2020-01

Exploring Context for Privacy Protection in North American Higher Education and Beyond

Wu, Leanne

UNIVERSITY OF CALGARY

Exploring Context for Privacy Protection in North American Higher Education and Beyond

by

Leanne Wu

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMPUTER SCIENCE

CALGARY, ALBERTA

JANUARY, 2020

Abstract

Undergraduate students in North American post-secondary institutions are subject to a wide range of data collection. It includes data generated in the course of teaching and learning, but also can include a wide range of other aspects of modern life, such as closed-circuit security cameras, internet and wireless network use, and what students buy and consume. This makes the post-secondary institution an ideal model for understanding the privacy impact of modern and future technologies, as a single organization which collects and potentially uses wide-ranging amounts and kinds of data about our daily lives.

This thesis proposes a framework which separates context into three interrelated layers so that systems can be designed which more fully protect the privacy of individuals, examines the ways in which we collect and use data about undergraduate students, and makes a quantitative study of undergraduate privacy behaviours and attitudes. Thus we present the case that context is a core concept for privacy protections which better protect undergraduate students and their privacy.

Acknowledgements

As lengthy as this journey has been, there are many people who have contributed some piece to this thesis, and I'm grateful for the opportunity to acknowledge their part in this work.

First of all, I'd like to thank my supervisor, Ken Barker, for his patience and willingness to see me through this process, and for many insightful conversations over the years.

I'd also like to thank the other members of my supervisory committee. Lora Oehlberg stepped in quite late in the process, but her insight on research involving human subjects and her comments on my thesis have strengthened my research considerably. Joerg Denzinger has also provided much advice over the years, and I value his forthrightness and expertise greatly.

There are many others in the Department of Computer Science who have provided their support, advice, and mentorship over the years. I would like to thank Nathaly Verwaal, Ben Stephenson, Rei Safavi-Naeini and Carey Williamson especially in this regard.

No journey like this happens alone, and I have many colleagues (nearly too many to mention) who have been invaluable fellow travelers with me. I'd like to particularly mention Maryam Majedi, Mishtu Banerjee, and Rosa Karimi Adl as close collaborators, but there are a plethora of graduate students who have engaged with my research and helped me develop ideas through PSec meetings, worked with me on planning and executing various events and initiatives, or crossed paths with me in other meaningful ways. Thank you to each and all. The world is lucky to have each and all of you.

My interest in the kinds of data post-secondary institutions collect and use really began when I took a temporary position to help faculty members transition from a thing to another thing. I'd like to thank Cindy Graham and Leslie Reid at the Faculty of Science, and Michael Ullyot at the Faculty of Arts for taking a chance on me, and for starting me on a journey that led to my being very curious about how universities work, especially once data comes into the picture.

At the Taylor Institute, I want to thank the Technology Integration Group, and particularly D'Arcy Norman, Isadora Mok-Kulakova and Kevin Saito for making me feel welcome and valuable, even as the ideas started to bounce off them at a frantic rate. They really are the best team, ever. There are many others in the Educational Development Unit and the Learning Technologies coaches team who have been good friends as well, especially Lin Yu, Haboun Bair, Patti Dyjur, Marlene Mansell, Brian Pshyk, Judy Tran, and Grace Whitehead. It was my time here which gave me a direction for this thesis and I am grateful.

The University of Calgary has been a wonderful professional home for me over the last

decade, and I am grateful to so many I have met in this community of who have given me a bit of advice, insight or encouragement over the years. This section is supposed to be shorter than the thesis, or I would thank each and every one.

I would also like to acknowledge the kindness and generosity of my friends. Quoc Nguyen, Paul Rodriguez, and John Tran have been there from the start, when I first moved back to Calgary. Along the way, Kate Chatfield-Reed and Chris Young, Sean Nichols, Jonathan and Stephanie Hudson, Andrew and Andrea Kuipers, Ryan Yee, Erika Harrison, Mark Koleszar, David Aikema, Man-Wai Chu, Jordan Kidney, Jagoda Walny and Simon Nix have all been there with advice, board games, and restorative daytime adventures, evenings and short trips. There are many others I should be thanking as well. I wouldn't have made it this far without their insight, understanding and willingness to kick back (and make me relax).

The same goes for all my extended family who have been there to remind me there's more to life than school. I am especially grateful for the support of Alvin and Mable Wong, Diana and Kam Wong, and Stella and Richard Li. And the cousins. I would list them all, but then I'd definitely run out of space.

Finally, Tyson Kendon, his parents, and my parents, Edward and Vera Wu. You're everything, and I don't have words. Thank you, thank you, thank you.

For my parents, who were always there.

Table of Contents

List of Figures and Illustrations

List of Tables

Chapter 1

Introduction

The last decade has seen a startling change in the role the Internet plays in the lives of citizens around the world. It has moved from an academic and scientific curiosity for researchers and hobbyists to an omnipresent force in everyday interactions. Internet access has shifted from being intentional, deliberate, and difficult to being nearly automatic, unthinking, and difficult to avoid. Where once one might have expected the Internet to be involved in only a small range of specialized tasks, such as accessing a webpage or posting messages to a newsgroup, now the likelihood exists that some data flows to the Internet even for commonplace tasks such as changing television channels, adjusting the lighting in a room, or driving a child to a friend's house. One of the forces powering the growth of the Internet has been the ability of organizations to harvest data from people performing these simple tasks, and the ability to leverage this data to drive economic growth.

Such growth comes at a cost, and that cost is paid by the integrity of the privacy of each individual touched by the Internet. The recognition of this cost is now being acknowledged by individuals, organizations, and lawmakers, spurring the creation of legislation in recent years such as the European Union's General Data Protection Regulation (GDPR)[Reg16].

Researchers have become increasingly occupied with the question of how to best protect privacy. There have been a variety of approaches from the information management commu-

nity, the most notable of which have focused on republishing data. This is a scenario which happens when some organization holding data about a population of individuals republishes that data so a third party can use it, with the primary concern being how to best obscure the identity of the individuals that provided the data, so that they cannot be identified by the third party using the data.

While this is an important problem, particularly in areas such as health care, it shifts the focus from the privacy of the individual. In this scheme, privacy beliefs or preferences of the individual are not taken into account. Indeed, there are many stories about privacy issues, in which organizations profess to be following the letter of the law, to be using state-of-the-art technologies for privacy protection, but are still at odds with individuals who claim that their privacy has been violated[Lom19].

We seek an approach which centers the problems it solves upon the privacy of the individual. This approach should take into account how organizations will govern their own data collection and processing, but provides a means to take into account individual circumstances and exigencies.

1.1 Motivation

This thesis examines contextual privacy with particular focus on institutions of higher education, a domain in which individuals already experience data collection in nearly every aspect of their lives. The data collection experienced by individuals who are part of and interact with these institutions is straightforward compared to the kind of data collection experienced by the majority of society, since the bulk of the data collection is dominated by one organization, the institution.

In post-secondary educational institutions in North America, colleges and universities not only are responsible for data about a student's education, but also in many aspects of life which occur on campus. This may include information about significant amounts of

private life at home, for those living in residence, but also includes information about what a student accesses via the Internet, who that student is with, what that student eats, and what they do for recreation. Collecting, storing, and processing that data requires considerable discretion, and contextual privacy is one means by which administrators and data analysts may understand how to systematically enforce such discretion to protect student privacy.

1.2 Research questions and contributions

This thesis proposes a way for computer scientists to move forward with contextual integrity by providing a framework which bridges context as understood by information systems (as a real-time aggregation of facts about the surrounding environment[DAS01]) and context as understood by other fields such as law (which includes not just physical facts about the surrounding environment, but other influences, such as social or cultural factors[Nis09]).

The first goal of this thesis is to build an understanding of privacy and its interactions with context to create a conceptual framework for contextual privacy which could be operationalized and implemented in the future. Next, this thesis builds an understanding of the extant issues in information management in post-secondary education, particularly with emphasis on potential privacy issues. We then apply the framework to a selected set of privacy issues to demonstrate the utility of the framework and show where gaps in our knowledge remain. Finally, we present a survey of undergraduate privacy behaviours and attitudes as a step towards filling in remaining gaps.

