THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF CYBERSECURITY ANALYTICS AND OPERATIONS

Communicating Privacy-Utility Trade-Offs of Differential Privacy Through Analogical Reasoning

EPHRAIM N.T.A. GOVERE
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Cybersecurity Analytics and Operations
with honors in Cybersecurity Analytics and Operations

Reviewed and approved* by the following:

Aiping Xiong
Assistant Professor of Cybersecurity Analytics and Operations
Thesis Supervisor

Anna C Squicciarini
Professor of Cybersecurity Analytics and Operations
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Differential privacy (DP) techniques have been proposed and applied to protect personal data. However, it remains unclear whether users will understand DP. One approach to address this problem is by effectively communicating the privacy-utility trade-offs of DP through analogical reasoning. This project assessed the effectiveness of an analogical reasoning approach, compared to a text description, in increasing understanding of privacy-utility trade-offs of DP and the willingness of users to share their private information. We conducted an online survey with college students (N=23) using Qualtrics. A subsample of the online participants was invited for a follow-up interview through Zoom. There were no significant differences between the conditions for text-only and analogical reasoning measures. The feedback on the survey instrument experience highlighted the need to minimize the length of the text and technical jargon in the survey. Results of our study show no significant differences between the text-only condition and analogical reasoning condition. Differences between our results and several other results reveal the need for continued research in the design of textual and visual approaches to effectively and accurately communicate differential privacy to increase understanding and willingness by users to share private information.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Aiping Xiong, Assistant Professor in the College of Information Sciences and Technology for giving me the opportunity to gain skills in conducting experiments related to cybersecurity and privacy through NSF award # 1931441 Research Experiences for Undergraduates (REU) program. I gratefully acknowledge the assistance and guidance from Dr. Aiping Xiong, my undergraduate thesis supervisor, and Professor Anna C Squicciarini, my Penn State's Schreyer Honors College advisor, Professor-in-Charge of Cybersecurity Analytics and Operations, College of Information Sciences and Technology.

Chapter 1

Introduction

Data leakage can occur when information is gathered, stored, and disseminated. Protecting personally identifiable information (PII) is one of the biggest challenges in today's cyber-driven society. The U.S. Department of Labor (n.d.) defined PII as information that is capable of indirectly or directly identifying an individual to whom the information applies. Some examples of information that is capable of directly identifying an individual include: name, address, email address, or social security number.

Under the United States (US) Department of Health and Human Services (HHS) Protection of Human Subjects Regulations Title 45 CFR, individually identifiable information means the information that can make it possible to ascertain the identity of, or associate with, an individual. To protect personal health information (PHI) from inappropriate use or disclosure, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule under Title 45 CFR, requires de-identifying the following 18 protected PHI elements that could be used to identify the individual or the individual's relatives, employers, or household members (United States National Institute of Health [NIH], 2022). Some of the protected PHI elements are: names, telephone numbers, social security numbers, medical record numbers, biometric identifiers, including fingerprints and voiceprints, and full-face photographic images and any comparable images.

The European Union established a General Data Protection Regulation (GDPR) (European Parliament and of the Council, 2016). The law defines PII as personal data, meaning

> "*...any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."* (GDPR 2016/679, Article 4).

There is a great need for techniques that anonymize an individual's PII in databases and protect them from having their PII stolen or leaked. One such privacy-preserving technique is Differential Privacy (DP). In the context of data release, differential privacy protects individuals' data privacy through adding noise to aggregated results such that the difference of whether or not an individual is included in the data is bounded (Dwork et al., 2006). While academic and industry are focusing on building tools for differential privacy systems, whether users understand the techniques, trust the techniques, and consequently increase their data participation, are unclear. In context of differential privacy, privacy-utility tradeoff would include the perceived privacy protection cost or risk of revealing data versus the cost of utility (Valdez & Ziefle, 2019). Proper communication of the privacy protection and utility cost of differential privacy will be critical to its adoption and implementation. Communication utilizing analogical reasoning and visualization could provide a better understanding of role of differential privacy on privacy protection and utility cost.

Chapter 2

Related Work

Differential Privacy

Differential privacy uses complex mathematical privacy-enhancing techniques to preserve the anonymity of data. It uses statistical data analysis to calculate and quantify privacy losses in large databases. Differential privacy adds noise to aggregated results in such a way that the difference of whether or not an individual is included in the data is bounded (Dwork et al., 2006). There are several variants of differential privacy, including pure differential privacy (Dwork et al., 2006), approximate differential privacy (Dwork et a., 2006), concentrated differential privacy (Dwork & Rothblum, 2016), and zero-concentrated differential privacy (zCDP) (Bun & Steinke, 2016).

Pure differential privacy ($(\varepsilon, 0)$-DP) anonymizes an individual's data when $\delta = 0$ and, approximate differential privacy ((ε, δ)-DP) is achieved when $\delta > 0$. Therefore, the definition of differential privacy uses mathematical theorems, randomized algorithms, and statistical queries to achieve data anonymity. Differential privacy techniques efficiently measure privacy loss which is the information revealed in its computations. In $(\varepsilon, 0)$-DP, parameter $\varepsilon > 0$ controls "privacy loss"; it is the privacy budget parameter. In (ε, δ)-DP, δ limits the probability of privacy loss exceeding ε (Bun & Steinke, 2016, p. 736). (ε, δ)-DP guarantees that privacy loss does not exceed ε due to the probability at most being $1-\delta$ (Dwork & Rothblum, March 2016, p.1). Similar to differential privacy, privacy loss uses complex mathematical concepts and techniques that may be too difficult to understand for non-technical users. In simpler terms, privacy loss is defined as "a random variable which quantifies how much information is revealed about an

individual by a computation involving their data." (Bun & Steinke, 2016, p. 635). After analyzing (μ, τ)-CDP, Bun & Steinke (2016) came up with their variant called zero-concentrated differential privacy (zCDP), an intermediate notion between (ε, 0)-DP and (ε, δ)-DP. It provides guarantees of (ε, δ)-differential privacy for all values of $\delta > 0$. Bun and Steinke also branded the (μ, τ)-CDP as mean-concentrated differential privacy (mCDP).

The differential privacy variants discussed above are not exclusive. Other types include personalized differential privacy (Jorgensen, Yu, and Cormode, 2015), Bayesian differential privacy (Yang, Sato, and Nakagaw, 2015), and Robust Local Differential Privacy (RLDP) (Lopuhaä-Zwakenberg, & Goseling, 2021). However, they all demonstrate that the probability distribution on the published results remains the same, "*independent of an individual's willingness to opt in or out of the dataset*" (Dwork, Naor, Reingold, & Rothblum, 2015, p. 735). The purpose of differential privacy is to add noise to aggregated results to limit the probability of PII identification (Dwork et al., 2006). If differential privacy is understood and made clear to the individual, they should have little fear of providing their PII where differential privacy is implemented. The second common element within differential privacy variants is their advanced mathematical and statistical computations saturated with theorems and randomized algorithms. According to Cuff and Yu (2016, p. 1), differential privacy "directly addresses the statistical distinguishability of the database and has led to algorithms for answering general queries with just the right amount of randomness used in order to preserve privacy". It is not clear whether end-users understand and therefore trust differential privacy techniques. Although differential privacy provides a successful method to protecting PII, there is difficulty gauging whether end-users trust the system enough to share their PII data. Formal definitions of differential privacy are difficult for non-technical users to comprehend and later trust the privacy protection

technique. Differences in differential privacy variants further exacerbate the problem of non-technical user comprehension. Wood et al. (2018) compiled a detailed document that aimed to explain differential privacy to non-technical users. Still, they could not avoid the mathematical equations associated with differential privacy. How can one communicate the complexity of differential privacy to non-technical users? The answer may lie in using analogical reasoning to communicate the privacy-utility trade-offs of differential privacy.

