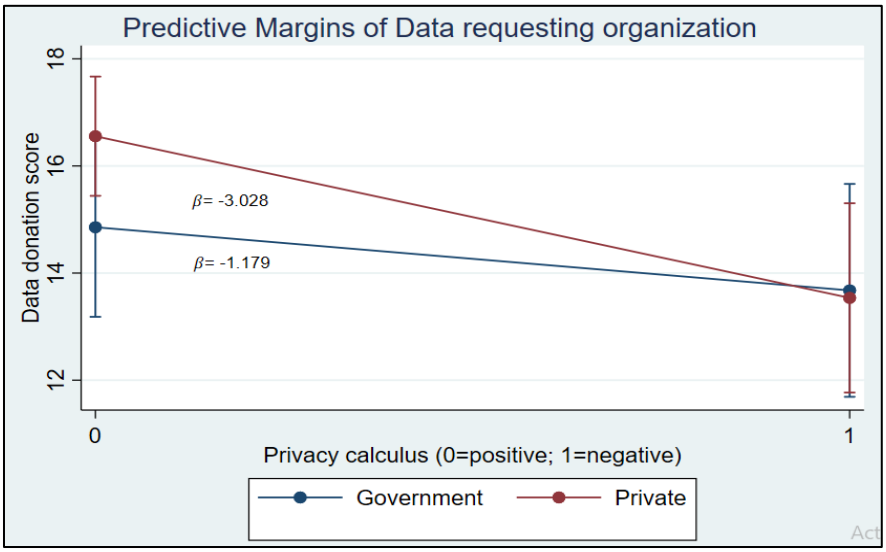

Figure 6: Margin plot of the conditions on Data donation score

6.5 Summary of results

Hypothesis	Support
H1: Perceived societal Privacy calculus will influence data donation decisions, such that individuals will donate more(less) data when they perceive net positive(negative) outcomes	Yes
H2: The data requesting organization will influence data donation decision, such that if the data requesting organization is a government one individuals will donate more data	No
H3: The type of data requesting organization will have an interaction effect on the privacy calculus, such that it will influence data donation decisions	Yes

Figure 7: Results of the hypothesis

The evidential support for the hypotheses were constructed through the ANOVA, t-test and the regression. *Figure 7* shows that H1 and H3 can be supported by empirical evidence. H1 is supported by the significant difference in means between individuals who received net negative privacy calculus and net positive privacy calculus for the dependent variable (*Score*). Individuals therefore will donate more personal health data when they perceive the outcome as a positive one. H2 is not supported as based upon the donation score there were not any significant differences in the means between individuals who received government and private. Lastly, support for H3 were seen throughout the analysis steps when the combined conditions were compared to each other. This was mostly seen for conditions in cases where the private condition was involved. The post hoc after regression concluded the findings that only the private organization had a significant interaction effect with the net privacy calculus on data donation scores. Thus, the evidential support for H3.

7 Discussion

7.1 Theoretical implications

The aim of the study was to examine the combined effects of the type of data requesting organization and net privacy calculus on health data donation decisions. Integrating the aspect of society on the privacy calculus and differentiating between the types of data organizations the study helps to expand on the privacy calculus and explore factors contributing to privacy behavior and disclosure. As such

distinction of the data organizations on the core privacy calculus relationships to observe health data donation has never been studied before.

First and foremost, in the study we found that participants age, gender, education and employment status did not correlate with privacy concerns and donation scores unlike some of the literature which supports this notion (Culnan & Armstrong, 1999; Hoy & Milne, 2010; Youn & Hall, 2008). (*Appendix 3*) The high amount of participants failure for government condition could be explained by them being more familiar with government interacions thus reducing attention (Mather, 2013). The results from the analysis section show consistency with the privacy calculus theory. An individual's privacy behavior and decision making depends on their assessment of the perceived risks and benefits associated with it. As seen with the support of H1, participants provided with the net positive privacy calculus had donated more health data than their counterparts. These participants also had fewer privacy concerns consistent to the findings in literature (Dinev et al., 2016). The evidence also does not show support for the privacy paradox as privacy concerns were inversely related to data donation scores. The study did not find a significant association of the type of data requesting organization on data donation scores, although the support of H3 shows that an interaction effect occurs when net privacy calculus is combined with the organizations. This was specifically significant for the private organization revealing that on interaction with positive privacy calculus data donation scores were highest among the four groups and were the lowest when interacting with negative privacy calculus. The difference in scores were also high between government and private when there is a positive privacy calculus. Linking to the Elaboration likelihood theory the stimuli of presenting the mock organization with varying risks and benefits makes individuals take the peripheral route of decision making through heuristics. The level of privacy concerns is therefore based on past experiences with government or private organizations among other factors. Based on the results, individuals having private organization as the data requesting entity showed to have highest data disclosure at the level of positive privacy calculus and lowest data disclosure at the level of negative privacy calculus. At the level of negative privacy calculus both the organizations had similar levels of data donation scores which implies similar perceived risks and benefits. On the other hand, the difference was greater at positive privacy calculus level with private condition leading in donation scores. This further implies that private organization affects the

privacy calculus of perceived benefits and risks to influence final decision-making process. This provides evidence for perceived risks and benefits in different situational contexts.

An interesting observation noted was that the mean data donation scores were above average in the overall experiment (14.87 out of max=21). This was also seen among all the manipulation conditions with the lowest score being for the negative privacy calculus group (12.72). Prior to starting the survey, we expected health donation scores to be low/below average, at least for all participants receiving negative privacy calculus as it consisted of higher risks of breaches and minimal benefits. Moreover, health data disclosure as Trumbo et al. (2007) argue is more personal and sensitive, having higher concerns and as presumed has less willingness for disclosure. The higher donation scores can therefore be associated with individuals weighing benefits greater than losses as Wilson & Valacich (2012) point out. While benefits in the study did not include any personal benefits it showed that increased data disclosure is associated with perceived societal benefits as in the mock scenario of our study which was to improve arthritis research. Theoretically increased scores could be attributed to introduction of the data requesting organization and its purpose thereby decreasing ambiguity and increasing engagement. However, the explanation for such observations of increased scores leans more towards literature stating individuals' social responsibility and prosocial behavior drives actions to help others specially in the healthcare context. (Skatova & Goulding., 2019; Steele et al., 2008; Clark et al., 1986).

The theoretical implications of results from the experimental study has provided privacy literature with a new understanding of societal privacy calculus and privacy decision making in the context of health data. The results suggest that the net outcome of the privacy calculus is a significant factor affecting privacy behavior and health data donation decisions. Furthermore, the study also reveals that the effects of the net privacy calculus are also enhanced if the data requesting organization is private which can influence the outcomes of data donation. Thus, the study helps contribute to theoretical and empirical privacy research as well as challenges faced in the healthcare context by individuals, organizations and policymakers which are elaborated in the next section.

7.2 Practical implications

A fine balance exists between individuals wanting to protect their personal data or using this data to help society through research or innovations. The research helps inform such dilemmas as there has been a limited number of studies discussing health data donations. On one hand in the information age data-driven healthcare provides us with massive improvements in research, treatment and personalization, whereas on the other data collection can lead to tremendous damages in both personal and societal levels through breaches and intrusion of privacy. While data donations inherently leave the decision-making process to the individuals, stakeholders like the data requesting organization, policy makers would benefit from being aware and understanding the decision-making process of the individuals.

The meaning of privacy is different for every individual is different since its value is highly dependent context specific (Acquisti et al., 2018). From the study we see that when private health data is shared individuals can expect varying levels of privacy concerns associated with it depending on the perceived risks and benefits. Individuals can help make a more desirable decision for themselves by understanding the whole process of data donation which includes information about the organization, purpose of donation and risks-benefit associated with it. Individuals can consider donations as a gift which transcends beyond the economic circle and what is deemed profitable as stated by Hénaff et al. (2010); although we add that such donations come with a tradeoff of benefits in the form of altruism and social belonging and risks in the form of privacy concerns. Individuals can therefore assess such long-term societal benefits and risks when making rational decisions. While the prosocial frame and altruistic tendencies of helping others overweigh the risks with data donation it is still important for individuals to understand risks are associated on a more personal level of the individual (Hummel et al., 2019).