This thesis presents three main contributions. The first is the creation of a conceptual framework for contextual privacy, providing a multi-layered approach to organize and coordinate both social and technical approaches to context. The second is a survey of post-secondary data collection and processing practices, as an area where the consideration of context is critical to understanding how to protect privacy. We apply the contextual framework as a lens to better understanding these practices. The third contribution is a survey

of undergraduate privacy behaviours and attitudes, to illustrate the importance of the other contributions.

1.3 The Structure of this Thesis

We begin this thesis with a discussion of privacy, examining both the philosophical and legal landscape (Chapter 2) and dominant technical approaches (Chapter 3). Chapter 4 introduces the concept and importance of context, in particular Nissenbaum's framework for contextual integrity[Nis04; Nis09], the state of research involving contextual integrity, and a discussion of the available opportunities to contribute to the further development of this concept.

Chapter 5 presents our first contribution, a conceptual framework for contextual privacy. In this chapter, we show the key components of the framework and how they interact, using detailed examples, both at a high and low level.

We move to our second contribution beginning in Chapter 6, which discusses data collection and use in post-secondary institutions, with special attention to privacy issues which are specific to this domain. Chapter 7 bridges the previous two chapters with a detailed case study which investigates particular instances of data collection and processing in the setting of a Canadian university. We situate these instances in contextual privacy by applying the framework of Chapter 5 to the data collection and processing practices discussed in Chapter 6.

This creates the case for a more detailed study of undergraduate privacy behaviours and attitudes to fill in the gaps of our knowledge. We will begin to address these gaps in Chapter 8, which presents our third contribution, a study on undergraduate privacy attitudes and behaviours.

Finally, we conclude this research in Chapter 9 with a discussion of our findings and an exploration of future work which will arise from this research.

Chapter 2

Foundations of Privacy

This chapter offers a quick tour of privacy, through history. It is by no means comprehensive, but aims to highlight some key aspects to privacy protection that are central to the contributions of the thesis. These foundations underlie the practices around how we handle data, especially the legislation and policy which govern it.

2.1 Classical origins

Privacy has long been recognized as a vital aspect of how individuals interact with the societies to which they belonged. In the classical world[Fai05], the Greeks had a division between *oikos*, the world of the hearth, home, and family, and *polis*, the world of public life. Thus, some things were kept close (to the home), and some things were made public, and shared with the surrounding community. This division has influenced the Western world since that time[Hor81].

As with many aspects of the humanities, ideas which have shaped privacy studies to date have centered upon Western thought, philosophy, and religious tradition. Although a significant degree of scholarship remains before these are brought into the common discourse, it is worthwhile to note that there are differing views of privacy around the world, highly dependent upon culture, tradition, and ways of living. For example, Confucian tradition

also separates the home from the public life, but restricts the degree of autonomy which individuals claim within both[Pen03], especially compared to Western ideas of privacy. This separate attitude towards privacy may acquire growing significance as China continues to incorporate an increasing degree of prevalent and automated surveillance to maintain order and power within its own borders in the near future. There is also a growing awareness and sensitivity that many indigenous peoples around the world also have diverse approaches to the consideration of what is private, what can be shared and what should not be shared, and have a history of suffering harms from colonial approaches to data collection which have not been respectful of indigenous peoples and communities[06].

Legislation and policy is based upon how society values specific rights. We turn now to how these have been implemented in Europe and North America.

2.2 The 19th Century

One of the most influential (and widely considered to be the first) treatments of privacy in modern American legal scholarship is in a legal brief written by Warren and Brandeis[WB90]. Written in response to the increasing use of photography by journalists to record the comings and goings of members of society from events, the brief describes specific threats to privacy, noting that the modern (for the time) technology had intensified the number and severity of threats to an individual's privacy. These threats are clearly of concern to society, yet exist without violating the laws of the time, most notably in the areas of defamation and property, both physical and intellectual.

Warren and Brandeis make an argument in favour of privacy rights which would exist outside of these extant rights, providing a definition of privacy as the"right to be left alone." Having defined this right, the brief finally discusses the types of limitations which should be applied to privacy, and discusses how to best integrate it into the legal framework of the time.

In addition to its influence on subsequent legislation of privacy (stretching even to the modern day), of particular relevance to our work is the authors' focus on the role of technology in pushing privacy rights to the forefront. At the time, they were discussing the advent of photography, which simplified the monitoring of those who attended social events, but of course, technology has quickly developed beyond these techniques to create new methods which can potentially harm the privacy of individuals.

2.3 The 20th Century

The events leading up to and during the Second World War saw the deaths of millions of noncombatants, in part due to the mass collection and processing of information of citizens by totalitarian governments, aided by technology such as the automated processing of punch cards. In response to these enormous harms, Article 12 of the Universal Declaration of Human Rights[Ass48] asserts the right to privacy and for legal protection of this right for all of humanity. The scope of this article explicitly includes aspects of an individual's life which could be understood to comprise part of an individual's private life, including the integrity of their family, home life, correspondence, reputation, and honour.

As consumer privacy and bulk processing of data became more commonplace in the later part of the 20th century with the rise of computation, most developed countries added privacy legislation to ensure safeguards were in place for their citizens. For example, in 1980, the OECD put forth principles of data protection[OEC80], the majority of which are included in most major privacy legislation enacted in the latter part of the twentieth century, primarily amongst the member nations of the OECD.

They are:

Collection Limitation. Data collectors should not collect more personal data than is required, ideally with the consent of the data provider.

Data Quality. Personal data should be as accurate, current, and complete as

possible

Purpose Specification. The purpose for which the data is to be collected and used should be specified at the point of collection, and furthermore, the use of such data should not exceed the purpose specified.

Use Limitation. Personal data should be kept confidential and not shared with other parties or used for purposes beyond what is originally specified at the point of collection.

Security Safeguards. Data should be kept secure and not vulnerable to unauthorized access, use, modification, or disclosure by other parties.

Openness. The practices which employ personal data should be transparent, so that data providers can understand where their personal data is, how it is being used, and by whom.

Individual Participation. Data providers must be able to confirm whether data collectors hold their personal data, be able to receive a copy of the personal data being held, and to request that this data be modified or deleted.

Accountability. There must be a means by which data collectors are held accountable for upholding the other principles.

Many member nations (including Canada) have enacted legislation which adopt some or all of these principles to protect their citizens. In Canada, two federal laws protect the privacy of citizens. The Privacy Act (first becoming law in 1983) governs how the federal government may handle personal information, and the Personal Information and Electronic Documents Act (commonly PIPEDA, first becoming in 2000)[Can00] governs how private bodies should handle personal information. Many public institutions such as municipal authorities, universities and health care providers fall under provincial jurisdiction, and

thus are subject to privacy legislation at the provincial level, such as Alberta's Freedom of Information and Privacy Act (which first came into law in 1995)[Alb00].

Schedule 1 of PIPEDA describes ten guiding principles of privacy protection which informs the main body of the act. These principles are clearly drawn from the OECD data protection principles, including much of the same vocabulary around (but not limited to) purpose specification, collection limitation, accountability, limitation of use, and accuracy. Included for each principle is an elaboration of the obligations and duties of each organization collecting and processing personal information under the auspices of PIPEDA.

For example, Principle 2 (Identifying Purposes) begins:

> *The purposes for which personal information is collected shall be identified by the organization at or before the time the information is collected.*

But continues with the following (and selected) clauses:

> ***4.2.3***
>
> *When personal information that has been collected is to be used for a purpose not previously identified, the new purpose shall be identified prior to use. Unless the new purpose is required by law, the consent of the individual is required before information can be used for that purpose. For an elaboration on consent, please refer to the Consent principle (Clause 4.3)...*

> ***4.2.5***
>
> *Persons collecting personal information should be able to explain to individuals the purposes for which the information is being collected.*

Because PIPEDA so clearly includes the data protection principles defined by the OECD, it is often held as an exemplar for privacy legislation. However, new developments on social, technical, legal, and political fronts are necessitating significant change in privacy legislation around the world.