Analogical Reasoning

Analogical reasoning is a type of reasoning or thinking that relies upon an analogy (Bartha, 2019). An analogy compares similarities, relations, or parallels across domains such as concepts, objects, elements, organs, systems, functions, situations, or phenomena. A strong analogy has many similarities across domains. The more differences or ignorance about the domains, the weaker the analogy (Bartha, 2019). The basic phenomena of an analogy include relational similarity, structural consistency, systematicity, inferences, alignable differences, interactive mapping, multiple interpretations, and cross-mapping (Gentner, 2003, p. 110).

Analogical reasoning is often used in everyday problem solving and uses analogies as an inductive tool to compare structured mental representations of two or more situations. Analogies can be represented visually to help analogically map – recognizing commonalities between source and target analogs – between situations (Holyoak, 2012). Analogical reasoning provides the ability for an individual to understand two or more supposedly unrelated concepts by mapping commonalities between them.

The analogical reasoning process entails analogical mapping, structural alignment, analogical inference, and evaluation (Gentner & Smith, 2013). Analogical reasoning uses examples of familiar concepts, objects, elements, organs, systems, functions, situations, or phenomena to enhance new learning, understanding, and application of unfamiliar ones. It enhances learning through analogical abstraction, inference projection, difference detection, and re-representation (Gentner & Maravilla, 2017). The ability to perceive like relational structure across different domains is a fundamental process to learning and generating new abstractions (Gentner & Hoyos, 2017) and is a core mechanism of human cognition (Gentner & Maravilla, 2017). The analogical reasoning approach has many advantages (Gentner & Smith, 2013; Sunstein,1993) that apply to communicating privacy-utility trade-offs of differential privacy. They include:

1. Facilitating the conceptualization of privacy-utility trade-offs of differential privacy.
2. Helping to inform end users' judgments about the privacy-utility trade-offs of differential privacy.
3. Enabling users to perceive the relational structure of privacy-utility trade-offs of differential privacy across different contexts.
4. Helping in explaining and understanding complex concepts and problems related to privacy-utility trade-offs of differential privacy.
5. Allowing end-users to understand complex differential privacy abstraction by reference to examples and enable them to agree to share private information.

Using analogical reasoning approach to enable users understand the privacy-utility trade-offs of differential privacy may increase their trust in the privacy method and increase their willingness to provide their PII for future research to solve problems. By providing a source analog, users

could potentially create mental commonalities between the source analog and the target analog (differential privacy). These commonalities will communicate the privacy and utility trade-offs that make DP a carefully constructed method to protect users' private information when providing it to a recipient.

Privacy-Utility Trade-offs

All privacy-preserving techniques have instances of privacy-utility tradeoffs. Privacy increase guarantees will consequently increase the loss of information and lower the accuracy of results and analysis utility. Within the differential privacy model, a substantial perturbation of data occurs when ε is smaller, reducing the accuracy of data analysis (Zhang-Kennedy et al., 2021). Privacy-Utility tradeoffs have been measured using visualization and the implementation of Laplacian noise – a fundamental component of noise application in DP. Some visualization products may suffer minimal information loss when injecting Laplacian noise, depending on the level of queries (Lee, 2017). Understanding the mechanical capabilities of different visual products and their ability to maintain privacy-utility may provide insight into the most efficient visual to portray privacy-utility tradeoffs.

Visualization

Visualization is the presentation or transformation of data or text information to a visual form or representation. Why should we use visualization in explaining a complex concept like differential privacy, privacy loss, and privacy-utility tradeoffs? It is because visualization translates or unveils abstract information or data into a visual form to gain a deeper

understanding and foster insight to amplify cognition and communication in solving problems (Chen et al., 2014). According to Zhang, Sarvghad, and Miklau (2020 p. 1788), “The goal of a privacy-preserving visualization is to protect individuals’ identities and sensitive information from exposure while still allowing users to make sense and gain knowledge from it.” A visual analogy is an imagistic, similarity-based reasoning strategy (Goldschmidt, 2001). Visualization types that have been investigated in privacy-preserving visualization experiments include line chart, bar chart, scatterplot, and pie (Saket, Endert, and Demiralp, 2018). Lee (2017) examined five types of visualization products: bar graphs, pie charts, heatmaps, linear plots, and scatterplots in their study on visualization and differential privacy. Nanayakkara et al. (2022) used quantile dotplot to visualize the un-noised and privacy-preserving displays. Their study provided evidence that visualizations helped users understand differential privacy.

Literature Review

We used the Population, Intervention, Compare, Outcome, Time (PICOT) approach (Riva et al., 2012) to formulate a clear, logical, and well-defined research question - "In the last ten years, how effective was analogical reasoning, compared to text description, in communicating privacy-utility trade-offs of differential privacy to users." The logic grid aligned with the PICOT Elements of our review question is shown in Table 2.

Table 1: The PICOT components of the review question.

PICOT element	Review question (basic terms)
Population	Users of information technology
Intervention	Communicating approaches for privacy-utility trade offs

Compare	Comparing analogical reasoning to non-analogical approaches
Outcome(s)	Effective communication of privacy-utility trade-offs
Time	Review of Publication Between 2010 - 2021

The following key terms were used to search abstracts for relevant studies: Differential Privacy AND Analogical Reasoning; Differential Privacy AND Privacy-Utility; Analogical Reasoning AND Privacy-Utility; Differential Privacy AND Analogical Reasoning AND Privacy-Utility. Literature databases. We searched in ten cybersecurity related databases and nine selected cybersecurity journals. (Table 2). For example, the predetermined Boolean/phrases in CiteSeerX (https://citeseerx.ist.psu.edu/index), and Web of Science (https://www.webofknowledge.com) databases that were accessed through Pennsylvania State University database licenses were: abstract:("Differential Privacy" AND "Analogical Reasoning" AND "Healthcare"), and ((AB=(Differential Privacy)) AND AB=(Analogical Reasoning)) AND AB=(Healthcare)), respectively. We used Zotero bibliographic software (Corporation for Digital Scholarship, 2021) for collecting, identifying duplicates, managing, and citing the literature search references.

Table 2: Selected Databases and cybersecurity journals for searching key words and their combinations.

	Selected Databases	URL
1	Academic Search Complete (EBSCOhost)	https://www.ebsco.com/products/research-databases/academic-search-complete
2	ACM Digital Library	https://dl.acm.org/
3	CiteSeerX	https://citeseerx.ist.psu.edu/index
4	DBLP Computer Science Bibliography	https://dblp.org/
5	Google Scholar	https://scholar.google.com/
6	Homeland Security Digital Library	https://www.hsdl.org/c/
7	IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp
8	Microsoft Academic	https://academic.microsoft.com/home
9	ScienceDirect	https://www.sciencedirect.com/
10	Web of Science	https://www.webofknowledge.com

	Selected Cybersecurity Journals	**URL**
1	Information Security Journal: A Global Perspective	https://www.tandfonline.com/toc/uiss20/current
2	International Journal of Cyber-Security and Digital Forensics	https://www.jdfsl.org/
3	Journal of Cyber Security Technology	https://www.tandfonline.com/toc/tsec20/current
4	Journal of Cybersecurity	https://academic.oup.com/cybersecurity/pages/About
5	Journal of Information Security and Applications	https://www.journals.elsevier.com/journal-of-information-security-and-applications
6	Journal of Security and Privacy	https://onlinelibrary.wiley.com/search/advanced?publication=24756725&text1=
7	Journal of Information Privacy and Security	https://link.springer.com/search?query=&search-within=Journal&facet-journal-id=10207
8	International Journal of Information and Network Security (IJINS)	http://ijins.iaescore.com/index.php/IJINS/search
9	Journal of Information Security	https://www.springer.com/journal/10207

We screened the studies based on:

Inclusion criteria

1. Published between January 2010 to December 1, 2021.
2. Peer-reviewed research articles published in scientific journals.

3. Full text articles published in English language.
4. Studies that focused on cybersecurity and differential privacy, or cybersecurity and analogical reasoning, or differential privacy and analogical reasoning.