The study also has implications for data donation organizations and policy makers in healthcare. The study shows that the effects of privacy concerns can be reduced by decreasing the privacy risks and increasing benefits. Data requesting organizations should look to improve benefits and reduce the risks as findings from the study reveal that such conditions would reduce privacy concerns and maximize data donation from individuals. It is also important for organizations to consider the societal aspects of data donation and highlight them. In the study we can assume that the mock scenario (improving

arthritis treatment) roused feelings of duty for some if not most individuals they were experimentally induced which Clark et al. (1986) highlights increase the frequency of actions to help others. Such findings could help improve voluntary data donations if organizations implement prosocial engagement scenarios provided they are true and within ethical boundaries. Equally important measures would be mechanisms to protect donated data. It is also important for both the government and private organizations to be transparent about the consequences data donation will provide to society. Transparency would help overcome asymmetric information about the disclosure and the inability to process costs/benefits which individuals have (Acquisti et al., 2013). This will lead will to increased trust, satisfaction and willingness for individuals to engage in data donation drives. Findings from the study indicate that individuals had lower and almost similar levels of data donation for both the private and government organizations if net privacy calculus was negative. When net privacy calculus was positive individuals preferred to significantly donate more data to a private organization than a government one. The study also found that private organizations have an interaction effect on the net privacy calculus which affects data donation scores. The results of the study have implications indicating that private organizations should work on negating risks associated with them while government organizations should work on creating assurances in the general population about the benefits associated with data donation. Given the variability in data donation decisions among participants one cannot fully assume to know what individuals want but instead call for organizations to have more control over their data. Organizational practices like easy opt out options of which type of health data they would not feel comfortable sharing would be a way to provide more control and a sense of security when in the process of making data donation decisions. Along with privacy protection mechanisms organizations can navigate dilemmas associated with health data donation by implementing privacy practices compliant with existing privacy laws.

7.2 Limitations and Future research

Prior to discussing the conclusion of the study, the limitations are discussed to suggest directions for future research. The first limitation of the study was that participants were selected only from a certain country (USA). Although it ensures quality it can be argued that the results would differ according to the country location. Bellman et al. (2004) mentions how different cultures of different countries

influences privacy behavior. Trust levels in the government are also different country wise (oecd.org, 2022). Moreover, different countries have different privacy regulatory laws and rights. Therefore, it can be assumed there would be a difference in the results if done for other countries and hence different implications. Such differences would help highlight whether the measures of extended privacy calculus and data donation organizations and contribute to a greater complexity of privacy behavior and decision making.

Second, all the participants were enrolled from the survey platform Prolific. They often share personal information on a daily basis through the platform and therefore the questions arise to the eternal validity and upward bias. Further studies could be conducted on another platform or a mock app to observe fluctuations from the result obtained. Third, the different conditions had varying sample sizes (Table 1). Although we accounted for it by analyzing and finding a significant effect of the data requesting organizations on the attention check failures, these variations in sample size between the four conditions affect statistical power.

Finally, although we were able to explain an amount of the variances in health data donation (Rsquared= 22%) the study did not account for other existing factors that can explain data donation decisions mainly personal privacy calculus. We examined four factors in a 2x2 factorial design (Gov.LRHB, GovHRLB, PrivLRHB and PrivHRLB) affecting decision making. Based on this it would be applicable to conduct an extension of study examining personal privacy calculus in the same study This alternate experiment would create the possibility to further contribute to the understanding and knowledge how privacy calculus and type data organization affect data donation decisions. The benefits of implementing personal privacy calculus in the model is that often the risks associated with data disclosure are personal therefore, it helps explore whether individuals would behave differently if personal benefits/risks were provided to them over societal benefits/risks. In such an instance would prosocial behaviors still prevail, and would there be an increase or decrease of data donation scores? This would potentially help to build on the findings of the study and help explain other variances unaccounted for in the study. The future studies could also implement a qualitative questionnaire to assess the reasons for variations in data donations.

8 Conclusion

As healthcare becomes increasingly digitized the promise to make use of health data is addressed by all the players involved with it: government, hospitals, private companies, and research organizations. Health data is growing exponentially and there is a lot of value derived from these health data such as identifying identify at-risk patients, personalized treatment, reducing scheduling times in hospitals, promoting research and development into novel treatment and drugs and best of all- saving lives. A novel and lesser-known concept for utilizing such data is through Health data donation. In the same way we donate blood or organs, data donation takes on the process where individuals are encouraged to donate their digital information for medical and academic research. Such drives provide individuals the liberty to choose whether and what they would like to donate. The concept of data donation is seemingly safer as it gives control over one's data. Nevertheless, negative consequences do exist when any kind of data is available and therefore individuals are faced with the dilemma of wanting to protect their personal data or using this data to help society and save lives. Healthcare as such provides an optimal platform for examining privacy concerns and decision-making behavior. Unlike other domains health data is unique in the way that it has emotions linked to it, is more sensitive and multiple stakeholders and present at different levels.

To understand data donation decisions, it becomes crucial to uncover the factors predicting and influencing this process. The most well recognized model in privacy literature is the Privacy calculus model which states that individuals behavior and decision-making process is dependent on the amount of risks and benefits associated with it. An individual weighs in both before making a final decision. In healthcare context the benefits associated with data donation are not on a personal level and often do not come with any incentives. Adding to this is that there is a limited number of studies that explore societal privacy calculus. Thus, it presents with an opportunity to add to privacy literature and understand its effect on donation decision. Few studies examine the distinction between the data requesting organization i.e., government and private, and those that do show variability in the results. Therefore, understanding its effect on behavior and decision-making process could uncover at least one of the many situational factors affecting decisions. The purpose of the study was hereby to examine

how the data requesting organizations in combination with the privacy calculus would impact data donation decisions.

The study consisted of a 2 (government vs. private) x 2 (net positive privacy calculus vs net negative privacy calculus) factorial experiment. As risk and benefits coexist, net positive privacy calculus was implied by Low risk & High benefit while net negative privacy calculus was High risk & Low benefit. A survey was carried out to gather data for analysis. The empirical findings show support for H1 and H3 as it was associated with a statistical significance and valid post hoc tests. Net positive privacy calculus was associated with increased donation scores (H1), while the data requesting organizations had similar levels of donation scores (H2) and was not statistically significant. The data requesting organization-specifically the private organization had an interaction effect on the privacy calculus and influenced data donation scores (H3). Private organization was associated with the highest as well as the lowest scores when combined with positive privacy calculus and negative privacy calculus respectively. Government organizations on the other hand had moderate score levels at both the positive and negative privacy calculus. Furthermore, participants who had less privacy concerns donated more data which was consistent with privacy literature. No indications of the privacy paradox were found. The donation scores were above average across all conditions which could be explained by prosocial behaviors. The results of the study thereby show that data requesting organization in combination with net privacy calculus show great implications on privacy behavior and data donation decisions. The study therefore fills the gap in privacy literature by expanding the privacy calculus and exploring the distinct data donation organizations.

The study contributes to the notion of more rational and desirable data donation decision making among the individuals, organization and policy makers' side. To address the variability in data donations among different organizations, the organizations are encouraged to provide more control to individuals. Key practices for better trust and participation would be being transparent about the process, including all the risks and benefits, and having data protection mechanisms in place. As observed from the study benefits outweigh risks when donating data; such findings call for data requesting organizations to adapt to practices that have increased societal benefit and creating

awareness about it. For policy makers the call would be to better balance privacy rights and prevent unauthorized data usage and sharing.