2.4 The 21st Century

Legislation has struggled to keep up with world events and technological developments which have threatened data privacy in recent years. Consider the stringent focus on national security adopted by Western states after the events of September 11, 2001, and the enhanced ability for a wide variety of actors (whether governmental or corporate) to gather and process previously unthinkable volumes and types of data. At the same time, lawmakers have not kept pace with the increasing pervasiveness of data collection in everyday life. In that time, we have seen social media, mobile devices, enhanced surveillance infrastructure (such as closed-circuit cameras deployed on a wide scale, as in the United Kingdom), and devices powered by the Internet of Things, all contribute to growing the sophistication of data collection, at the same time that data processing has become massively enhanced by the application of machine learning and other analytic techniques at unprecedented scale.

At the same time, many organizations apparently feel that the penalties imposed by legacy privacy legislation do not warrant alterations to their current data handling practices. For example, despite specific and urgent concerns raised by the Canadian federal government with regards to Facebook's marketing of data to a wide range of other parties, senior leaders at Facebook have repeatedly turned down the government's invitations in 2019 to attend parlimentary hearings on the matter[Can19]. Other attempts to protect privacy in Canada such as the Do-Not-Call list are seen as ineffective[Pad19], in part due to the difficulty of imposing measures which are proportional to the harms borne by the public.

2.4.1 The General Data Protection Regulation

One attempt to update privacy legislation in member countries of the European Union is the General Data Protection Regulation[Reg16], passed in 2016 with enforcement beginning in 2018. It requires organizations collecting and processing personal data to provide stringent protections for consumer data and enacts significant penalties for non-compliance. GDPR

is a modernization of previous privacy legislation in the EU and its member states, with an explicit acknowledgement of the growing importance of data, data analysis, and the increasing difficulty of protecting individual privacy in the face of technological progress.

As such, the wording of the regulation provides more clarity than in previous legislation, defining concepts such as *personal data*, *consent*, and *transparency* explicitly, and drawing reference to approaches such as pseudonymization which are well-researched and implemented (and will be discussed in detail in the following chapter). One of the major differences between the GDPR and pre-existing privacy legislation is the amount of clarity with which it defines how organizations should be responsible for the data they collect and the data they process (or release to a third party for future processing), delineating the parties in each organization who should be responsible for the data, and the procedures by which they should be approached for potential remedies.

The regulation also attempts to be more forward-thinking in its protections, for example, making frequent reference to genetic and biometric data as types of data which require special protection, especially in Article 9 (Processing of special categories of protected data). This attempts to implement protections far beyond the current *status quo* which sees infrequent use and processing of such data compared to other kinds of data which are more readily available, but which the framers of the GDPR clearly see as a serious future threat, since such data is irrevocably and irreversibly linked to a single individual.

This legislation has attracted a great deal of attention worldwide because it is applicable to any organization which may collect or process data about citizens of EU member states, and thus a great number of organizations are subject to its regulation. There are also improvements to its enforcement framework, and the promise of significant fines that can be levied on violators, which means that most enterprises cannot ignore its strictures.

As the GDPR begins to be enforced in the EU, many other jurisdictions will be observing its impact to determine how to revise their own privacy legislation. However, it remains to be seen whether GDPR-style legislation can be adopted in other jurisdictions such as the United

States, since there remain vast differences in how each society weighs the balance between privacy and other considerations, such as free speech, national security, and individual rights.

Chapter 3

Privacy and Data Management

This chapter examines the interaction of privacy and data management, with reference to some of the most influential works in this area. We begin with what is perhaps the broadest proposal for privacy protection in the field of data management. We use this proposal to explore subsequent work which was intended to address more specific questions and issues surrounding privacy.

3.1 Hippocratic Databases

A seminal paper by Agrawal *et al.*[Agr+02] in 2002 proposes that the data management community build a database management system which commits to respecting the privacy rights of the individual in much the same way that physicians commit to when they vow to uphold the Hippocratic Oath. Dubbing it the *Hippocratic Database*, the paper sketched a "straw-man" design for a database which would implement some of the major legal guidelines for privacy preservation in a database system.

The contributions of the paper were three-fold: providing the field with a vision for a privacy-preserving database and motivating it with urgency and importance; outlining the legal guidelines for which the system must be able to provide a technical solution; providing a rough template to design a privacy-preserving database and thereby highlighting some of

the most important problems yet to be solved.

The ten privacy principles elucidated by the paper align well with modern international legislative standards for privacy which have been described in the previous chapter, although some of the principles used in the paper are more rigidly defined. For example, Agrawal *et al.* define the principle of limited collection as follows:

> The personal information collected shall be limited to the minimum necessary for accomplishing the specified purposes.

However, we can look to the OECD's privacy principles[OEC80], one of the legislative foundations for Hippocratic Databases, to find a broader definition for the principle of collection limitation :

> There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

This example illustrates that there may be agreement as to what privacy principles should be observed and enforced, but that there also exist differences surrounding how specific those principles are, and what protections those principles should afford. The version of this principle proposed by Agrawal *et al.* is more readily supported by information management systems, while that proposed by the OECD is more far-reaching and thus more protective of individual privacy rights, and could support many interpretations (and proposals for implementation).

Thus, in our analysis of accompanying and ensuing work, it is critical to recall that the principles laid out by Agrawal *et al.* in their work on Hippocratic databases have been seen as seminal, but they do not encompass the only approach to operationalizing privacy legislation in data management.

The straw-man design proposed by Agrawal *et al.* provides for a number of components which are built around the core of what is primarily a traditional database management

system (typically referred to in the literature as 'DBMS'). These components provide privacy-specific functionality for inputs arriving to the DBMS such as assistance with authoring and storing privacy policies, validating incoming data and queries, as well as privacy-specific functionality on an ongoing basis, such as checking on limitations to data retention, and performing longer-term analysis on data access in the system.

A few enhancements to the DBMS are proposed by the design, primarily allowing for the storage of privacy-specific metadata, data pertinent to enhanced auditing provided for by the Hippocratic database as a whole, as well as enhanced access control mechanisms.

3.1.1 Building Hippocratic Databases

The power of the work surrounding Hippocratic databases was such that immediately thereafter, the data privacy research community witnessed several efforts to build a Hippocratic database. These tended to immediately highlight some of the flaws inherent in the original design. In this section, we discuss some of the more well-known attempts to point out areas where additional work was required for a complete implementation.

Agrawal *et al.* extended their vision of a Hippocratic database with several follow-up papers[LeF+04] and at least one patent. In these, access control is a major focus in the development of the Hippocratic database, building upon well-established schemes already implemented in major DBMSs, so that they can quickly provide support for privacy policies. There are two specific contributions in their follow-on work. The first provides cell-level access controls in a DBMS, where traditionally (and most sensibly, for a relational data management system), the granularity for most access control mechanisms is at the relation (or row) level. The second provides a way to convert (as simply as possible) privacy policies written using a pre-existing standard (the Platform for Privacy Preferences, also known as P3P[Cra+02]) into a format that could be used by a relational DBMS.

Another effort undertook to build a Hippocratic database using PostgreSQL[Pad+09], a popular open-source DBMS. This implementation includes support for privacy policies,

similar to LeFevre *et al.*[LeF+04], in which primitives were modeled to fit both SQL as well as P3P. The main contribution of this work, however, is in extending ideas about data minimization which had been developing in the privacy community.

This area of research begins roughly contemporaneously with Hippocratic databases, but originates with statistical databases and aggregation, particularly for use in a healthcare setting (and thus, for American researchers, relying on more stringent legislation found in HIPAA). The most well-known techniques in this research broadly support the principles of limited collection and limited use which are described in law, and Padma *et al.*[Pad+09] in their work on Hippocratic PostgreSQL include an implementation of some of the most common techniques (such as k-anonymity[Swe02]) in a database engine.

As an attempt to translate the straw-man design in the original paper into an implemented, operational system, there still remain issues with the Hippocratic PostgreSQL system. Perhaps the most significant is that it continues to downplay the importance of defining what privacy should mean for data collectors, those processing private data, as well as for individual data providers.

The proposal to create a Hippocratic database which integrates privacy as a primary and intrinsic design principle was ambitious but has not yet been completely accomplished, due to the scope of the work, and how the initial straw-man design influenced subsequent attempts at implementation. This effort also pointed out the need to consider the semantics of privacy as a nuanced concept which requires detailed definition before attempting to operationalize a system.