Exclusion criteria

1. Studies not written in English.
2. Studies unrelated to the research question.
3. Studies published before 2010.
4. Unavailable full texts
5. Unpublished studies.

Search Outcomes

The literature search outcomes are shown in the PRISMA2020 R flow diagram (Page et al., 2021) decision tree (Figure 1). The search revealed a lack of reported studies on the analogical reason as it related to differential privacy and or privacy-utility trade-off. Results also reveal the scarcity studies combining differential privacy and privacy-utility trade-off. Only 44 studies (about 2 per year) were reported in the last 20 years under the inclusion and exclusion literature search criteria for the search strategy.

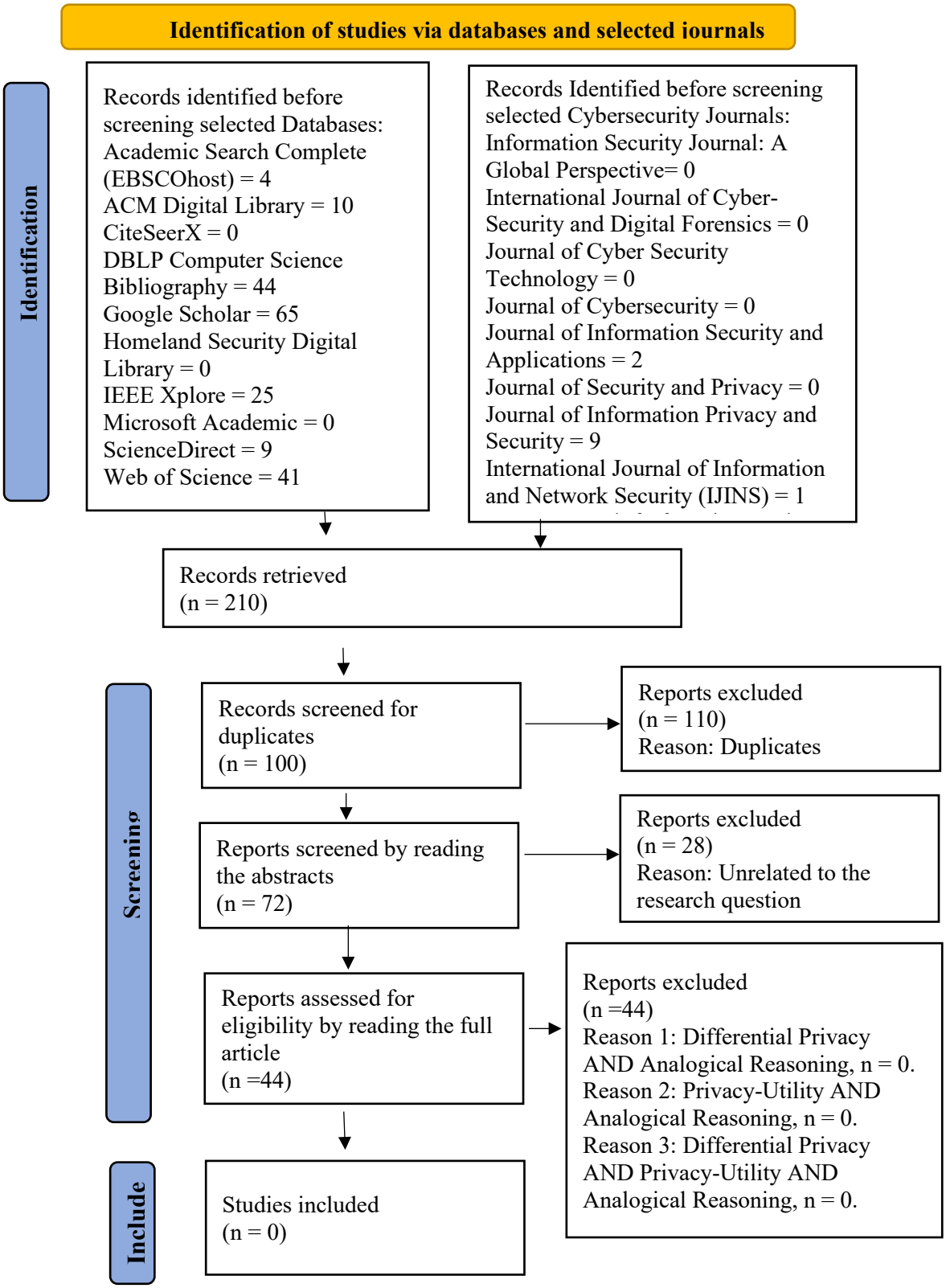


Figure 1: A PRISMA flow diagram of the systematic search, inclusion, and exclusion of retrieved studies.

Chapter 3

Research Objectives

The purpose of this study was to find effective approaches to communicating privacy-utility trade-offs of differential privacy to internet users. Having users understand the privacy protection and utility cost of differential privacy will increase their trust in the privacy method and increase their willingness to provide their private information for future research. We focused on assessing the effectiveness of analogical reasoning in helping users understand the privacy protection and utility cost of differential privacy. Specifically, this study addressed the question: How effective is analogical reasoning, compared to text description, in communicating privacy-utility trade-offs of differential privacy to users? Is the analogical reasoning approach more effective than a text description in increasing the willingness of users to share private information? We hypothesized that having users understand the privacy-utility trade-offs of differential privacy would increase their trust in the privacy method and increase their willingness to provide their private information for future research. The more people participate, the more information gathered, and the more reliable the research outcomes are to make evidence-based decisions and practices.

Chapter 4

Rationale for Hypothesis

Analogical reasoning is often used in everyday problem solving and uses analogies as an inductive tool to compare structured mental representations of two or more situations. Analogies can be represented visually to help analogically map – recognizing commonalities between source and target analogs - between situations (Holyoak, 2012). Analogical reasoning provides the ability for an individual to understand two or more supposedly unrelated concepts by mapping commonalities between them. This method of reasoning could prove effective at helping users understand the trade-offs between privacy and the utility of differential privacy. By providing a source analog, users could potentially create mental commonalities between the source analog and the target analog (differential privacy). These commonalities will communicate the privacy and utility trade-offs that make differential privacy a carefully constructed method to protect users' private information when providing it to a recipient. Based on this rationale, the project addresses the following specific hypotheses:

Hypothesis one: Analogical reasoning as an effective approach to communication differential privacy.

- Null hypothesis (H_0): There are no differences in understanding differential privacy between participants exposed to analogical- and non-analogical-reasoning method of differential privacy communication.
- Alternative hypothesis (H_1): There are differences in understanding differential privacy between participants exposed to analogical and non-analogical-reasoning method of differential privacy communication.

Hypothesis two: Analogical reasoning as an effective approach to communication differential privacy.

- Null hypothesis (H_0): There are no differences in willingness to share PII between participants exposed to analogical- and non-analogical-reasoning method of differential privacy communication.
- Alternative hypothesis (H_1): There are differences in willingness to share PII between participants exposed to analogical- and non-analogical-reasoning method of differential privacy communication.

Chapter 5

Methodology

Study Design

We conducted an online survey. This survey was prepared using the Qualtrics survey tool licensed by The Pennsylvania State University. Qualtrics is an online tool used for creating and publishing surveys and analyzing feedback from participants. A subsample of the survey participants was invited for a follow-up online interview through Zoom.

Participants and Stimuli

We recruited 23 participants with a minimum age requirement of 18 years old. Each participant had the possibility of winning one of six Amazon gift cards ($25) for completing the survey. Participants received either the text-only condition or the analogical reasoning condition. The text-only condition consisted of four paragraphs about a specific concept regarding DP: privacy protection, noise, privacy-utility tradeoffs, and risks associated with DP. The analogical reasoning condition consisted of four images. Each image portrayed privacy protection, noise implementation, privacy-utility tradeoffs, and DP-associated risks. After the DP conditions, participants received five comprehension questions to gauge their understanding of DP. Each question presented a topic: privacy understanding, utility understanding, and privacy-utility tradeoffs understanding. The last six questions of the survey involved data sharing questions. Participants were asked questions about their willingness to share private information. Each

question covered one of the following categories of information: health, social media, salary, and monetary spending (Appendix B).