Despite the significant empirical research, the study includes limitations about the external validity and sample population. All the participants were based in the USA and were members of Prolific platform. Thus, the study was limited to account for cultural differences and had an upward bias as the participants were used to sharing information on the survey platform. Furthermore, not accounting for personal privacy calculus limited greater evidential support. Suggestions for future research would be implementing personal privacy calculus in the model and examining the effects of personal risks and benefits when combined with the data requesting organization and societal risks and benefits. This would provide capability to strengthen the study and provide additional findings.

It can be concluded that the data requesting organizations when combined with privacy calculus have significant impact on the privacy behavior and health data donation of individuals. The study therefore contributes with new knowledge and understanding of privacy behavior but further it also contributes to the challenges faced in healthcare by identifying decisions which would make data donation more desirable for individuals and improve data driven healthcare.

9 References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2018). *Privacy and human behavior in the age of information*.

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making [Article]. *IEEE Security & Privacy*, *3*(1), 26–33. https://doi.org/10.1109/MSP.2005.22

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What Is Privacy Worth? *The Journal of Legal Studies*, *42*(2), 249–274. https://doi.org/10.1086/671754

Acquisti, A., Taylor, C., & Wagman, L. (2016). *The Economics of Privacy *. https://doi.org/10.1257/jel.54.2.442

Alashoor, T., Fox, G., & Smith, H. J. (2017, December 9). *The Priming Effect of Prominent IS Privacy Concerns Scales on Disclosure Outcomes: An Empirical Examination*. https://www.researchgate.net/publication/321341361_The_Priming_Effect_of_Prominent_IS_Privacy_Concerns_Scales_on_Disclosure_Outcomes_An_Empirical_Examination

Alashoor, T., Han, S., & Joseph, R. C. (2017). Familiarity with Big Data, Privacy Concerns, and Self-Disclosure Accuracy in Social Networking Websites: An APCO Model. *Communications of the Association for Information Systems*, *41*, 62–96. https://doi.org/10.17705/1CAIS.04104

Alashoor, T., Keil, M., & Jiang, Z. (Jack). (2019). Data Donations for Advancing Medical Research: Mitigating the Negative Effect of Privacy Concerns. *Academy of Management Proceedings*, *2019*(1), 13405. https://doi.org/10.5465/AMBPP.2019.13405abstract

Anderson, C. L., & Agarwal, R. (2011). The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information [Article]. *Information Systems Research*, *22*(3), 469–490. https://doi.org/10.1287/isre.1100.0335

Angst, C. M. (2009). Protect My Privacy or Support the Common-Good? Ethical Questions About Electronic Health Information Exchanges [Article]. *Journal of Business Ethics*, *90*(Suppl 2), 169–178. https://doi.org/10.1007/s10551-010-0385-5

Angst, C. M., & Agarwal, R. (2009). Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood modeland individual persuasion. *MIS Quarterly: Management Information Systems*, *33*(2), 339–370. https://doi.org/10.2307/20650295

Ash, T. (2012, August 13). *Sharing is in our nature*. https://www.salon.com/2012/08/13/paul_seabright_on_evolution_salpart/

Awad, N., & Krishnan, M. (2006). The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization [Article]. *MIS Quarterly*, *30*(1), 13–28. https://doi.org/10.2307/25148715

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/J.TELE.2017.04.013

Batson, C. D., & Powell, A. A. (2003). Altruism and Prosocial Behavior. *Handbook of Psychology*, 463–484. https://doi.org/10.1002/0471264385.WEI0519

Beckerman, J. Z., Pritts, Joy., Goplerud, Eric., Leifer, J. C., Borzi, P. A., Rosenbaum, Sara., & Anderson, D. R. (2008). *A delicate balance: behavioral health, patient privacy, and the need to know - Digital Collections - National Library of Medicine*. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101513987-pdf

Bellman, S., Johnson, E. J., Kobrin, S. J., & Lohse, G. L. (2004). International Differences in Information Privacy Concerns: A Global Survey of Consumers. *The Information Society*, *20*(5), 313–324. https://doi.org/10.1080/01972240490507956

Buck, C., Dinev, T., & Anaraky, R. G. (2022). Revisiting APCO. *Modern Socio-Technical Perspectives on Privacy*, 43–60. https://doi.org/10.1007/978-3-030-82786-1_3

Cake, C., Ogburn, E., Pinches, H., Coleman, G., Seymour, D., Woodard, F., Manohar, S., Monsur, M., Landray, M., Dalton, G., Morris, A. D., Chinnery, P. F., Hobbs, F. D. R., & Butler, C. (2022). Development and evaluation of rapid data-enabled access to routine clinical information to enhance early recruitment to the national clinical platform trial of COVID-19 community treatments. *Trials*, *23*(1). https://doi.org/10.1186/S13063-021-05965-4

Cespedes, F. v, & Smith, H. J. (1993). Database Marketing: New Rules for Policy and Practice [Article]. *MIT Sloan Management Review*, *34*(4), 7.

Chong, I., Ge, H., Li, N., & Proctor, R. W. (2018). Influence of privacy priming and security framing on mobile app selection. *Computers & Security*, *78*, 143–154. https://doi.org/10.1016/j.cose.2018.06.005

Clark, M. S., Mills, J., & Powell, M. C. (1986). Keeping Track of Needs in Communal and Exchange Relationships. *Journal of Personality and Social Psychology*, *51*(2), 333–338. https://doi.org/10.1037/0022-3514.51.2.333

cloudian. (n.d.). *Data Protection and Privacy: 12 Ways to Protect User Data*. Retrieved December 29, 2022, from https://cloudian.com/guides/data-protection/data-protection-and-privacy-7-ways-to-protect-user-data/

Cox, J., Nguyen, T., Thorpe, A., Ishizaka, A., Chakhar, S., & Meech, L. (2018). Being seen to care: The relationship between self-presentation and contributions to online pro-social crowdfunding campaigns [Article]. *Computers in Human Behavior*, *83*, 45–55. https://doi.org/10.1016/j.chb.2018.01.014

Culnan, M. J., & Armstrong, P. K. (1999). Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation [Article]. *Organization Science (Providence, R.I.)*, *10*(1), 104–115. https://doi.org/10.1287/orsc.10.1.104

Culnan, M. J., & Bies, R. J. (2003). Consumer Privacy: Balancing Economic and Justice Considerations. *Journal of Social Issues*, *59*(2), 323–342. https://doi.org/10.1111/1540-4560.00067

Dhami, S., Thompson, D., el Akoum, M., Bates, D. W., Bertollini, R., & Sheikh, A. (2022). Data-enabled responses to pandemics: policy lessons from COVID-19. *Nature Medicine*, *28*(11). https://doi.org/10.1038/S41591-022-02054-0

Dienlin, T., & Metzger, M. J. (2016). An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Dinev, T., Albano, V., Xu, H., D'Atri, A., & Hart, P. (2016). *Individuals' Attitudes Towards Electronic Health Records: A Privacy Calculus Perspective*. 19–50. https://doi.org/10.1007/978-3-319-23294-2_2

Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Https://Doi.Org/10.1287/Isre.1060.0080*, *17*(1), 61–80. https://doi.org/10.1287/ISRE.1060.0080

Dinev, T., McConnell, A. R., & Jeff Smith, H. (2015). Research Commentary- Informing Privacy Research Through Information Systems, Psychology, and Behavioral Economics: Thinking Outside the "APCO" Box. *Https://Doi.Org/10.1287/Isre.2015.0600*, *26*(4), 639–655. https://doi.org/10.1287/ISRE.2015.0600