The field of data management has largely moved onto other areas of research which provide opportunities to define more formal definitions of privacy. We move to three areas in which privacy research has been focused in the last decade: policy-based approaches, techniques in data minimization, and the semantics of privacy.

3.2 Policy-based approaches

In this section, we discuss privacy-centric approaches to creating and enforcing policies for accessing resources (typically data) in a system. Access control models are an area of research which had been well-explored prior to the Hippocratic database proposal, and continue to be developed well beyond what is described by Agrawal *et al.*[Agr+02] in their vision.

3.2.1 Access Control

Since nearly the inception of relational databases, access control is the standard approach to securing access to information in databases. In the simplest form of access control, rules are defined so that a *subject* is allowed or denied access to an *object*.

As size and complexity of information systems has grown considerably over time, so too have access control schemes adapted to meet these new demands. The remainder of this section sketches some of the developments most relevant to privacy.

One well-known access control scheme is role-based access control [San+96; FCK95], which has been adopted by most major operating systems and commercial DBMSs. In role-based access control schemes, subjects are grouped according to their *roles* in the organization, with rules created to define for each role the permitted (or disallowed) discretionary access of individual data items. Role-based access control schemes also provide additional constraints for securing data. Common constraints ensure that an individual cannot simultaneously be granted and denied access to the same resource, even when grouped into more than one role in the system, nor that an individual (again, by simultaneously inhabiting multiple rules) cannot acquire too much responsibility in the system. Some role-based access control schemes also support the composition of roles into a hierarchy to further reduce the number of rules to be defined.

This has been amended by introducing conditions and obligations[Bet+02]. These provide a way for policies to require actions immediately preceding or following the access of

an object. Conditions and obligations provide a means to implement some of the privacy principles which are central to designing privacy-preserving systems, specifically purpose specification (and other principles which are often strongly related to the concept of purpose), and those requiring the ability to audit access to private data. The OECD principles describe this as "auditability," whereas Agrawal *et al.* refer to it as compliance. In either case, this principle is intended to allow data providers to verify that the other privacy principles are applied appropriately by the data collector for the private data which has been collected from them.

Privacy-aware Role-Based Access Control[Ni+07] creates space for privacy in role-based access control by providing support for purpose. By supporting purpose, conditions, and obligations, most of the privacy principles described by the OECD can be supported. However, the specification of purpose, conditions, and obligations in an operational system can cause the number of rules which must be created, checked and maintained to grow in an unmanageable fashion. There have been attempts to control this growth in the number of rules by specifying purpose in different ways, for example, as a workflow[Jaf+11], but creating access control schemes which are adequately expressive, but also decidable and relatively efficient still remains a challenge.

3.2.2 P3P: The Platform for Privacy Preferences

Practitioners realized in the early 21st century that there was an increased need for privacy policies which could be attached to webpages so that consumers could readily determine whether a given service provider would respect their privacy preferences. The intent was to create a standard for specifying policies which could be easily understood both by laypersons (such as consumers) and experts (such as those in organizations who specialize in data governance and compliance).

The Platform for Privacy Preferences (or P3P)[Cra+02] is an XML-based standard intended to allow internet users to specify their privacy preferences, for easy comparison to

published privacy policies of websites they visited. Tools would be provided to automatically filter out websites whose privacy policies did not match their privacy preferences, or to provide a warning to users that their privacy preferences would not be observed.

However, a number of flaws with this platform restricted its adoption over the intervening years. For example, the focus on compliance is with a data collector's stated privacy policy, which may not actually conform to their operational privacy *practices*. Hence, there is no way to dynamically align a data collector's privacy practices with the consumer's stated set of privacy preferences.

We also find that with P3P, many of the privacy primitives defined by the platform do not require detailed definition. For example, it is possible, when specifying elements for purpose, to specify `<OTHER>` or `<CUSTOM>`. While the standard stipulates that this must be accompanied by a detailed explanation, data collectors can easily skirt this stipulation by providing insufficient information.

This ability to keep privacy primitives at a general level of specification means that blanket policies which are essentially not at all protective of privacy can be used at most times by data collectors to skirt the privacy concerns of individual consumers.

A significant portion of the Hippocratic databases paper by Agrawal *et al.*[Agr+02] is devoted to the conversion of P3P into policies understood by the Hippocratic database, which enables the database to quickly support standards already in use. However, there is a missed opportunity here to deeply consider the semantics of privacy, and specific purpose. While P3P saw wide adoption in the early 2000s, there are ways in which it approaches privacy which, while pragmatic, are not useful. For example, purpose is quickly represented using merely a string, which can in fact read "any." Thus while purpose is represented in P3P, many practitioners may argue that it is not truly specified or enforceable.

We also observe the danger in supporting standards which are then not rapidly adopted by the technical community, as the last major update to P3P took place in 2006, and the standard is no longer supported by most modern consumer applications.

3.3 Data minimization

We use the term "data minimization" to refer to techniques which address principles of privacy protection which exist in both the legal and technical realms. These include the principles of limited collection and limited disclosure (the idea that the least amount of personal information should only be used and disclosed for the purposes which were specified at the time of collection), as well as that of safety (that personal information should be protected from misuse).

The primary goal of these techniques is to ensure that the data disclosed by individuals keeps separate the pieces of data which may be *personally identifiable* from those that are *sensitive* as much as possible. While sensitive attributes often indicate that an individual (and their data disclosed in the system) may be of interest for further analysis, the intent is to obscure those pieces of data required to identify an individual from a larger set of their peers.

3.3.1 Anonymization

Anonymization broadly refers to a series of techniques by which a dataset is stripped of selected pieces of personally identifiable data before being released for further analysis. These techniques are generally proposed with two goals in mind. First of all, to maintain the utility of the entire dataset, anonymization techniques seek to minimize the amount of data lost through this process. Secondly, anonymization techniques seek to produce a redacted dataset which can pass a series of attacks based upon statistical analysis, which vary, depending on the specific technique chosen.

k-anonymity

First proposed by Sweeney in 2002[Swe02], k-anonymity is the best-known anonymization technique. The primary goal is to prevent the suppression of more than just the minimal

data required to keep data private. To do this, Sweeney identifies groups of attributes which in combination, function as *quasi-identifiers*, meaning that they can be used to identify data providers even when attributes which conventionally are considered to be personally identifiable must be suppressed. Thus, to protect the identity of individual data providers, Sweeney defines the property of k-anonymity to be such that each quasi-identifier must exist in the data set at least k times.

l-diversity

Although Sweeney identifies several potential types of attack in her work with k-anonymity, there remain many other flaws. For example, the proponents of a subsequent anonymization method, l-diversity point out one such flaw[Mac+07]. Much is done to prevent the identification of individual data providers, but very little is done to unlink sensitive values from those identified by a quasi-identifer. They point out that k-anonymity in effect, divides datasets into equivalence classes of at least size k, and without careful attention to the sensitive values in a dataset, it is possible to assume that each equivalence class can be linked to a sensitive value. Thus, l-diversity is defined such that there must be at least l distinct sensitive values for each of these equivalence classes.

Recent work in anonymization

The work into anonymization continues much in this vein: A work claims that privacy protection is achieved through the application of a particular technique, and then a follow-on work exploits a statistical property of the data generated by the results of this technique to subsequently point out a critical flaw, and proposes a new modification of the technique, which is in turn critiqued and used to produce another follow-on work. As additional measures must be considered to achieve better anonymization of a dataset, increasingly, the utility of data processed in this manner decreases, as does the difficulty of achieving a good anonymization scheme.

Additional follow-up work also seeks to optimize the performance of different anonymization techniques. Because the techniques have been developed in response to numerous attacks based upon statistical properties identified in different datasets, identifying suitable sets of identifiers for anonymization can be a time-consuming exercise. Data taken from operational environments seldom exhibit the same statistical properties as datasets which have been carefully authored and chosen to fit specific schemes for anonymization, and are often significantly larger, both in cardinality and in the number of fields being included.

3.4 Differential privacy

Ever since Dwork's seminal work[Dwo06][Dwo+06], differential privacy has been widely hailed as a gold standard in the area of data privacy. This approach seeks to provide a formal guarantee for the privacy of a dataset.