We recruited six participants who completed the survey to be a part of the follow-up interview. The goal of the interview study was to obtain feedback regarding the existing survey design. Each participant was asked to answer eight interview questions, which addressed participants' general opinion on the survey and their understanding of the descriptions, conditions, and questions (Appendix C).

Procedure

The research proposal and protocols (HRP-591 - Protocol for Human Subject Research) were reviewed and approved by the Institutional Review Board (IRB) in the office of the Human Research Protection Program (HRPP) at Penn State. The eligibility screening process started before obtaining the informed consent. The invitation letter (see Appendix A) clearly stated that only students who are at least 18 years old, at Penn State undergraduate student, within the College of Information Sciences and technology, enrolled in the Spring 2022 semester, and fluent in English are eligible to participate. In addition, the first part of the survey contained the eligibility criteria to complete the survey, and students who did not meet the criteria were instructed not to proceed.

After informed consent, participants were randomly assigned to one of two conditions. One condition had the analogical reasoning approach that use visualization to communicating differential privacy. The other condition had the written text descriptions of differential privacy (control) (see Appendix B). The procedure of the two conditions was the same. At the beginning

of the survey, participants saw a description of the purpose of this study. After answering demographic questions, participants who were eligible for our study proceeded. Participants in the control group viewed the written descriptions. Participants in the other condition saw a visualization in the form of a picture that used an already constructed source analog to help participants understand the target analog, differential privacy. Participants in both conditions then answered a few comprehension questions about differential privacy, evaluating the effectiveness of the communication. Then, participants made data-sharing decisions across six different scenarios (two about health information, two about social media information, and two about personal finance information) in both conditions (Valdez & Ziefle, 2019). The survey was designed to take less than half an hour to complete.

When the participants finished the survey, a thank you message popped up indicating the completion of the survey. The thank you message contained a link to a separate survey where students entered their PSU email address that was used to distribute the gift cards.

Measures

We measured the effectiveness of analogical reasoning to communicate differential privacy bason on a questionnaire filled out by the recruited participants. We evaluated the participants' 1) comprehension of differential privacy and 2) the effect of communication in their data-sharing decisions. We expected that participants in the analogical-reasoning condition would show a better understanding of differential privacy, increase their data-sharing decisions, and reveal more trust in differential privacy.

Data Analysis

Correct answer rate for comprehension questions was determined for each participant and grouped as a function of 2 (condition: text only, analogical reasoning-visualization) for analyses of Chi-squared test. The binary data-share decision was determined for each participant and grouped as a function of 3 (information type: health, social media, personal finance) and 2 (condition: text only, analogical reasoning-visualization) for a chi-squared test.

We conducted null-hypothesis testing ($\alpha = 0.05$) for those measures. The null hypothesis was rejected when the obtained results among the conditions were significantly different from each other. Post-hoc tests with Bonferroni correction were performed, testing pairwise comparisons with corrected p values for possible inflation.

The statistical significance of collected data between the treatment group and control group were analyzed using analysis of variance (ANOVA) and chi-squared test by using Microsoft Excel.

Chapter 6

Results

We summarize the survey responses in Table 3. Chi-square Test analysis of the comprehension question responses in Table 3 between text-only condition and the analogical reasoning comprehension questions revealed no significant difference [X^2 (1, N = 23) = 0.65, p > 0.05]. However, when comparing comprehension between the text-only condition and analogical reasoning condition, the text-only responses had a numerically higher correct answer rate (68.06%) than analogical reasoning (54.85%) [Table 3]. The greatest differences among the comprehension questions, between text-only and analogical reasoning, occurred within Q2. Q2 gauged respondents' comprehension of noise application when implementing differential privacy. Respondents given the text-only condition performed better (77.78%) when answering Q2 then the analogical reasoning condition (25.00%) [Table 3]. When asked how DP affects one's privacy within the dataset, the analogical reasoning condition performed better [100.00%] compared to the text-only condition [66.67%].

The Chi-squared test analysis comparing the results [Table 3] of the data sharing questions between the text-only condition and analogical reasoning condition showed no significant differences between the conditions [X^2 (1, N = 23) = 0.74, p > 0.05]. Comparing data sharing between the text-only responses and the analogical reasoning responses, the text-only condition had the largest percentage of respondents agreeing to share their PII (83.34%) compared to the analogical reasoning condition (75.00%) [Table 3]. Among the data sharing questions, Q6 had the greatest difference between the text-only responses (88.89%) and the analogical reasoning responses (57.14%). Participants exposed to the text-only condition exhibited a better understanding of DP and more willingness to share their private information.

Table 3: Percent responses on the comprehension and data sharing.

Comprehension Questions				Data Sharing Questions		
Percentage of Correct Answers				**Percentage of "Yes" Answers**		
Treatment	Text	Analogical			Text	Analogical
Q1	66.67%	100.00%		Q6	88.89%	57.14%
Q2	77.78%	25.00%		Q7	77.78%	78.57%
Q3	66.67%	45.45%		Q8	88.89%	71.43%
Q4	66.67%	58.33%		Q9	88.89%	64.29%
Q5	62.50%	45.45%		Q10	66.67%	85.71%
				Q11	88.89%	92.86%
Participants	9	14			9	14
Mean %	68.06%	54.85%		**Mean %**	83.34%	75.00%

The results of our interview study revealed similar opinions among respondents with regard to the survey. Respondents brought up the opinion of the survey being sensible the most frequently (16 occurrences) [Table 5]. Four out of the eight interview questions involved respondents' opinions of when the survey attempted to explain a concept of differential privacy. Respondents believed the survey explanations were too lengthy (17 occurrences) and had technical jargon (10 occurrences) which made the explanations more difficult to understand [Table 5]. Three out of the six respondents stated that their answers to the data sharing questions would have remained the same even without taking the survey. However, the other three out of six respondents stated that before taking the survey, at least one of their answers would have changed.

Table 4: Respondents' opinions on interview questions

Opinions	Respondent 1	Respondent 2	Respondent 3	Respondent 4	Respondent 5	Respondent 6
Q1: General DP Explanation Opinions	Sensible	Sensible; Lengthy Explanation	Sensible; Lengthy Explanation	Sensible; Lengthy Explanation	Technical Jargon; Lengthy Explanation;	Technical Jargon; Lengthy Explanation
Q2: Description Comprehension Opinions	Sensible	Lengthy Explanation	Lengthy Explanation	Lengthy Explanation	Technical Jargon; Lengthy Explanation;	Technical Jargon; Sensible; Lengthy Explanation
Q3/Q4: Condition Explanation Comprehension Opinions	Sensible	Lengthy Explanation	Sensible	Lengthy Explanation	Technical Jargon; Lengthy Explanation	Technical Jargon; Sensible; Lengthy Explanation
Q5: Comprehension Questions Opinions	Sensible	Sensible	Sensible	Sensible	Technical Jargon	Two Questions were Confusingly Similar
Q6: Combination of Descriptions and Comprehension Questions Opinions	Sensible	Lengthy Explanation	Sensible	Sensible	Technical Jargon; Lengthy Explanation	Technical Jargon; Lengthy Explanation
Q7: Data Sharing Opinions	DP gave more privacy confidence	Would have chosen the same answers regardless of survey	Would have chosen the same answers regardless of survey	DP gave more privacy confidence	Would have chosen the same answers regardless of survey	DP gave more privacy confidence
Q8: Survey Influence on Data Sharing Opinions	Slight Influence	No Influence	No Influence	Slight Influence	No Influence	Slight Influence

Chapter 7

Discussion

Based on our results, respondents were willing to share their private information within the text-only condition (83.34%) and the analogical reasoning condition (75.00%). Bullek et al. (2017) found that some individuals associate obfuscating privacy mechanisms with deception and prefer less data privacy. Respondents may have been more willing to share their private information and preferred low levels of privacy during the data sharing portion of our survey due to equating differential privacy as deceptive. The research found that some users come with predetermined privacy concerns, and their willingness to share information is likely tied to having their privacy concerns addressed (Cummings et al., 2021). Respondents in our survey may have been more willing to share their private information because DP addressed some of their privacy concerns. The non-significant differences between the text-only condition and the analogical reasoning condition could be due to the small sample size.