Dron, L., Kalatharan, V., Gupta, A., Haggstrom, J., Zariffa, N., Morris, A. D., Arora, P., & Park, J. (2022). Data capture and sharing in the COVID-19 pandemic: a cause for concern. *The Lancet Digital Health*, *4*(10), e748–e756. https://doi.org/10.1016/S2589-7500(22)00147-9

el Majdoubi, D., el Bakkali, H., Sadki, S., Maqour, Z., & Leghmid, A. (2022). The Systematic Literature Review of Privacy-Preserving Solutions in Smart Healthcare Environment. *Security and Communication Networks*, *2022*, 1–26. https://doi.org/10.1155/2022/5642026

Evans, R., & Ferguson, E. (2014). Defining and measuring blood donor altruism: a theoretical approach from biology, economics and psychology. *Vox Sanguinis*, *106*(2), 118–126. https://doi.org/10.1111/vox.12080

Fridman, A., Gershon, R., & Gneezy, A. (2022). Increased generosity under COVID-19 threat. *Scientific Reports* |, *12*, 4886. https://doi.org/10.1038/s41598-022-08748-2

gdpr.eu. (n.d.). *What is GDPR, the EU's new data protection law?*

Gellert, R., & Gutwirth, S. (2013). The legal construction of privacy and data protection [Article]. *The Computer Law and Security Report*, *29*(5), 522–530. https://doi.org/10.1016/j.clsr.2013.07.005

Gerber, N., Gerber, P., & Volkamer, M. (2018). Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, *77*, 226–261. https://doi.org/10.1016/j.cose.2018.04.002

Girdhari, S., & Ndayizigamiye, P. (2022). Adoption of Blockchain to Support the National Health Insurance Implementation in South Africa: An Integrative Review. *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/978-1-7998-8915-1.Ch009*, 194–228. https://doi.org/10.4018/978-1-7998-8915-1.CH009

Glaser, J., Henley, D. E., Downing, G., & Brinner, K. M. (2008). Advancing Personalized Health Care through Health Information Technology: An Update from the American Health Information Community's Personalized Health Care Workgroup [Article]. *Journal of the American Medical Informatics Association : JAMIA*, *15*(4), 391–396. https://doi.org/10.1197/jamia.M2718

Glynn, S. A., Busch, M. P., Schreiber, G. B., Murphy, E. L., Wright, D. J., Tu, Y., Kleinman, S. H., & for the NHLBI REDS Study Group. (2003). Effect of a National Disaster on Blood Supply and Safety. *JAMA*, *289*(17), 2246. https://doi.org/10.1001/jama.289.17.2246

Gu, J., Xu, Y. (Calvin), Xu, H., Zhang, C., & Ling, H. (2017). Privacy concerns for mobile app download: An elaboration likelihood model perspective. *Decision Support Systems*, *94*, 19–28. https://doi.org/10.1016/j.dss.2016.10.002

Harris, P. (2007). *Many US adults are satisfied with use of their personal health information*. https://harris-interactive.co.uk/

Harvey, J., Smith, A., Goulding, J., & Branco Illodo, I. (2019). Food Sharing, Redistribution, and Waste Reduction via Mobile Applications: A Social Network Analysis. *Industrial Marketing Management*, *88*, 437–448. https://doi.org/10.1016/J.INDMARMAN.2019.02.019

Hénaff, Marcel., Morhange, J.-L., & Feenberg-Dibon, A.-M. (2010). *The price of truth : gift, money, and philosophy*. 466.

Hermalin, B. E., & Katz, M. L. (2006). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics*, *4*(3), 209–239. https://doi.org/10.1007/S11129-005-9004-7

Hoy, M. G., & Milne, G. (2010). Gender Differences in Privacy-Related Measures for Young Adult Facebook Users [Article]. *Journal of Interactive Advertising*, *10*(2), 28–45. https://doi.org/10.1080/15252019.2010.10722168

Hummel, P., Braun, M., & Dabrock, P. (2019). *Data Donations as Exercises of Sovereignty* (pp. 23–54). https://doi.org/10.1007/978-3-030-04363-6_3

Kim, D., Park, K., Park, Y., & Ahn, J.-H. (2019). Willingness to provide personal information: Perspective of privacy calculus in IoT services. *Computers in Human Behavior*, *92*, 273–281. https://doi.org/10.1016/j.chb.2018.11.022

Kitsos, P., & Pappa, P. (2015). Mobile Communications Privacy. *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/978-1-4666-5888-2.Ch202*, 2097–2105. https://doi.org/10.4018/978-1-4666-5888-2.CH202

Kost, E. (2022, December 22). *14 Biggest Healthcare Data Breaches*. https://www.upguard.com/blog/biggest-data-breaches-in-healthcare

Krasnova, H., Veltri, N. F., & Günther, O. (2012). Self-disclosure and privacy calculus on social networking sites: The role of culture intercultural dynamics of privacy calculus. *Business and Information Systems Engineering*, *4*(3), 127–135. https://doi.org/10.1007/S12599-012-0216-6/METRICS

Krutzinna, J., & Floridi, L. (2019). *The Ethics of Medical Data Donation*. *137*, 198. https://doi.org/10.1007/978-3-030-04363-6

Kung, F. Y. H., Kwok, N., & Brown, D. J. (2018). Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology*, *67*(2), 264–283. https://doi.org/10.1111/apps.12108

Laerd statistics. (n.d.). *One-way ANOVA in SPSS Statistics*. https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

Laudon, K. C. (1996). Markets and privacy. *Communications of the ACM*, *39*(9), 92–104. https://doi.org/10.1145/234215.234476

Laufer, R. S., & Wolfe, M. (1977). Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. *Journal of Social Issues*, *33*(3), 22–42. https://doi.org/10.1111/J.1540-4560.1977.TB01880.X

Li, H., Luo, X. (Robert), Zhang, J., & Xu, H. (2017). Resolving the privacy paradox: Toward a cognitive appraisal and emotion approach to online privacy behaviors. *Information & Management*, *54*(8), 1012–1022. https://doi.org/10.1016/j.im.2017.02.005

Lin, X., Huang, X., Liu, S., Li, Y., Luo, H., & Yu, S. (2021). Social Welfare Analysis under Different Levels of Consumers' Privacy Regulation. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(7), 2943–2964. https://doi.org/10.3390/jtaer16070161

Luccasen, A., & Grossman, P. J. (2017). WARM-GLOW GIVING: EARNED MONEY AND THE OPTION TO TAKE. *Economic Inquiry*, *55*(2), 996–1006. https://doi.org/10.1111/ecin.12417

Lukács, A. (2016). *WHAT IS PRIVACY? THE HISTORY AND DEFINITION OF PRIVACY*.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Marr Bernard. (2016, March 9). *Big Data: How A Big Business Asset Turns Into A Huge Liability*. https://www.forbes.com/sites/bernardmarr/2016/03/09/big-data-how-a-big-business-asset-turns-into-a-huge-liability/?sh=52af548c7761

Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. C. (2017). "Now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization*, *140*, 1–17. https://doi.org/10.1016/j.jebo.2017.03.024

Mather, E. (2013). Novelty, attention, and challenges for developmental psychology. Frontiers in Psychology vol. 4 https://doi.org/10.3389/fpsyg.2013.00491

Mcgraw, D., & Mandl, K. D. (n.d.). *Privacy protections to encourage use of health-relevant digital data in a learning health system*. https://doi.org/10.1038/s41746-020-00362-8

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Montgomery, J. (2017). Data Sharing and the Idea of Ownership. *The New Bioethics*, *23*(1), 81–86. https://doi.org/10.1080/20502877.2017.1314893

Nickerson, C. (2022). *Elaboration Likelihood Model. Simply Psychology*. https://misq.umn.edu/influence-processes-for-information-technology-acceptance-an-elaboration-likelihood-model.html

oecd.org. (2022). *Governments seen as reliable post-pandemic but giving citizens greater voice is critical to strengthening trust- OECD*. https://www.oecd.org/newsroom/governments-seen-as-reliable-post-pandemic-but-giving-citizens-greater-voice-is-critical-to-strengthening-trust.htm