Differential privacy was developed as an extension of a wide body of research in statistical databases, in which a published dataset is used to perform a specific analytic task. Therefore, the goal of differential privacy is to perturb that dataset in a way which mimics the contribution of each individual to the dataset, in such a way that should that dataset be published with one fewer individual, it would not be possible to determine which individual was missing.

To produce this effect for a given dataset, there are two general approaches. Anonymization (or other related techniques) can be used to reduce the number of distinct features in the dataset, or alternately, noise can be added to the data to make distinct features less apparent. In both cases, the utility of the dataset decreases, since it is less accurate than before the perturbations were applied.

We offer some observations about this approach, before discussing subsequent work. First of all, the definition of differential privacy applies to a very specific set of uses of private data. It refers to the result of one query of one dataset, used in statistical databases, which handle

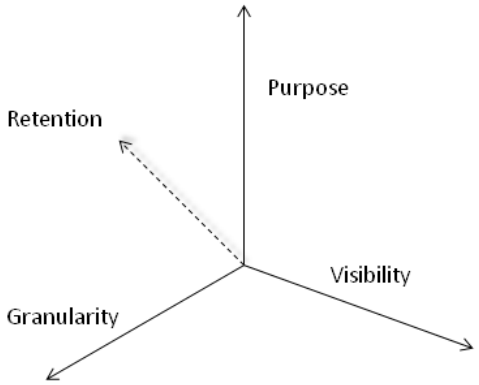
data in a specific manner, rather than for a general purpose workload in a conventional database. Secondly, this approach is effective only if the contribution of each individual of a dataset is relatively insignificant - either because the data set is very large, or not very private. Finally, the output of a differentially private dataset is problematic, since the utility of the data can be compromised, sometimes seriously, depending on how and to what degree the dataset was perturbed.

Since differential privacy has a formal definition which appeals strongly to the research community, a significant amount of follow-up work has built upon Dwork's original ideas. Additional work has sought to establish more efficient means of determining the level of differential privacy provided by a dataset, while other research groups have sought to apply these techniques to many areas of data analysis for which masses of data are required.

As contrasted against principles of privacy identified in legal theory and legislation (used as a foundation by most modern privacy policies), differential privacy addresses relatively few privacy issues. It is useful for considering limited disclosure and perhaps limited collection, but is not concerned with individual privacy, but instead the privacy of individuals in a aggregated and anonymized set of data.

Because differential privacy arises from a background of statistical databases and information theory, it is important to keep in mind that the typical threat model used by differential privacy assumes an external adversary who would ordinarily not be able to access the dataset as microdata, and who may only have a few (perhaps only one) accesses to the data. However, there is a rich area of potential research to examine how to enforce the ethical access and reuse of data by those who already have legitimate access to the private data of individual users, who can in practice retrieve the same dataset repeatedly to satisfy their queries and curiosity.

Figure 3.1: A simplified schematic of the privacy taxonomy by Barker *et al.*[Bar+09]



3.5 The Privacy Taxonomy

A scan of the extant work in data privacy led Barker's research group[Bar+09] to conclude that the field was strongly lacking in a robust definition of privacy. They proposed that a privacy taxonomy be used to frame the many different ways in which researchers coming from a data management background discuss privacy. This taxonomy consisted of four predicates: Purpose, Visibility, Granularity and Retention.

As shown in Figure 3.1, each predicate becomes an orthogonal dimension in the same geometric space, with values for each predicate arranged such that those which indicate more privacy protection (because they became more specific or restrictive) lay closer to the origin for each dimension than the other values used for that predicate.

Purpose This predicate defines why data is being collected, stored, or used. We can describe two extremes for purpose - a case in which data is collected for any and all purposes, and a case in which there is no purpose for the data to be collected, stored, or used. The first extreme describes purpose at its least protective state, whereas the second describes purpose at its most protective state, in which data cannot be used by the system. There are a number or proposals to structure the predicate, in terms of the points which exist in between these two extremes. The original work by Barker *et al.* proposes that we consider the number of purposes for which a data value could

be used, whether this is for one instance of usage, for a single purpose, or for multiple purposes.

Visibility This predicate shows which parties are permitted to access the datum. Again, there are two extremes, where no one at all is permitted access at the most protective point of the predicate, and at the opposite end, where anybody (in the original work, "the world") is permitted access. Between these two extremes, different levels of visibility must exist, although much more work has been done to specify these levels, from the work that has been done in the field of access control.

Granularity This predicate specifies the level of detail at which the datum may be accessed. This may vary from not seeing the datum at all (or indeed, any information about its existence), to being able to access the datum in its entirety (i.e. the"microdata"). In between, it is possible that only information about the datum's existence be released, or that the datum can be accessed with some detail suppressed.

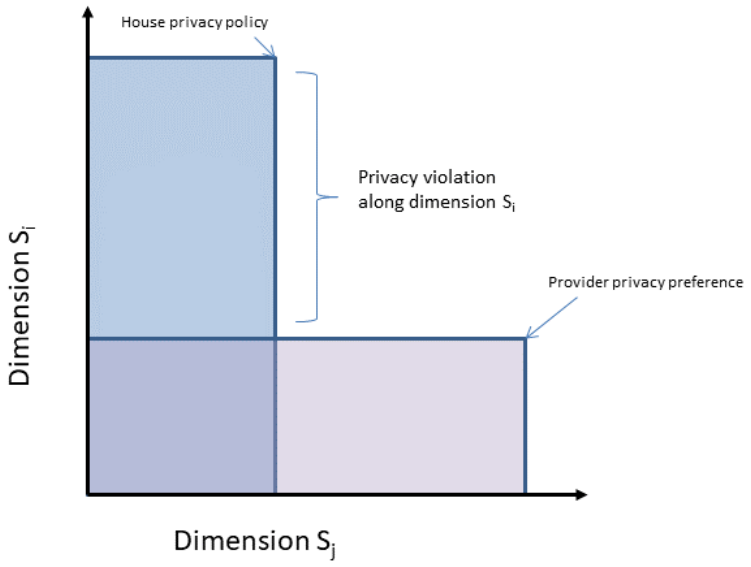
Retention This predicate determines how long it is possible to access the datum within the system. It may not be kept at all, or kept indefinitely, or retained only for a short period of time (which could be specified by the policy).

Although the original intention of the taxonomy was to provide a way to frame different forms of privacy research in a similar light, there are other applications for which the taxonomy can be used, allowing us to compare the privacy preference of the data provider with the policies of the organization collecting and using the data, which Barker *et al.* style *the house*.

3.5.1 Using the privacy taxonomy to determine privacy violations

Building on the idea that the privacy taxonomy developed by Barker *et al.* uses a geometric space to order the four identified privacy predicates, Banerjee *et al.*[Ban+11] further extend the model to quantify privacy violations.

Figure 3.2: A simplified schematic of privacy violations by Banerjee *et al.*[Ban+11]



As shown in Figure 3.2, the authors observe that for an arbitrary dimension (such as S_j in the figure), we can compare the privacy policy set by the house with the privacy preferences of the provider for a given data item. The value of this predicate which is closer to the origin point of the taxonomy is considered more private. Therefore, a privacy violation exists when the stated preferences of the data provider lies closer to the origin than that of the house, for a particular predicate. No privacy violation exists when the stated preferences of the data provider for a given predicate are less restrictive (and therefore lie further away from the origin, as shown by Dimension S_i in Figure 3.2) than that of the house.

This idea is further extended, for multiple dimensions, incorporating three of the four predicates of the privacy taxonomy: Visibility, Granularity, and Retention. Because values in the remaining dimension (Purpose) are evidently not comparable in this fashion, this technique treats Purpose as a grouping variable for the data items. Because this comparison only tests for the existence of privacy violations, the comparison of privacy policy to privacy preferences in each dimension can be further scaled by adjusting with *sensitivity values* provided by both the data provider and the house, for the predicate, the field, and the specific value of the datum.

While the primary contribution of providing a means to identify privacy violations and determining their severity is important, another key idea generated by this work is the consideration that individual data providers may have personal privacy preferences which may be entirely distinct from those of their peers.

3.5.2 Moving away from the privacy taxonomy

While the privacy taxonomy provides a way to quickly situate pre-existing privacy frameworks and models, as we move to attempting to use the taxonomy against specific privacy policies and individual data provider preferences, issues with the robustness of the taxonomy arise. We examine issues with two predicates in particular.