We conducted an interview study to receive feedback from respondents about their experience when taking our survey. A common theme among all participants was the length of the contextual scenario. At the beginning of our survey, respondents were presented with a scenario to help contextualize the entire survey. Respondents suggested the contextual scenario be more concise and shortened because its length contributed to a loss of focus when proceeding through the rest of our survey. Another common theme among respondents was that the technical jargon within our survey made it more difficult to comprehend concepts. Some respondents stated they are unfamiliar with privacy-preserving technology and found the technical jargon to

hinder their understanding process. McDonnell et al. (2016) found that introducing concepts first and technical vocabulary (jargon) second can increase student learning. If respondents were learning DP for the first time through our survey, their ability to understand DP could have been helped or hindered depending on the learning structure.

However, other research shows that visualization could be more effective and more persuasive than text-only content. Participants exposed to the analogical reasoning reported their willingness to share their PII. On their study on visualizations to teach about mobile online privacy, Mekhail, Zhang-Kennedy, and Chiasson (2014) concluded that infographic conditions enhanced knowledge and usefulness of the information gained. In the study by Kiernan et al (2018), the infographic letter (letter with illustrations) improved knowledge and trust when compared to control letter (text only) (88.7% vs. 66.7%, absolute percentage difference 22.0%, $p < .0001$). The study reported significantly higher transparency, perceived value for retention, and trust (Cohen's ds = 0.4–1.0, ps < .0001) among infographic participants.

Beveridge and Parkins (1987) found that analogues can be effective if they represent the appropriate features of source-target analog relationship. Hegarty listed several principles that could guide the visual display. They included principles related to task specificity, expressiveness, perception, semantics, and usability of displays, and the principle of visual momentum (Hegarty, 2011). Our visualization display fulfilled some of these principles. For example, semantic principle of compatibility by making the visual display (a person) compatible with its meaning "individual" whose information is added to a large pool of individuals. This could have enhanced the effectiveness of the analogical reasoning approach compared to the text-only.

Visualization improves cognition. According to Hegarty (2011) advantages of displays include external storage of information. organization of information, and the offloading of cognition on perception. The analogical mapping and the integration of information take place in the frontopolar cortex of the brain which is responsible for mapping the structural similarities between the base analog and the target analog (Green, 2012). The portion of the participants whose outcome did not show the effectiveness of analogical reason when exposed to the analogical condition may be explained by inherent differences in the frontopolar cortex among the participants. The degree of activation in regions of frontopolar cortex by the visualizing the displays may not be uniform across the participants. The stronger the activation in regions of frontopolar cortex, the higher decision-making efficiency (Laureiro-Martínez et al., 2014). Although visualization could improve one's ability to understand differential privacy, its features may distract users from accurately and honestly answering questions (Couper et al., 2001).

Limitations and Future Directions

Even though visualization could be effective, there is room to compare our static visualization with dynamic visualizations. Differences between our results and several other results reveal the need for continued research in the design of textual and visual approaches to effectively and accurately communicate differential privacy to increase understanding and willingness by users to share private information. Our future work will involve testing text-only approach versus multiple different visualizations. The work will involve a larger sample size. We intend to recruit undergraduate students enrolled in the College of Information Sciences and

Technology (sample size = 1665). The participants will be recruited from the following six College of Information Sciences and Technology programs at the PSU University Park campus:

1. Bachelor of Science in Cybersecurity Analytics and Operations
2. Bachelor of Science in Data Sciences
3. Bachelor of Science in Enterprise Technology Integration
4. Bachelor of Science in Human-Centered Design and Development
5. Bachelor of Science in Information Sciences and Technology
6. Bachelor of Science in Security and Risk Analysis

We expected a minimum number of about 330 students to complete the research procedures. We calculated the sample size for this future research using Taro Yamane (Yamane, 1973) formula at 95% confidence level.

Yamane's formula:

n = N/(1-(N(e^2))

Where:

n= sample size required

N = number of participants in the population (1665 = all undergraduate students)

e = allowable error (%) or precision error (= 5% to give our sampling a 95% confidence level that the sample is a good representation of the population).

Based on this formula the sample size (n) should be 323 students. However, we will recruit all the undergraduate students in the College of Information Sciences and Technology in order to obtain at least 323 participants. The more the participants the more reliable the results become.

Chapter 8

Conclusion

In our study, the proposed analogical reasoning approach, compared to the text description, showed no statistical difference in helping users understand the privacy-utility trade-offs of differential privacy and increase their willingness to provide share private information. Based on the interview study, we conjecture the ineffectiveness of the proposed analogical reasoning approach is due to lengthy explanations and, and technical jargon which made the explanations more difficult to understand. We plan to improve the current analogical reasoning approach and modify the textual descriptions and visualization. We also plan examine the effectiveness of updated approaches in a formal study with a large sample size.

Appendix A

Invitation Letter

Research Participants Needed

Researchers at the College of Information Sciences and Technology are conducting a research study to investigate the effective ways to communicate privacy enhancing technologies. Research participants are needed in this study. You will complete an online survey, in which we will explain a privacy enhancing technology at first, then you will be asked to answer a few questions about it. The survey will take approximately 15 minutes. As an incentive to participate, we will randomly select participants who complete the study and each of them will receive a $25 Amazon gift card. When the survey is submitted a thank you message will popup indicating you completed the survey. The thank you message will contain a link to a separate survey where you will provide your Penn State email address and your current or expected degree. An email will be sent to the winners of the gift cards containing the gift card codes.

Guidelines will be followed to ensure that your privacy will be protected.

You are eligible to participate in this study if you are:

- At least 18 years old
- A Penn State undergraduate student
- Within the College of Information Sciences and technology
- Enrolled in the Spring 2022 semester
- Fluent in English

Please click this link to begin the survey:

https://pennstate.qualtrics.com/jfe/form/SV_3qMEgT0oCDrKfVI

Thank you for your interests in this study. Please reply to email exg5084@psu.edu.

Any questions, please feel free to contact us.

Email: exg5084@psu.edu

Appendix B

Online Survey

Survey Criteria

Please choose the answers that apply to you. If the following answer does not apply to you, please do not proceed with this survey:

- ☐ I am at least 18 years old

Demographic Questions

What is your gender?

- o Male
- o Female
- o Non-binary / third gender
- o Prefer not to say

How would you best describe yourself?

- o American Indian or Alaska Native
- o Asian
- o Black or African American
- o Native Hawaiian or Other Pacific Islander
- o White
- o Prefer not to say

Are you of Hispanic/Latino/Spanish origin?

- o Yes
- o No
- o Prefer not to say

What year are you as an undergraduate student?

- o Freshmen
- o Sophomore
- o Junior
- o Senior

- Prefer not to say

Which undergraduate program are you studying in?

[Enter Here]

Introduction

The purpose of this study is to conduct research on the most efficient method to communicate differential privacy. You will be presented with a scenario that helps contextualize the concepts in this survey. You will be given an explanation of differential privacy. Questions will be asked to help gauge your understanding of differential privacy and understand your willingness to share information if it is protected by the privacy protection technique.