Ozdemir, Z. D., Jeff Smith, H., & Benamati, J. H. (2018). Antecedents and outcomes of information privacy concerns in a peer context: An exploratory study. *Https://Doi.Org/10.1057/S41303-017-0056-z*, *26*(6), 642–660. https://doi.org/10.1057/S41303-017-0056-Z

Petty, R. E., & Cacioppo, J. T. (1986). *The Elaboration Likelihood Model of Persuasion* (pp. 123–205). https://doi.org/10.1016/S0065-2601(08)60214-2

Posner, R. (1978). The Right of Privacy. *Sibley Lecture Series*. https://digitalcommons.law.uga.edu/lectures_pre_arch_lectures_sibley/22

Posner, R. (1981). The Economics of Privacy. *American Economic Review*, *71*. https://chicagounbound.uchicago.edu/journal_articles/620

Prainsack, B. (2019). *Data Donation: How to Resist the iLeviathan* (pp. 9–22). https://doi.org/10.1007/978-3-030-04363-6_2

Princi, E., & Krämer, N. C. (2020). Out of Control – Privacy Calculus and the Effect of Perceived Control and Moral Considerations on the Usage of IoT Healthcare Devices. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.582054

Rachels, J. (1975). Why Privacy is Important [Article]. *Philosophy & Public Affairs*, *4*(4), 323–333.

Radu, V. B. (2022). "Building and Maintaining Trust in Public Institutions During the Coronavirus Pandemic. A Theoretical Perspective." *Transylvanian Review of Administrative Sciences*, *66E*, 64–80. https://doi.org/10.24193/tras.66E.4

RBC Capital Markets | The healthcare data explosion. (n.d.). Retrieved December 28, 2022, from https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion

Rodrigues, S. M., LeDoux, J. E., & Sapolsky, R. M. (2009). The Influence of Stress Hormones on Fear Circuitry. *Annual Review of Neuroscience*, *32*(1), 289–313. https://doi.org/10.1146/annurev.neuro.051508.135620

Rohm, A. J., & Milne, G. R. (2004). Just what the doctor ordered: The role of information sensitivity and trust in reducing medical information privacy concern [Article]. *Journal of Business Research*, *57*(9), 1000–1011. https://doi.org/10.1016/S0148-2963(02)00345-4

Schwartz, S. H., & Bilsky, W. (1990). Toward a Theory of the Universal Content and Structure of Values: Extensions and Cross-Cultural Replications. *Journal of Personality and Social Psychology*, *58*(5), 878–891. https://doi.org/10.1037/0022-3514.58.5.878

Sharyl J . Nass, Laura A . Levit, & Lawrence O . Gostin. (2019). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*.

Shaw, D. M. (2020). Defining Data Donation After Death: Metadata, Families, Directives, Guardians and the Route to Big Consent. *The Ethics of Medical Data Donation*. http://europepmc.org/books/NBK554074

Sheringham, J., Kuhn, I., & Burt, J. (2021). The use of experimental vignette studies to identify drivers of variations in the delivery of health care: a scoping review. *BMC Medical Research Methodology*, *21*(1), 81. https://doi.org/10.1186/s12874-021-01247-4

Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. *PLOS ONE*, *14*(11), e0224240. https://doi.org/10.1371/JOURNAL.PONE.0224240

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly: Management Information Systems*, *35*(4), 989–1015. https://doi.org/10.2307/41409970

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996a). Information Privacy: Measuring Individuals' Concerns about Organizational Practices [Article]. *MIS Quarterly*, *20*(2), 167–196. https://doi.org/10.2307/249477

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996b). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly: Management Information Systems*, *20*(2), 167–195. https://doi.org/10.2307/249477

Sojka, B. N., & Sojka, P. (2008). The blood donation experience: self-reported motives and obstacles for donating blood. *Vox Sanguinis*, *94*(1), 56–63. https://doi.org/10.1111/J.1423-0410.2007.00990.X

Solove, D. J. (2002). Conceptualizing Privacy [Article]. *California Law Review*, *90*(4), 1087–1155. https://doi.org/10.2307/3481326

Stanford Medicine Health Trends. (2017).

Steele, W. R., Schreiber, G. B., Guiltinan, A., Nass, C., Glynn, S. A., Wright, D. J., Kessler, D., Schlumpf, K. S., Tu, Y., Smith, J. W., & Garratty, G. (2008). The role of altruistic behavior, empathetic concern, and social responsibility motivation in blood donation behavior. *Transfusion*, *48*(1), 43–54. https://doi.org/10.1111/J.1537-2995.2007.01481.X

Stigler, G. (1980). An Introduction to Privacy in Economics and Politics. *Journal of Legal Studies*, *9*(4). https://chicagounbound.uchicago.edu/jls/vol9/iss4/2

Taddeo, M. (2016). Data philanthropy and the design of the infraethics for information societies. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083). https://doi.org/10.1098/RSTA.2016.0113

Tech Republic. (2020, July 30). *Facebook data privacy scandal: A cheat sheet*. TechRepublic.

Terracciano, A., McCrae, R. R., Hagemann, D., & Costa, P. T. (2003). Individual Difference Variables, Affective Differentiation, and the Structures of Affect. *Journal of Personality*, *71*(5), 669–704. https://doi.org/10.1111/1467-6494.7105001

The Economist | The world's most valuable resource is no longer oil, but data. (2017). https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Trumbo, C. W., McComas, K. A., & Kannaovakun, P. (2007). Cancer Anxiety and the Perception of Risk in Alarmed Communities. *Risk Analysis*, *27*(2), 337–350. https://www.academia.edu/29008526/Cancer_Anxiety_and_the_Perception_of_Risk_in_Alarmed_Communities

Vaidya, M. (2014). Ice bucket challenge cash may help derisk ALS drug research. *Nature Medicine*, *20*(10), 1080. https://doi.org/10.1038/NM1014-1080

Varian, H. R. (2002). Economic Aspects of Personal Privacy. *Cyber Policy and Economics in an Internet Age*, 127–137. https://doi.org/10.1007/978-1-4757-3575-8_9

Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Law Review*, *4*(5), 193–220. http://links.jstor.org/sici?sici=0017-811X%2818901215%294%3A5%3C193%3ATRTP%3E2.0.CO%3B2-C

Westin, A. F. (1976). *Computers, Health Records, and Citizen Rights*. https://books.google.dk/books?id=mi1CAAAAIAAJ&printsec=frontcover&source=gbs_atb&redir_esc=y#v=onepage&q&f=false

Wilson, D., & Valacich, J. (2012). Unpacking the Privacy Paradox: Irrational Decision-Making within the Privacy Calculus. *ICIS 2012 Proceedings*. https://aisel.aisnet.org/icis2012/proceedings/ResearchInProgress/101

Witti, M., & Konstantas, P. D. (2019). *SECURE AND PRIVACY-AWARE DATA COLLECTION ARCHITECTURE APPROACH IN FOG NODE BASED DISTRIBUTED IOT ENVIRONMENT*. 19–32. https://doi.org/10.5121/csit.2019.91302

Xiang, D., & Cai, W. (2021). Privacy Protection and Secondary Use of Health Data: Strategies and Methods. *BioMed Research International*, *2021*, 1–11. https://doi.org/10.1155/2021/6967166

Xu, F., Michael, K., & Chen, X. (2013). Factors affecting privacy disclosure on social network sites: an integrated model. *Electronic Commerce Research*, *13*(2), 151–168. https://doi.org/10.1007/s10660-013-9111-6

Xu, H., Teo, H.-H., Tan, B. C. Y., & Agarwal, R. (2012). Effects of Individual Self-Protection, Industry Self-Regulation, and Government Regulation on Privacy Concerns: A Study of Location-Based Services. *Information Systems Research*, *23*(4), 1342–1363. https://doi.org/10.1287/isre.1120.0416

Youn, S., & Hall, K. H. (2008). *Gender and online privacy among teens: Risk perception, privacy concerns, and protection behaviors.*

Zhu, M., Wu, C., Huang, S., Zheng, K., Young, S. D., Yan, X., & Yuan, Q. (2021). Privacy paradox in mHealth applications: An integrated elaboration likelihood model incorporating privacy calculus and privacy fatigue. *Telematics and Informatics*, *61*, 101601. https://doi.org/10.1016/j.tele.2021.101601

Appendix

Appendix 1. Experimental design

Consent

The ART Foundation Health Data Donation Research

ABOUT THIS RESEARCH
You are being asked to participate in a research study. Scientists do research to answer questions and learn new information. Some research might help change or improve the way we do things in the future. This consent information will tell you more about the study to help you decide whether you want to participate. Please read this information before agreeing to be in the study.