Purpose is central to the concept of privacy, as captured by legislation (such as by the OECD principles[OEC80] and PIPEDA[Can00]). Indeed, of all of the predicates of the taxonomy, purpose is the one which is explicitly specified by legislation (such as by the OECD or in PIPEDA). Its importance requires that care be taken in its specification, but each organization has its own conventions for the definition and specification of purpose, for which no common standard has been adopted.

More nuanced approaches expose the largest flaw with purpose in this taxonomy, which is the criticism that"purpose is a string." This statement summarizes the majority of definitions of purpose in privacy, in which purpose is difficult to be specified, much less put in order. How can an information management system determine which of these purposes should take priority? What happens should purposes conflict, for example, if the same individual is a clinician, a patient, a donor, and a research subject all at once? How do we assign specific purposes to only specific personnel (for example, patients should not receive information for the purposes of being employed at the hospital, finance staff should not receive patient data for the purposes of providing treatment). Are the assigned purposes adequately descriptive? How much detail is enough?

We also point out that retention can be treated simply but in practice, requires more

complexity to meet the needs of legislation. Retention rules in legislation map back to the principle of collection limitation, which is the idea that data is only kept in the system "as long as it is needed." Because different organizations apply different criteria to satisfy this principle, complex records management workflows have been developed to meet the retention requirements of each organization. As cloud computing becomes commonplace, meaning that data is increasingly stored offsite and housed with one or more third parties, retention requirements can be extremely onerous and cumbersome.

Thus, customer data exists in systems in a wide range of states, such as being stored "raw" (i.e. as microdata, unaltered), aggregated and joined with other data, common for data warehousing and business intelligence applications, and in archival formats, such as on other media such as hardcopies and tape drives, often inaccessible because they are stored separately, offline, and often offsite. Even as data retention regimes require that records be deleted or destroyed, we find that there are disparate outcomes for such data, as there is a difference between the simple deletion of records, and ensuring that they are truly destroyed so that no trace is retrievable. In an era where databases are commonly distributed on servers across the Internet, in locations which may be unknown even to the database's administrators, the destruction of electronic records is not guaranteed.

Although the privacy taxonomy is useful for the abstraction of privacy for use in data management, there should be a way to simplify potential scenarios so that complex organizations and enterprises do not have to create abstractions for everything at a single point in time. The next chapter discusses context as a different approach to privacy.

Chapter 4

Privacy and Context

The approaches to privacy we have covered thus far have been developed for organizations which have typically collected private data belonging to numerous data providers, and will be stored for later processing. The organization is required by legislation to protect the privacy of those data providers by restricting the later processing to only accepted purposes and accepted processors, in such a way that the original data providers cannot be re-identified based upon the publication of a redacted form of their data.

However, the privacy protection afforded by these approaches will not be sufficient in the future. Data providers find that their data is now being collected continuously. Consider some of these examples which are now part of our everyday life: services tracking their web use, corporations making sense of our purchasing patterns, systems engineered to harvest information about our location and application use from our mobile devices, vehicles and home appliances, governments, institutions, and private organizations who are using increasingly more automated surveillance systems to ensure the safety of citizens, and social media services seeking to leverage the networks of our friends and family for profit. The prevalence of organizations who reuse and trade in data is growing as novel methods of processing data become increasingly common. These methods can use facts and features, which were once considered innocuous, to reveal aspects of everyday life which many data providers may wish

to keep private. Data providers are also increasingly aware of privacy issues, and often find that the privacy polices created by organizations to protect them are insufficient to meet their privacy expectations.

The plethora of information systems with which we all interact on a regular basis are also not necessarily adversarial in nature. It is often convenient, or necessary, to provide these systems with enough data so that we can make use of them. Consumers would like their financial services providers to be able to detect fraudulent attempts to access credit cards or bank accounts in real time. When negotiating congested traffic, it is useful to have apps which map alternate routes (and the likelihood that using them will save time) in real time. Health providers should have as much pertinent information as possible to provide timely, safe and effective treatments. Devices which can sense when spaces are occupied promise to provide considerable savings in the operating costs of a residence and increased efficiencies for those with concerns about the environment.

Thinking about *context* helps us to design systems which can use the correct information for the correct application at the correct point in time. While it may not be appropriate for a financial services provider to take into account the credit rating of a parent when handling the records of a given data provider, that same parent's history of certain conditions may be highly useful and relevant for a health provider. While knowledge of a data provider's current location may help a traffic app to plan an alternate route, it may not be appropriate or useful for that data provider's current location to be provided to a health care provider.

For a financial services provider, knowledge of the data provider's current location may be useful but not appropriate in all circumstances - for example, should the data provider's current location at a casino be taken into account the next time their credit rating is assessed? What if that data provider's location was regularly recorded at a casino, several times a week for months or years? What if the financial services provider also had information that the data provider was also employed by that casino, or by an organization which regularly did business with the casino?

We envision a privacy-preserving system which will protect the privacy of individual data providers, but also ensure that systems have enough data to guarantee utility, that will be flexible enough to handle the way data is processed in the present and future world, and to negotiate the nuance which is required as we interact with information systems in different ways. To accomplish this, this chapter interrogates the relationship between privacy and context in information systems. We first introduce the idea of context and how it has been historically used in computer science, especially in data privacy.

4.1 Context

The term context can have a variety of meanings depending on the discipline.

People have an innate understanding of social contexts. Children are often mindful about the difference between an "inside voice" and an "outside voice." Families often share intimacies and confidences with each other, but not when non-family members are present. Co-workers may leave the office for lunch together, but decline to discuss work matters while out of the office.

4.1.1 Context outside of Computer Science

Context serves as a key to understanding most works of literature. Critics and scholars generally hold that understanding the context in which each work was set, but also the context in which each work is written, leads to a richer understanding of the work. For example, a contemporaneous account of an event set in the past should likely be understood in a different way than an account of the same event written decades or centuries later.

Context is obviously critical in law. The process of establishing what is just in both civil and criminal matters relies not only on the alleged offense, but also upon the circumstances, the situation, and the relationships between those involved.

4.1.2 Context in Computer Science

There are a number of ways in which computer science refers to context. Commonly, operating systems define a context as the minimum amount of information about a task so that it can be suspended and resumed later. Also informally, some may refer to context as the current state of a program.

A significant body of research in context-aware computing has developed from research in human-computer interaction, particularly in pervasive computing. The goals of context-aware computing is to enable users to acquire more function or useful information from their environment based on systems which use sensors to automatically detect environmental factors, as described by Dey *et al*[DAS01]. For example, most mapping applications will return the nearest locations when a user searches for "gas station" or "restaurant." This is in contrast to most privacy-preserving systems, which aim to use systems to detect environmental (and potentially social) factors to restrict access to private data as accurately as possible.

Systems which leverage this body of research are often concerned with privacy. However, the focus is often upon usability, the feasibility of a potential hardware implementation and integration with pre-existing technologies. There is relatively little examination (as in Jiang *et al.*[JL02] or Hull *et al.*[Hul+04]) as to what privacy means, how privacy should be defined, and what protections should be afforded the system's users.

4.2 Contextual Integrity

Much of the early work to formalize privacy for implementation in information systems led to the recognition that a much more nuanced approach to privacy is necessary. As everyday consumers negotiate the provision of a massive amount of data about their ordinary lives, approaches which allow and recognize for change are increasingly important.

The framework of *contextual integrity* developed by Nissenbaum[Nis04; Nis09] provides a different approach from more traditional methods based upon access control. As origi-

nally proposed, contextual integrity illustrated how the flow of information from one actor to another can trigger concerns over privacy. In its most simple terms, contextual integrity suggests that the act of data transmission is always situated in a social context containing specific, and often unwritten, informational norms. Nissenbaum's framework describes these informational norms as a composition of three elements engaged in the act of data transmission: the actors, which include the subject of the data; the sender and recipient of the the data; the type of data; and the "transmission principles," which constrain the transmission in some way.