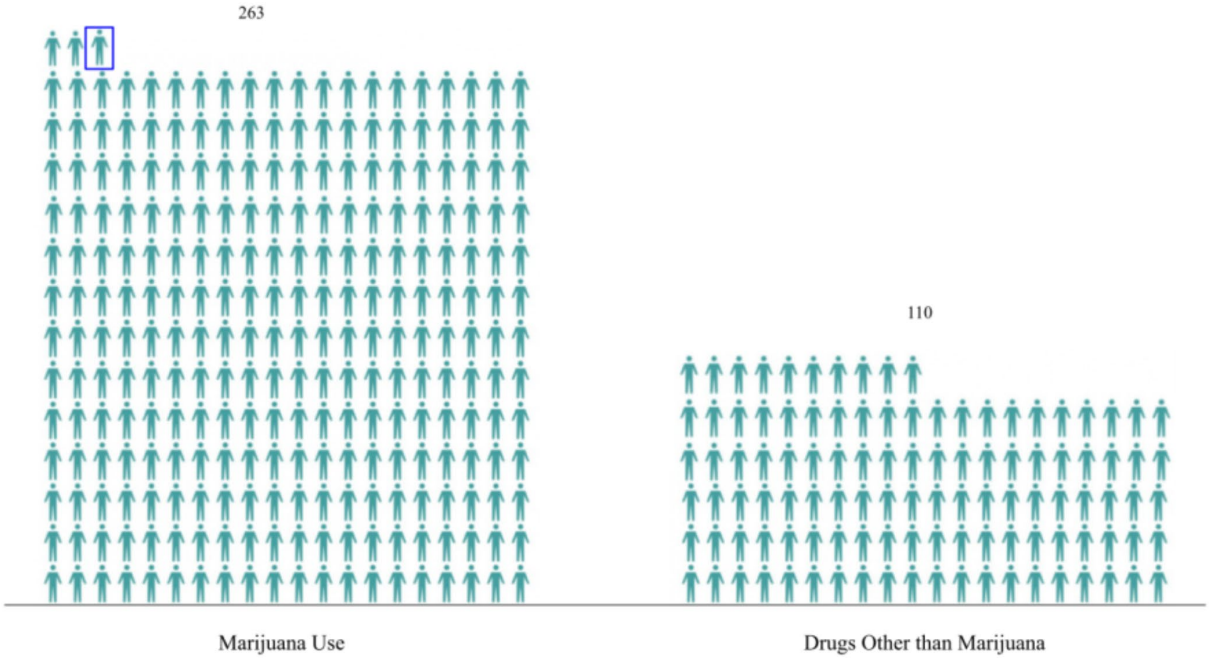
Scenario

Alex and Bill are researchers at the medical research center of State University. Both have access to a database that contains medical outcomes about students at State University, including information related to the drug use of each student. Because it contains protected personal health information, access to the database is restricted. To gain access, Alex and Bill were required to demonstrate that they planned to follow the regulations and protocols from federal, state, and university for handling personal data, by undergoing confidentiality training and signing data use agreements proscribing their use and disclosure of personal information obtained from the database. In October, Alex published an article based on the information in this database and wrote that " the current freshman class at State University is made up of 3,005 students. 373 of them have substance use issues, with 263 responses being marijuana use." Alex reasons that, because the figure in her article is an average taken over 3,005 people, no individual's personal information will be exposed. The following month, Bill published a separate article exploring the relationship between students' enrolled year and medical outcomes at State University. It contains these figures: "the current freshman class at State University is made up of 3,004 students, with 262 marijuana

use.” A student Zoe, read both articles and noticed the discrepancy. From the published information, Zoe concluded that one freshman withdrew from State University between the two months and that the student is a marijuana user. Zoe asked around and determined that a student named John had dropped out in October.

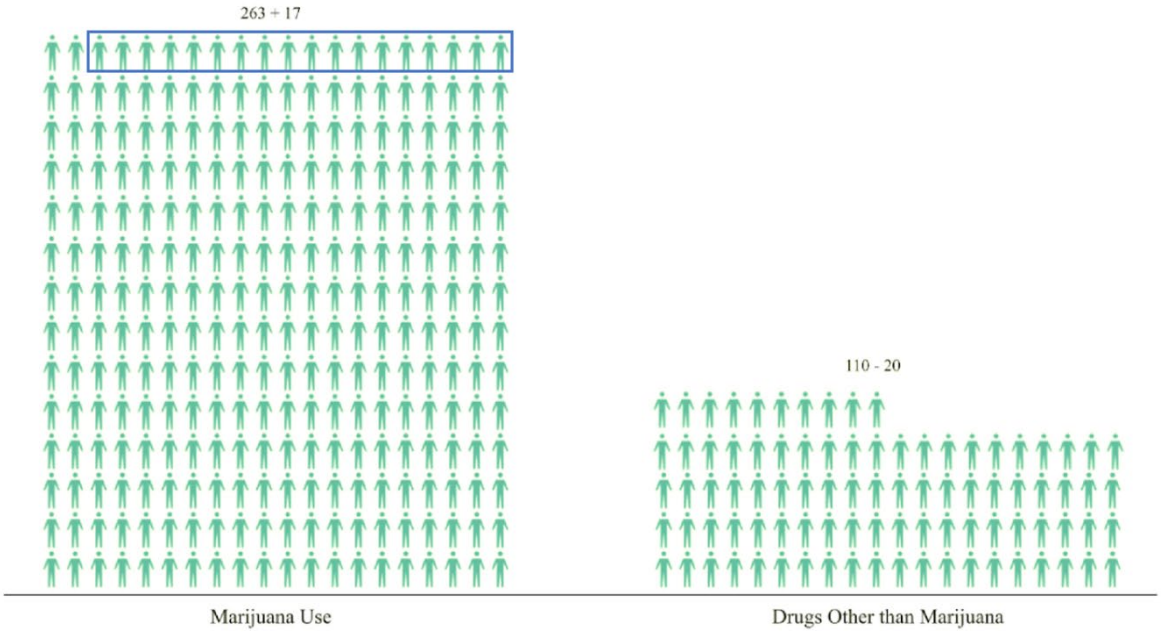
Analogical Reasoning Condition

Thus, in combination, the results of multiple analyses using information about the same people may enable one to draw inferences about individuals in the data, e.g., John’s marijuana use (see image below). Although individually published information that, in isolation, seems innocuous. However, when combined, the information compromised individuals' privacy.

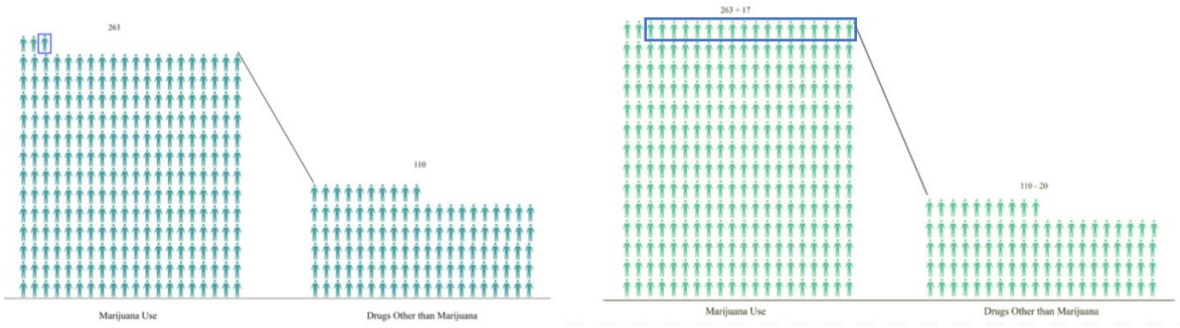


When differential privacy is added, there is a lower probability to identify John’s drug usage answer. There is positive noise and negative noise added to each variable. The added noise included an additional