TAKING PART IN THIS STUDY IS VOLUNTARY
You may choose not to take part in the study or may choose to leave the study at any time. Deciding not to participate, or deciding to leave the study later, will not result in any penalty and will not affect your relationship with the researchers conducting this study. Participation is voluntary and you can stop the survey at any time without penalty. As an alternative to participating in the study, you may choose not to take part.

WHY IS THIS STUDY BEING DONE? The purpose of this study is to investigate the ART foundation health data donation research.

WHAT WILL HAPPEN DURING THE STUDY? If you agree to be in the study, you will be asked to do the following things:
• You will be asked to answer demographic questions.
• You will be provided information about the ART foundation research.
• You will be asked whether you wish to make a data donation.
• You will be asked a number of questions about your attitude towards data donation.
• The study will take approximately 7 minutes of your time.

WHAT ARE THE RISKS OF TAKING PART IN THE STUDY? We do not anticipate any risks to you participating other than those encountered in daily life. Nevertheless, while completing the survey, you can tell the researchers that you feel uncomfortable or that you do not want to answer a particular question.

WHAT ARE THE POTENTIAL BENEFITS OF TAKING PART IN THE STUDY? We hope to contribute to science and health data research.

HOW WILL MY INFORMATION BE PROTECTED? We will not be accessing any personally identifying information about you. Efforts will be made to keep your personal information confidential. Your personal information may

be disclosed if required by law. No information which could identify you will be shared in publications about this study. Organizations that may inspect and/or copy your research records for quality assurance and data analysis include groups such as the study investigators and their research associates and the research ethics committee.

WILL MY INFORMATION BE USED FOR RESEARCH IN THE FUTURE? Information collected from you for this study may be used for future research studies or shared with other researchers for future research. If this happens, information which could identify you will be removed before any information are shared. Since identifying information will be removed, we will not ask for your additional consent.

WHO SHOULD I CALL WITH QUESTIONS OR PROBLEMS? For questions about the study, contact the researcher, Pratyush Khanra at prkh20ab@student.cbs.dk who would be happy to address your questions or concerns.

PARTICIPANT'S CONSENT

In consideration of all of the above, I give my consent to participate in this research study. By proceeding, I confirm that I am 18 years old, and agree to take part in this study.

Basic information

How old are you? Answer: _____

Which gender do you identify to? Answer: _____

What is your nationality? Answer: _____

What is your highest level of education? Answer: _____

Are you currently employed? Answer: _____

Introduction and Scenario randomization

Condition 1: Government + Net positive outcome	Condition 2: Government + Net negative outcome
 The ART Foundation, a **government-funded project**, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals.	 The ART Foundation**,** a **government-funded project**, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals.

<table>
<tr><td>Experts believe that this data donation governmental research is expected to have:
3. Low Privacy Risks to Society. The reason is that this research has a low likelihood of information breaches because it is subjected to data protection laws.
4. High Benefits to Society. The reason is that this research could significantly reduce the prevalence of arthritis in the population.</td><td>Experts believe that this data donation governmental research is expected to have:
3. High Privacy Risks to Society. The reason is that this research has a high likelihood of information breaches because it is not subjected to data protection laws.
4. Low Benefits to Society. The reason is that this research could overestimate projections and only have slight improvement in arthritis treatment.</td></tr>
<tr><th>Condition 3: Private + Net positive outcome</th><th>Condition 4: Private + Net negative outcome</th></tr>
<tr><td> 
The ART Foundation, a private-funded project, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals.

Experts believe that this data donation private research is expected to have:
3. Low Privacy Risks to Society. The reason is that this research has a low likelihood of information breaches because it is subjected to data protection laws.
4. High Benefits to Society. The reason is that this research could significantly reduce the prevalence of arthritis in the population.</td><td> 
The ART Foundation, a private-funded project, is undertaking research with the goal of reducing arthritis prevalence by 85% within the year 2027. To achieve this, the ART foundation needs to collect and analyze thousands of medical information donated by individuals.

Experts believe that this data donation private research is expected to have:
3. High Privacy Risks to Society. The reason is that this research has a high likelihood of information breaches because it is not subjected to data protection laws.
4. Low Benefits to Society. The reason is that this research could overestimate projections and only have slight improvement in arthritis treatment.</td></tr>
</table>

<u>Privacy concerns and Data donation Questionnaire</u> (Randomized order)

<u>Privacy concerns</u>: For each of the following indicate whether you agree or disagree with the statements. *(7-point Likert scale where 1 is Strongly disagree and 7 is Strongly agree)*.

1. I am concerned that the information I donate to the ART foundation research could be misused.
2. I am concerned about donating information to the ART foundation because of what others might do with it.

3. I am concerned about donating information to the ART foundation because it could be used in a way I did not foresee.
4. I am concerned about donating information to the ART foundation because others can find private information about me.

Health data donation questionnaire

Below is a list of health data that you can donate to the ART foundation research.
We would like to let you know that the data you donate will be completely anonymous and made available only to licensed researchers at participating institutions. Your personal data will never be shared with or sold to third parties.

What is your height?

Donate below I prefer not to donate this information

What is your weight?

Donate below I prefer not to donate this information

What is your blood type?

Donate below I prefer not to donate this information

How much do you sleep on an average night? (Hours per night)

Donate below I prefer not to donate this information

What is your diet?
(Vegetarian, non-vegetarian, vegan, others)

Donate below I prefer not to donate this information

Do you have any known allergies?

Donate below I prefer not to donate this information

Have you been vaccinated for Covid-19?

Donate below I prefer not to donate this information

Do you take any medication?

Donate below I prefer not to donate this information

Have you undergone any surgery within the past 10 years?

Donate below I prefer not to donate this information

How frequently do you drink alcohol per week?

Donate below I prefer not to donate this information

How frequently do you smoke cigarettes per week?

Donate below I prefer not to donate this information

How frequently do you smoke/consume marijuana per week?

Donate below I prefer not to donate this information

Have you ever used any drugs/narcotics?

Donate below I prefer not to donate this information

How many hours do you spend exercising per week?

Donate below I prefer not to donate this information

How many hours do you spend on social media per day?

Donate below I prefer not to donate this information

Do you currently have any acute disease? (e.g.: Common cold, pneumonia, measles, flu, etc.)

Donate below I prefer not to donate this information

Have you been diagnosed with any chronic disease?
(e.g.: Hypertension, diabetes, arthritis, coronary heart disease, etc.)

Donate below I prefer not to donate this information

Do any of your immediate family members have any chronic diseases?

Donate below I prefer not to donate this information

Have you been diagnosed with any mental conditions/disorders?
(e.g.: Depression, anxiety, OCD, PTSD, etc.)

Donate below I prefer not to donate this information

How frequently do you go for a health check-up? (Per year)

Donate below I prefer not to donate this information

Are you covered by a health insurance?