Since the framework relies on a normative view of privacy, privacy is viewed from a personal lens as well as an organizational one. Nissenbaum draws a distinction between explicit or injunctive norms, and norms which are implicit, or descriptive. Explicit norms are those which have been defined as a rule, as a published privacy policy or legislation. Implicit norms are those which derive from the interplay of personal views and experience of the actors involved in the context, comprising privacy behaviours which are not necessarily well-defined, but which actors "know." When these informational norms remain intact, then the contextual integrity of the transmission remains intact. When the norms are broken in some way, then the contextual integrity of the transmission is violated.

For example, members of a family may share pictures with each other on a social networking service to keep family members and friends updated. When that social networking service takes those pictures, the re-purposing of which is allowable by the service's privacy policies, to market a local childcare service, family members (and those connected to them) may have a sense of unease. This sense of unease may grow even greater should the service attempt to use those pictures in another way - for example, to promote a local political campaign. If a family member uses the advertised local childcare service and finds it to be poor, then that experience shared amongst the family members may lead to an even greater sense that the pictures in question have been misused.

In this example, the family (and others connected to them) share norms about how data

(family pictures) will be used by each other, and by extension, the social networking service. When the photos are used for another use (promotion of a local childcare service), some members of the family may be disturbed by the re-purposing of these pictures for another use. Other members of the family may see this use more pragmatically, understanding that the social networking service must be able to pay their costs by offering effective advertising services, or finding the information about the advertised services to be useful. The social networking service, by the practice of re-purposing family pictures, illustrates that its informational norms differ from that of their users.

However, as these pictures are reused for purposes more distant from the everyday lives of this family, it is quite likely that more of the individuals who were the originators or intended audience for those pictures would object to how the pictures have been re-purposed. From this behaviour, we might infer that the users of the social networking service share certain norms about what personal pictures might be re-purposed for - that there is a distinction between a business whose clientele might involve those featured by the pictures, and a political campaign which might not be as concerned with those featured in the pictures. The social networking service, in the meantime, may view the reuse of these photos for promoting a political campaign as no different than the reuse of the photos for promoting a local business.

4.3 Adoption of Contextual Integrity by Computer Science

Contextual integrity has been influential amongst computer scientists working on privacy, especially after references to context have begun appearing in legislation and privacy guidelines such as the US Consumer Privacy Bill of Rights[The12].

In addition to data privacy and attempts to produce a formalized, operational implementation of contextual integrity, the framework has been influential in human-computer

interaction, especially in the area of usable privacy[Bar12].

4.3.1 Formalizing Contextual Integrity

While largely a framework so that legal scholars could think about privacy in nuanced terms, an attempt to formalize a part of this framework was made by Barth *et al.*[Bar+06]. In this work, Barth *et al.* developed a formal language they called CI, using linear temporal logic (LTL) to express privacy policies which fit the framework. In the paper, CI is evaluated by using it to encode carefully selected examples taken from pre-existing privacy law, and by comparing to other policy-based models for privacy protection, including traditional role-based access control, the eXtensible Access Control Markup Language (XACML[1]), and P3P[2]. The comparison serves to demonstrate the additional features implemented by CI which were not available (and likely not considered to be required) for the other models.

One of the co-authors offers another variation on the formalization of contextual integrity by proposing PrivacyLFP[Dat+11]. Instead of LTL, this formalization builds a formal language based upon first-order logic, attempting to incorporate audit mechanisms as a means for supporting additional language features which would otherwise be completely unenforceable.

While LTL is convenient for formalizing parts of contextual integrity, providing certain guarantees for the complexity for checking properties such as consistency and compliance, there are some shortcomings. The authors do not comment on the difficulty of encoding privacy policies as logical rules, although they are the only individuals to have done so, in their paper. Contextual integrity is also notable for insisting upon norms rather than policies, as often it is when unwritten norms are in violation that providers feel that their privacy has been threatened. CI assumes that all norms have been recorded as a privacy policy or pieces of privacy legislation.

[1] `http://docs.oasis-open.org/xacml/3.0/errata01/os/xacml-3.0-core-spec-errata01-os-complete.pdf`

[2] `https://www.w3.org/P3P/`

4.3.2 Surveying Contextual Integrity in Computer Science

A literature survey written by Benthall, Gürses, and Nissenbaum[BGN17] documents a wide range of work within computer science which has followed up on Nissenbaum's original writings on contextual integrity. The survey observes some notable differences between Nissenbaum's original work, and the research that builds on it, with some useful categorizations of many of the subsequent work which has adopted contextual integrity as a key idea.

Research produced by computer scientists is primarily concerned with how a framework can be operationalized. Concepts within that framework must be represented in a way which can be implemented on computers, and so an ideal starting point is to consider only explicit norms, which have already been defined by an organization or written into legislation. Not only is it easier to locate exemplars of explicit norms, but they tend to be subject to less change than implicit norms, and have been carefully written to minimize confusion, conflict, and contradiction.

Secondly, the definition of context has become more malleable than hoped for by Nissenbaum and her collaborators. Benthall *et al.* highlight specifically how some works have avoided embedding the concept of norms in their definition of context; and for contextual integrity this is an inappropriate definition. An example of this, as observed by Bernthall *et al.* is in a study conducted by Wijesekera *et al.*[Wij+15], which attempts to encapsulate contextual integrity to mean that "privacy violations occur when personal information is used in ways that defy users' expectations." Their work builds around studying the permissions end users grant applications installed on a smartphone based upon their environmental context: time, location, connectivity. While this is one aspect of contextual integrity, it does not reflect the entirety of the framework.

The authors of this review also found that Nissenbaum's conception of a context as "social sphere" is often conflated with definitions of context from elsewhere in computer science, which stress measurable physical variables rather than social norms and mores which are frequently in flux, according to the individuals who are presently sharing in a particular

context.

In the next chapter, I take these findings into account, by proposing a solution which offers room for both context as understood by the humanities and social sciences, and as typically defined by technical disciplines. While technical disciplines assume that context can be defined by ambient factors which can be measured by sensors, application use, and other tangible inputs, the humanities and the social sciences tend to assume that context is not necessarily a computable quantity. Creating a conceptual framework for an operational system which can relate context to both is a critical step as we consider how to implement systems which can use context for privacy protection.

Chapter 5

A Framework for Contextual Privacy

As a framework grounded in legal scholarship and thought, there are some aspects of contextual integrity which have eluded computer scientists since its first proposal. We present this framework as a way to address issues identified with contextual integrity in the previous chapter. The goal is to resolve the different definitions of context which have evolved from technical practice as well as legal and social thought and provide a means for both to contribute to privacy protection.

The central concept in both contextual integrity and in this framework is that of *context*. Nissenbaum[Nis04] identifies context as a social sphere, which encompasses the data transmission between the actors (who might include the data provider, the data collector, or the data processor) in accordance with norms which are understood within that sphere. Over time, computer science researchers have shifted this understanding of context to a construct which is more familiar to them, using data about the surrounding environment such as location, time and connectivity, as well as data about the applications being used to collect or access data, and the permissions assigned to these applications by the end user. While inferences may be made about what social sphere is relevant based upon environmental data (for example, if somebody is standing in a private residence or near a public space), these data points collected about the surrounding environment may not necessarily define a social

sphere completely (the private residence may contain a business such as a piano teacher's studio, or an accountant's home-based office, whereas the public space may be occupied in part by those who do not live in the area).

Another significant aspect of contextual integrity often misunderstood by computer science researchers is that of the *norm*. Nissenbaum distinguishes between explicit norms such as those recorded by privacy policies and legislation, and implicit norms, which are formed by those within a given context.

Explicit norms tend to be handled well by modern data governance, as legislation or policy whose rules, enforcement and penalties are clearly defined. However, there is a gap between the legislation described in Chapter 2 of this thesis and the technologies and techniques described in Chapter 3 which have not yet been satisfactorily made operational and still depends upon the judgement of human operators and practitioners. This gap remains a source of significant effort by those engaged in privacy research.

By contrast, implicit norms are often not evident to information systems until they have been violated. Implicit norms might include individual privacy preferences, as well as norms implicitly held by a group. These might be well-established ones, such as religious groups who are uncomfortable with having their images recorded for posterity, or temporary ones, such as a teenager who is uncomfortable with a friend's social media post because they feel awkward about their physical appearance that day. For computer scientists, implicit norms are problematic, with obvious implications with respect to decidability and completeness, because implicit norms can differ with each individual active in a system, in accordance with their beliefs, history, and present decision-making process.