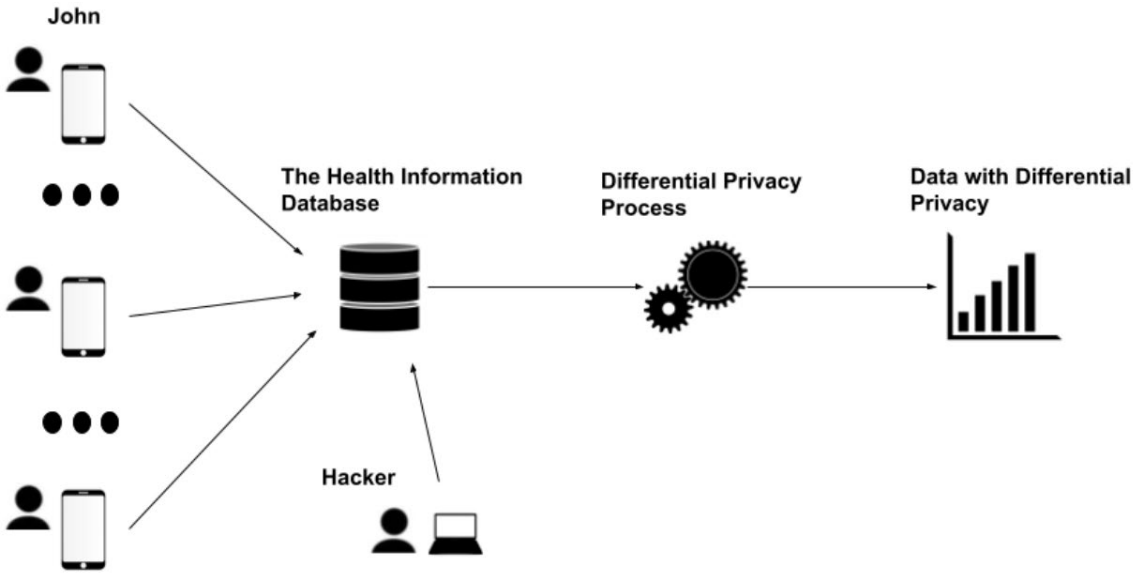
17 people to better mask the answers of those who answered "Marijuana Use". The negative noise removed 20 people, better masking the answers for those who answered "Drugs Other than Marijuana". The change in color within the data indicate the added noise, both positive and negative. Both positive and negative noise reduce the probability of identifying John's drug answers and the answers of others in the data. The change in color within the data indicate the added noise, both positive and negative, within the data.



The added noise alters the data, indicated by the increase and decrease in total individuals present within each variable and a shift in variable height. Having more positive or negative noise can change the accuracy of the data. However, the noise alteration does not obscure the data enough to make it unusable. The utility of the data is maintained when adding a controlled level of positive or negative noise.



Central differential privacy is applied after the health app organization collects the data. Therefore, there is a possibility of data leakage if the organization's database is compromised or hacked by attackers before the differential privacy process.



Text-only Condition

Thus, in combination, the results of multiple analyses using information about the same people may enable one to draw inferences about individuals in the data. Although individually published information in isolation seems innocuous, when combined, the information compromised individuals' privacy. To prevent compromised information, researchers at the medical center in State University will deploy the privacy protecting technique, central differential privacy (DP). DP protects users' privacy by employing random noise to aggregated data to anonymize user information.

When using DP, random noise can either be positive or negative. The researchers at State University found that 263 students had drug issues regarding marijuana usage. When implementing positive noise, the researchers may include an additional 14 dummy users to the marijuana group to help decrease the probability of someone identifying a specific individual. When implementing negative noise, the researchers may remove 10 students from the marijuana category to decrease the probability of identifying a specific individual.

When DP is implemented, the random noise will affect the usability of the data and the amount of privacy protection provided to the user data. Having more positive or negative noise within a dataset will provide increased privacy protection but decrease the utility of the data. Researchers at State University would have less accurate results if they decided to increase either positive or negative noise but, will also be giving users more privacy protection. Having less positive or negative noise will decrease privacy protection but increase the utility of the data. Researchers at State University would have more accurate results with less positive or negative noise but, users would have less privacy protection.

There are possible risks associated with using DP. DP adds noise after gathering all the necessary data. Once the data is gathered into one database, the DP process is applied and anonymizes the entire dataset.

Since the data is gathered into one database, a hacker could intercept the data before DP is applied, compromising the entire dataset.

Comprehension Questions

Q1: Suppose you participated in the medical research at State University. The answers were collected and differential privacy was applied to the dataset. How does differential privacy affect your privacy within the dataset?

- Decreases the likelihood of being identified
- Increases the likelihood of being identified
- Differential privacy does not affect privacy
- Differential privacy makes you completely undetectability within a dataset
- Prefer not to answer

Q2: Suppose you participated in the medical research at State University. You answered truthfully to information relating to student drug use and the answers were collected using differential privacy. If positive or negative noise is added to the dataset by using differential privacy, will the accuracy of the data be changed?

- Yes, positive or negative noise applied by differential privacy will change the accuracy of the data
- No, positive or negative noise applied by differential privacy will not change the accuracy of the data
- Yes, but only positive noise will change the accuracy of the data
- Yes, but only negative noise will change the accuracy of the data
- Prefer not to answer

Q3: Suppose you participated in the medical research at State University. You answered truthfully to information relating to student drug use and the answers were collected using differential privacy. If noise is added to the dataset by using differential privacy, how accurate would the modified results be compared to the true results (without differential privacy)?

- Less accurate
- More accurate

- No change
- Differential privacy does not affect data accuracy
- Prefer not to answer

Q4: Suppose medical research at State University collected information on students drug usage. The research collected information on 263 students. After applying differential privacy, 37 fake users were added to the collected data, making the total 300 students. How has the privacy and utility changed?

- Privacy has Increased, Utility has Decreased
- Privacy has Decreased, Utility has Increased
- Privacy has not changed, Utility has not changed
- Only Privacy has Increased, Utility has not changed
- Prefer not to answer

Q5: Suppose medical research at State University collected information on students drug usage. The research collected information on 263 students. After applying differential privacy, the researchers decided 37 fake students were too many and only added 2 fake students, making the total 265 students. How has the privacy and utility changed compared to having 37 fake students?

- Privacy has Increased, Utility has Decreased
- Privacy has Decreased, Utility has Increased
- Privacy has not changed, Utility has not changed
- Only Privacy has Increased, Utility has not changed
- Prefer not to answer

Data Sharing Questions

Q6: Suppose researchers at State University want to collect information on student's drug use. If the privacy protection technique (differential privacy) is applied, would you be willing to share that information with the research?

- Yes
- No
- Prefer not to answer

Q7: Suppose researchers at State University want to collect information on student's vaccination card to see what percent of the university is vaccinated. If the privacy protection technique (differential privacy) is applied, would you be willing to share that information with the research?

- Yes
- No
- Prefer not to answer

Q8: Suppose researchers at State University want to collect information on student's social media posting habits. If the privacy protection technique (differential privacy) is applied, would you be willing to share that information with the research?

- Yes
- No
- Prefer not to answer

Q9: Suppose researchers at State University want to collect information on student's social media comments to improve their anti-bully and anti-hate speech algorithm. If the privacy protection technique (differential privacy) is applied, would you be willing to share that information with the research?

- Yes
- No
- Prefer not to answer

Q10: Suppose researchers at State University want to collect information on the average salaries of current university students. If the privacy protection technique (differential privacy) is applied, would you be willing to share that information with the research?

- Yes
- No
- Prefer not to answer

Q11: Suppose researchers at State University want to collect information on the average amount of money students spend on textbooks. If the privacy protection technique (differential privacy) is applied, would you be willing to share that information with the research?

- Yes
- No
- Prefer not to answer

Appendix C

Interview

Question 1: What were your general opinions on part of the survey that explained differential privacy? Were there any obvious spelling errors or sentences that did not make sense grammatically, etc.?

Question 2: Were the descriptions easy to understand? If not, what part of the description could use more clarification?

Question 3: (Visual Group): Were the images easy to comprehend? If not, what part of the image made it difficult to comprehend?

Question 4: (Text Group): Were the text descriptions describing differential privacy difficult to understand? If so, what portion of the description was hard to understand?