Donate below I prefer not to donate this information

Manipulation check

Please indicate how much you agree with the following statements: (*7-point Likert scale where 1 is Strongly disagree and 7 is Strongly agree*).

1. I believe that the ART foundation research will be protected against data privacy breaches
2. I believe that the ART foundation research will provide a lot of benefits to society

Attention check

These questions test whether you attentively read the questions and understood the survey scenario.

The ART foundation is a:

[Single selection: Government organization; Private organization; I don't remember]

The ART foundation research will have:

[Single selection: Low privacy risks and High benefits to society; High privacy risks and Low benefits to society; I don't remember]

Conclusion

Individuals' responses to health data donation vary with various situational factors like age, education, etc. One such situation is when an individual is faced between the choice of donating data to a government compared to a private organization. This survey is part of a research that aims to explore and contribute to this topic.

Thank you for participating in our study. Please tell us if you have any feedback about this study.

Please feel free to provide your thoughts on this topic and feedback on the survey below.

End of survey

Appendix 2. Results without exclusion

```
. tabstat Score, statistics( count mean sd semean min max cv ) by(Priv)

Summary for variables: Score
     by categories of: Priv (private)

      Priv |         N      mean        sd  se(mean)       min       max        cv
-----------+--------------------------------------------------------------------
Government |       203  14.77833  7.197024   .505132         0        21  .4869986
   Private |       199  14.67337  7.405215  .5249418         0        21  .5046705
-----------+--------------------------------------------------------------------
     Total |       402  14.72637  7.291901   .363687         0        21  .4951595
-------------------------------------------------------------------------------

. tabstat Score, statistics( count mean sd semean min max cv ) by(HRLB)

Summary for variables: Score
     by categories of: HRLB (HRLB)

  HRLB |         N      mean        sd  se(mean)       min       max        cv
-------+--------------------------------------------------------------------
  LRHB |       202  16.19802   6.22628  .4380795         0        21  .3843853
  HRLB |       200     13.24  7.973095   .563783         0        21  .6021975
-------+--------------------------------------------------------------------
 Total |       402  14.72637  7.291901   .363687         0        21  .4951595
-------------------------------------------------------------------------------
```

```
. tabstat PC_avg, statistics( count mean sd semean min max cv ) by(Priv)

Summary for variables: PC_avg
     by categories of: Priv (private)

     Priv |         N      mean        sd  se(mean)       min       max        cv
----------+--------------------------------------------------------------------
Government |       203  3.881773  1.874843  .1315882         1         7  .4829863
   Private |       199  4.204774  1.644693  .1165892         1         7   .391149
----------+--------------------------------------------------------------------
    Total |       402  4.041667  1.769874  .0882733         1         7  .4379071
----------------------------------------------------------------------------------


. tabstat PC_avg, statistics( count mean sd semean min max cv ) by(HRLB)

Summary for variables: PC_avg
     by categories of: HRLB (HRLB)

  HRLB |         N      mean        sd  se(mean)       min       max        cv
-------+--------------------------------------------------------------------
  LRHB |       202  3.649752  1.748357   .123014         1         7  .4790345
  HRLB |       200    4.4375  1.706246  .1206498         1         7  .3845061
-------+--------------------------------------------------------------------
 Total |       402  4.041667  1.769874  .0882733         1         7  .4379071
----------------------------------------------------------------------------------
```

Table 22: Descriptive statistics of manipulation conditions (separated)

```
-> priming = 0

Summary statistics: N, mean, sd, se(mean), min, max, cv
  by categories of: HRLB (HRLB)

  HRLB |     Score    PC_avg
-------+----------------------
  LRHB |        99        99
       |  17.15152  3.414141
       |  5.815819   1.76132
       |  .5845118  .1770193
       |         0         1
       |        21         7
       |  .3390848  .5158895
-------+----------------------
  HRLB |        97        97
       |  12.60825   4.32732
       |  8.214008  1.731059
       |  .8340062  .1757624
       |         0         1
       |        21         7
       |   .651479  .4000302
-------+----------------------
 Total |       196       196
       |  14.90306  3.866071
       |  7.443172  1.801064
       |  .5316552  .1286474
       |         0         1
       |        21         7
       |  .4994391   .465864
-------------------------------
```

```
-> priming = 1

Summary statistics: N, mean, sd, se(mean), min, max, cv
  by categories of: HRLB (HRLB)

  HRLB |     Score    PC_avg
-------+----------------------
  LRHB |       103       103
       |  15.28155  3.876214
       |  6.493462  1.713869
       |  .6398198  .1688725
       |         0         1
       |        21         7
       |  .4249216  .4421502
-------+----------------------
  HRLB |       103       103
       |  13.83495  4.541262
       |  7.732155  1.684378
       |  .7618718  .1659667
       |         0         1
       |        21         7
       |  .5588856  .3709052
-------+----------------------
 Total |       206       206
       |  14.55825  4.208738
       |  7.159098  1.727502
       |   .498798  .1203608
       |         0         1
       |        21         7
       |  .4917553  .4104562
-------------------------------
```

Table 23: Descriptive statistics of manipulation conditions (separated with primed=1)

```
. tabstat Score, statistics( count mean sd semean min max cv ) by(Cond)

Summary for variables: Score
     by categories of: Cond (FL_16_DO)

     Cond |         N      mean        sd  se(mean)       min       max        cv
----------+--------------------------------------------------------------------
 Gov.LRHB |       101  15.69307  6.523408  .6491034         0        21  .4156872
 Gov.HRLB |       102  13.87255  7.732753  .7656567         0        21   .557414
Priv.LRHB |       101  16.70297  5.903464  .5874167         0        21   .353438
Priv.HRLB |        98  12.58163  8.203499  .8286785         0        21  .6520218
----------+--------------------------------------------------------------------
    Total |       402  14.72637  7.291901   .363687         0        21  .4951595
----------+--------------------------------------------------------------------
```

```
. tabstat PC_avg, statistics( count mean sd semean min max cv ) by(Cond)

Summary for variables: PC_avg
     by categories of: Cond (FL_16_DO)

     Cond |         N      mean        sd  se(mean)       min       max        cv
----------+--------------------------------------------------------------------
 Gov.LRHB |       101  3.576733  1.819596  .1810565         1         7  .5087312
 Gov.HRLB |       102  4.183824  1.888582  .1869974         1         7  .4514009
Priv.LRHB |       101  3.722772  1.679993  .1671655         1         7  .4512747
Priv.HRLB |        98  4.701531   1.45626  .1471045         2         7  .3097417
----------+--------------------------------------------------------------------
    Total |       402  4.041667  1.769874  .0882733         1         7  .4379071
----------+--------------------------------------------------------------------
```

Table 24: Descriptive statistics of manipulation conditions (combined)

```
. oneway Score Priv

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      1.10702275      1   1.10702275      0.02     0.8855
 Within groups      21320.7935    400   53.3019837
------------------------------------------------------------------------
    Total           21321.9005    401   53.1718217

Bartlett's test for equal variances:  chi2(1) =   0.1622  Prob>chi2 = 0.687

. oneway Score HRLB

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups       879.34129      1    879.34129     17.21     0.0000
 Within groups      20442.5592    400    51.106398
------------------------------------------------------------------------
    Total           21321.9005    401   53.1718217

Bartlett's test for equal variances:  chi2(1) =  12.0875  Prob>chi2 = 0.001
```

```
. oneway Score Cond

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      1014.13616      3   338.045388      6.63     0.0002
 Within groups      20307.7643    398   51.0245335
------------------------------------------------------------------------
    Total           21321.9005    401   53.1718217

Bartlett's test for equal variances:  chi2(3) =  13.3461  Prob>chi2 = 0.004
```

```
. oneway PC_avg Cond

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      76.8360586      3   25.6120195      8.64     0.0000
 Within groups      1179.27852    398   2.96301137
------------------------------------------------------------------------
    Total           1256.11458    401   3.13245532

Bartlett's test for equal variances:  chi2(3) =   7.4421  Prob>chi2 = 0.059
```