Question 5: Did you find this question difficult to understand at all? If so, what portion of the question did not make sense?

Question 6: Did the previous description of differential privacy help with this question at all? If not, what do you believe should be changed during the description portion of the survey?

Question 7: Based on our hypothesis, we expected 'Yes' to be the chosen answer. Why did you choose "Yes" or "No" in this situation?

Question 8: How different would your answer have been before taking the survey?

BIBLIOGRAPHY

Bartha, P. (2019). "Analogy and Analogical Reasoning", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/reasoning-analogy/>.

Beveridge, M., & Parkins, E. (1987). Visual representation in analogical problem solving. *Memory & Cognition, 15*(3), 230-237.

Bullek, B., Garboski, S., Mir, D. J., & Peck, E. M. (2017, May). Towards understanding differential privacy: When do people trust randomized response technique?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3833-3837).

Bun, M., & Steinke, T. (2016, November). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Theory of Cryptography Conference (pp. 635-658). Springer, Berlin, Heidelberg.

Chen M., Floridi L., Borgo R. (2014) What Is Visualization Really For?. In: Floridi L., Illari P. (eds) The Philosophy of Information Quality. Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science), vol 358. Springer, Cham. https://doi.org/10.1007/978-3-319-07121-3_5

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*(2), 230-253.

Cuff, P., & Yu, L. (2016, October). Differential privacy as a mutual information constraint. *In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 43-54).

Cummings, R., Kaptchuk, G., & Redmiles, E. M. (2021, November). " I need a better description": An Investigation Into User Expectations For Differential Privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security* (pp. 3037-3052).

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), Theory of Cryptography (pp. 265–284). Springer. https://doi.org/10.1007/11681878_14

Dwork, C., Naor, M., Reingold, O., & Rothblum, G. N. (2015, November). Pure differential privacy for rectangle queries via private partitions. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 735-751). Springer, Berlin, Heidelberg.

Dwork, C., & Rothblum, G. N. (2016). Concentrated differential privacy. arXiv preprint arXiv:1603.01887.Gentner, D. (2003). Analogical reasoning, psychology of. Encyclopedia of Cognitive Science, 1, 106-112.

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science, 9*(3), 672-693.

Gentner, D., & Maravilla, F. (2017). Analogical reasoning. In The Routledge International Handbook of Thinking and Reasoning (pp. 186-203). Routledge.

Gentner, D., & Smith, L. A. (2013). Analogical learning and reasoning. *The Oxford Handbook of Cognitive Psychology*, 668-681.

Goldschmidt, G. (2001). Visual analogy—a strategy for design reasoning and learning. In *Design knowing and learning: Cognition in design education* (pp. 199-219). Elsevier Science.

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 264-272. doi:http://dx.doi.org/10.1037/a0025764

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*(3), 446-474.

Holyoak, K. J. (2012). Analogy and relational reasoning. https://psycnet.apa.org/record/2012-08871-013

Kiernan, M., Oppezzo, M. A., Resnicow, K., & Alexander, G. L. (2018). Effects of a methodological infographic on research participants' knowledge, transparency, and trust. *Health Psychology, 37*(8), 782.

Laureiro-Martínez, D., Canessa, N., Brusoni, S., Zollo, M., Hare, T., Alemanno, F., & Cappa, S. F. (2014). Frontopolar cortex and decision-making efficiency: comparing brain activity of experts with different professional background during an exploration-exploitation task. *Frontiers In Human Neuroscience, 7*, 927.

Lee, H. B. (2017). *Visualization and differential privacy* [Master's thesis, University of Illinois at Urbana-Champaign]. https://www.ideals.illinois.edu/bitstream/handle/2142/99106/LEE-THESIS-2017.pdf?sequence=1&isAllowed=y

McDonnell, L., Barker, M.K. and Wieman, C. (2016), Concepts first, jargon second improves student articulation of understanding. *Biochem. Mol. Biol. Educ., 44*: 12-19. https://doi-org.ezaccess.libraries.psu.edu/10.1002/bmb.20922

Mekhail, C., Zhang-Kennedy, L., & Chiasson, S. (2014). Visualizations to teach about mobile online privacy. *In Persuasive Technology Conference (poster).*

Nanayakkara, P., Bater, J., He, X., Hullman, J., & Rogers, J. (2022). Visualizing Privacy-Utility Trade-Offs in Differentially Private Data Releases. *arXiv preprint arXiv:2201.05964*.

Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews *BMJ 372 :n71* doi:10.1136/bmj.n71

Riva, J. J., Malik, K. M., Burnie, S. J., Endicott, A. R., & Busse, J. W. (2012). What is your research question? An introduction to the PICOT format for clinicians. *The Journal of the Canadian Chiropractic Association, 56*(3), 167.

Saket, B., Endert, A., & Demiralp, Ç. (2018). Task-based effectiveness of basic visualizations. *IEEE Transactions On Visualization And Computer Graphics*, *25*(7), 2505-2512.

Sunstein, C. R. (1993). On analogical reasoning. *Harvard Law Review, 106*(3), 741-791.

The European Parliament and of the Council (2016, Article 4). The General Data Protection Regulation (GDPR) (EU) 2016/679 of The European Parliament and of the Council of 27 April 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434

United States Department of Labor. (n.d.). Guidance on the protection of personal identifiable information. United States Department of Labor. Retrieved March 30, 2022, from https://www.dol.gov/general/ppii#:~:text=Personal%20Identifiable%20Information%20(PII)%20is,either%20direct%20or%20indirect%20means.

United States National Institute of Health [NIH]. (2022) Summary of the HIPAA Privacy Rule. https://www.hhs.gov/sites/default/files/privacysummary.pdf

Valdez, A. C., & Ziefle, M. (2019). The users' perspective on the privacy-utility trade-offs in health recommender systems. *International Journal of Human-Computer Studies*, *121*, 108-121.

Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., ... & Vadhan, S. (2018). Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.*, 21, 209.

Yamane, Taro (1973). "Statistics: an introductory analysis." New York: Harper & Row.

Yang, B., Sato, I., and Nakagawa, H. (2015). Bayesian differential privacy on correlated data. In *SIGMOD International Conference on Management of Data*, 747–762. ACM, 2015.

Jorgensen, Z., Yu, T., & Cormode, G. (2015, April). Conservative or liberal? Personalized differential privacy. In *2015 IEEE 31St International Conference On Data Engineering* (pp. 1023-1034). IEEE.

Zhang, D., Sarvghad, A., & Miklau, G. (2020). Investigating visual analysis of differentially private data. *IEEE Transactions On Visualization And Computer Graphics*, *27*(2), 1786-1796.

Zhang-Kennedy, L., & Chiasson, S. (2021). A systematic review of multimedia tools for cybersecurity awareness and education. *ACM Computing Surveys (CSUR), 54*(1), 1-39.

ACADEMIC VITA

Ephraim Nkosilathi Tendaishe Alexander Govere

Education

The Pennsylvania State University, University Park, PA
BS Cybersecurity Analytics and Operations, May 2022

- Dean's List 7/7 Semesters
- Scholar, Schreyer Honors College
- Honors Thesis: Communicating Privacy-Utility Trade-Offs of Differential Privacy Through Analogical Reasoning.

Work Experience

The Pennsylvania State University, University Park, PA
Intern, Research Experiences for Undergraduates (REU), June 2021 – Present

The Pennsylvania State University, University Park, PA
Information Technology Support Specialist, Aug 2019 – May 2020

Certifications

Cisco Certified Network Associate (CCNA)
Cisco Certified Entry Network Technician (CCENT)

Awards

Research Experiences for Undergraduates (REU) Internship Award
Penn State Capital Scholar's Scholarship Award
Penn State Harrisburg Scholarship Award
Commonwealth Campus First Year Student Award

Associations

The Honor Society of Phi Kappa Phi
Information Systems Security Association (ISSA)
Competitive Cybersecurity Organization (CCSO)
The National Society of Leadership and Success