Table 25: One way ANOVA of all conditions

Comparison of Data donation score by Combined manipulation conditions
(Bonferroni)

Row Mean-Col Mean	Gov.LRHB	Gov.HRLB	Priv.LRH
Gov.HRLB	-1.82052 0.421		
Priv.LRH	1.0099 1.000	2.83042 0.030	
Priv.HRL	-3.11144 0.014	-1.29092 1.000	-4.12134 0.000

Comparison of privacy concerns by Combined manipulation conditions
(Bonferroni)

Row Mean-Col Mean	Gov.LRHB	Gov.HRLB	Priv.LRH
Gov.HRLB	.607091 0.074		
Priv.LRH	.14604 1.000	-.461051 0.343	
Priv.HRL	1.1248 0.000	.517707 0.205	.978758 0.000

Table 26: Bonferroni post- hoc test between the combined conditions

```
. ttest Score, by(Priv)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Governme	203	14.77833	.505132	7.197024	13.78232	15.77433
Private	199	14.67337	.5249418	7.405215	13.63817	15.70856
combined	402	14.72637	.363687	7.291901	14.0114	15.44134
diff		.1049583	.7282997		-1.326815	1.536732

```
    diff = mean(Governme) - mean(Private)                         t =   0.1441
Ho: diff = 0                                     degrees of freedom =      400

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.5573         Pr(|T| > |t|) = 0.8855          Pr(T > t) = 0.4427
```

Table 27: t- test by data requesting organization

```
. ttest Score, by( HRLB )

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    LRHB |     202    16.19802    .4380795    6.22628    15.3342    17.06184
    HRLB |     200       13.24     .563783    7.973095   12.12824    14.35176
---------+--------------------------------------------------------------------
combined |     402    14.72637     .363687    7.291901    14.0114    15.44134
---------+--------------------------------------------------------------------
    diff |              2.95802    .7131157                1.556097    4.359943
------------------------------------------------------------------------------
    diff = mean(LRHB) - mean(HRLB)                                t =   4.1480
Ho: diff = 0                                     degrees of freedom =      400

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Table 28: t- test by data net privacy calculus

```
. reg Score Priv HRLB

      Source |       SS           df       MS      Number of obs   =       402
-------------+----------------------------------   F(2, 399)       =      8.60
       Model |  881.160462         2  440.580231   Prob > F        =    0.0002
    Residual |   20440.74        399  51.2299249   R-squared       =    0.0413
-------------+----------------------------------   Adj R-squared   =    0.0365
       Total |  21321.9005       401  53.1718217   Root MSE        =    7.1575

------------------------------------------------------------------------------
       Score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Priv |  -.1345541   .7140392    -0.19   0.851    -1.538303    1.269195
        HRLB |  -2.959365   .7140127    -4.14   0.000    -4.363062   -1.555668
       _cons |    16.2653   .6173139    26.35   0.000     15.0517    17.47889
------------------------------------------------------------------------------
```

Table 29: Regression analysis – Model 1

```
. reg Score Priv HRLB PC_avg

      Source |       SS           df       MS      Number of obs   =       402
-------------+----------------------------------   F(3, 398)       =     33.78
       Model |  4326.87789         3  1442.29263   Prob > F        =    0.0000
    Residual |  16995.0226       398  42.7010618   R-squared       =    0.2029
-------------+----------------------------------   Adj R-squared   =    0.1969
       Total |  21321.9005       401  53.1718217   Root MSE        =    6.5346

------------------------------------------------------------------------------
       Score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Priv |   .4302601    .654923     0.66   0.512    -.8572807    1.717801
        HRLB |  -1.609156    .668978    -2.41   0.017    -2.924328   -.2939835
      PC_avg |  -1.706843   .1900086    -8.98   0.000    -2.080389   -1.333297
       _cons |   22.21245   .8694476    25.55   0.000     20.50316    23.92173
------------------------------------------------------------------------------
```

Table 30: Regression analysis – Model 2

```
. reg Score Priv HRLB PC_avg Cond

      Source |       SS           df       MS      Number of obs   =       402
-------------+----------------------------------   F(4, 397)       =     25.78
       Model |  4396.84001         4  1099.21      Prob > F        =    0.0000
    Residual |  16925.0605       397  42.6323942   R-squared       =    0.2062
-------------+----------------------------------   Adj R-squared   =    0.1982
       Total |  21321.9005       401  53.1718217   Root MSE        =    6.5293

------------------------------------------------------------------------------
       Score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Priv |   -1.52833    1.66307    -0.92   0.359    -4.797854    1.741194
        HRLB |  -3.020777   1.288828    -2.34   0.020    -5.554558   -.4869971
      PC_avg |   -1.69365   .1901348    -8.91   0.000    -2.067447   -1.319854
        Cond |   .5571142   .4348932     1.28   0.201    -.2978673    1.412096
       _cons |  -15.57585     29.511    -0.53   0.598    -73.59321    42.44152
------------------------------------------------------------------------------
```

Table 31: Regression analysis – Model 3

Appendix 3: Pairwise correlation matrix

```
. pwcorr Score Priv HRLB Prim PC_avg age gend edu empl, star(0.05) sig
```

	Score	Priv	HRLB	Prim	PC_avg	age	gend
Score	1.0000						
Priv	0.0529 0.3885	1.0000					
HRLB	-0.2671* 0.0000	-0.0136 0.8248	1.0000				
Prim	-0.0241 0.6945	-0.0100 0.8703	-0.0099 0.8724	1.0000			
PC_avg	-0.4399* 0.0000	0.0195 0.7509	0.2724* 0.0000	0.1096 0.0732	1.0000		
age	0.0320 0.6018	0.0699 0.2541	0.0452 0.4617	-0.0522 0.3950	0.0348 0.5708	1.0000	
gend	-0.0972 0.1126	-0.0567 0.3555	-0.0938 0.1255	0.1791* 0.0033	0.0135 0.8254	-0.1440* 0.0184	1.0000
edu	0.0604 0.3246	0.1098 0.0727	-0.0169 0.7834	0.0063 0.9189	0.0568 0.3543	0.2259* 0.0002	-0.0879 0.1515
empl	0.0464 0.4498	-0.0780 0.2029	0.0103 0.8665	-0.0101 0.8695	0.0050 0.9350	0.1510* 0.0133	-0.0085 0.8897

Table 32: Pairwise correlation matrix

Appendix 4. Score for donation questions

Total estimation Number of obs = 267

	Total	Std. Err.	[95% Conf.	Interval]
height	213	6.575759	200.0528	225.9472
weight	199	7.132474	184.9567	213.0433
blood_group	102	7.954287	86.33863	117.6614
sleep	221	6.182074	208.828	233.172
diet	195	7.265119	180.6955	209.3045
allergies	186	7.525895	171.1821	200.8179
vaccination	218	6.337026	205.5229	230.4771
medications	168	7.907358	152.431	183.569
surgery	197	7.200146	182.8235	211.1765
alcohol	195	7.265119	180.6955	209.3045
cigarette	213	6.575759	200.0528	225.9472
marijuana	201	7.062024	187.0954	214.9046
drugs	181	7.649758	165.9382	196.0618
exercise	195	7.265119	180.6955	209.3045
social_media	180	7.672821	164.8928	195.1072
acute_disease	201	7.062024	187.0954	214.9046
chronic_disease	186	7.525895	171.1821	200.8179
family_disease	151	8.114778	135.0226	166.9774
mental_health	178	7.717279	162.8053	193.1947
chekups	175	7.779866	159.682	190.318
hlth_insurance	211	6.664912	197.8773	224.1227

Table 33: Score for each donation questions