However, many implicit norms are guided by explicit norms, or grouped together as representative for a set of individuals, who have some shared beliefs or history. Thus, privacy-protecting systems may be able to enforce implicit norms, if they are provided with a framework that allows them to create a template for norms which apply to subsets of individuals who do not share the same role as officially identified within the system. Because these

individuals do not share the same role, traditional access control approaches are not entirely appropriate for the enforcement (much less the discovery) of implicit norms. A typical solution for such a problem might be to assign a "temporary" or "interim" role which would have a different set of access control rules, but this can create considerable complexity as the number of rules generated in the system and attached to each individual grows, and as it becomes necessary to track which of these temporary roles is currently in play.

Computer scientists intend techniques such as access control to automate the application of rules to requests for access to data. However, in the fields of law and compliance, the preferred standard when applying norms (as encoded by legislation or policy) is not to immediately prescribe a remedy, but instead to consider each situation separately, with its own individual circumstances, so that an appropriate remedy may be identified. Consequently, implementing a system which respects contextual integrity should not be done by automatically applying rules without considering specific situational characteristics. Thus, this framework proposes a multi-layered approach so that abstract conceptions of contexts, norms, and the roles interacting with both can be joined with real-time observations so that rules regarding the use of personal data are generated and applied appropriately.

5.1 An overview of the framework

Our framework creates an operational version of contextual integrity by defining context as a multi-layered entity, shown in Figure 5.1, composed of the *abstraction*, the *operation*, and the *observation*. This allows us to define context as an abstract entity without reference to specific individuals and circumstances but grounded in a social sphere with normative values, an intermediate space in which operational rules can be created and enforced by a data management system, and as a collection of environmental conditions which surround a data collection or processing operation.

Each of the framework's three layers has three aspects, which are shown in Figure 5.2.

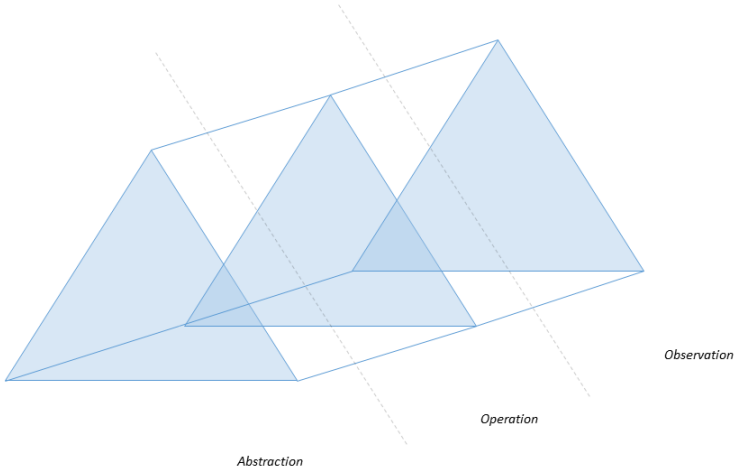


Figure 5.1: An overview of the framework

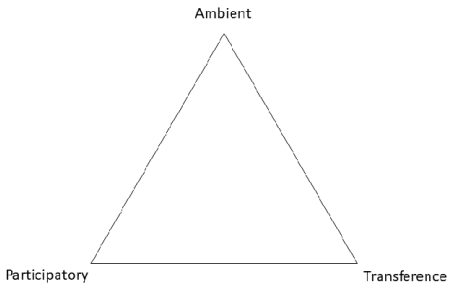


Figure 5.2: Aspects of each layer of the framework

Ambient elements describe the surroundings (the "where"), whether it is Nissenbaum's social sphere which defines the domain in which we are operating, and the vocabulary we use in it, or the immediate physical environment of the data collection. *Transference* elements describe the data operation and whether the context permits it (the "why" and the "how"). *Participatory* elements describe the elements undertaking the actions, or having actions performed on their data (the "who").

	Ambient	Participatory	Transference
Abstraction	Social sphere	Role	Norm
Operation	Situation	Actor	Rule
Observation	Environment	Person	Practice

Table 5.1: Terminology used by the framework

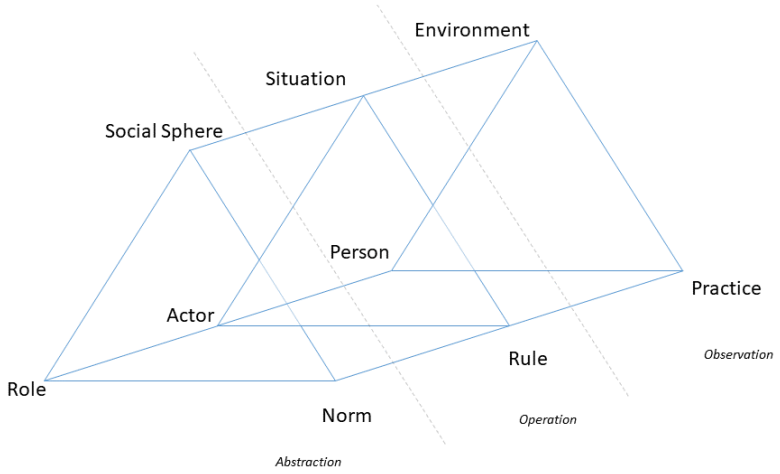


Figure 5.3: The contextual privacy framework

Thus, for each layer, we have three terms, each representing a different aspect, resulting in nine labels for different parts of the framework. To keep these labels organized in the following sections, Table 5.1 lays out terms for each layer and aspect in the framework. We apply these labels to the framework, with the result shown in Figure 5.3.

First, we consider the *social sphere* to be an abstract representation, attached to other abstractions: *role*, which represents those participating in the operation, and *norms*, which govern the operation.

We also have a series of real-time observations which occur at the time of the operation. These might be observations about who is currently authenticated to access the system, a time stamp, a location, or information about the client accessing the server, such as an IP, MAC address, or operating system. They may be observations about what is being accessed or processed, for example, in the form of a database query.

These observations are grouped into three concepts: the *practice*, the *person*, and the *environment*.

A middle layer (the *operation*) is required to fuse together real-time observations with the abstract concepts. This layer makes specific the concepts in the abstract layer and applies to them real-time observations. Thus, a social sphere and environment combine to create a

situation, a person and a role define an *actor*, and norms and practices combine as *rules*.

The following sections describe each element of the framework in detail, especially as they relate to each other.

5.1.1 Abstraction: Social Sphere, Role, Norm

We envision that the abstraction layer would be used by those responsible for defining norms, whether implicit or explicit. This would involve legal experts, policymakers, and others involved in the authoring and enforcement of legislation, policy and custom.

Social Sphere *A social sphere* is the aggregation of roles, norms, data, and the social domain to which they are attached. Social spheres restrict which roles can exist and the vocabulary in which norms can be expressed. For example, in a clinical treatment setting, journalists do not exist, nor should any identifying personal data be disclosed to satisfy the public interest.

Role Each *role* represents an archetype of parties who interact within a social sphere. They typically share similar information-handling norms.

Norm *Norms* specify general behaviours for how different roles may interact with data within a specific social sphere. Norms can be stated with some parameters which are unknown and uncertain. These parameters are furnished by other parts of the framework when more specific knowledge about how data will be collected or processed is known.

For example, when addressing the question of the appropriateness of retrieving information about a student's whereabouts on campus after hours, we might consider:

- Who is the specific student? (Would we expect them to be on campus at that time, whether they live in residence or are registered to a course which meets after hours? Has the student reached the age of majority, or do privacy constraints applying to minors also apply?)

- Where on campus is the student? (There may be differing norms depending on the precise location. Is it a public area, a restricted laboratory with valuable equipment, or the student's own residence room or building?)

- Who else is at the location? (There may be differing norms depending on the immediate social environment. Is it a group of students studying together or working on a class project, a meeting of a student organization of a religious or political nature, or a casual group of friends who spend part of the evening at a campus bar?)

- Do any other conditions exist which must be taken into account? (For example, is there an emergency associated with a time and location on campus?)

Few of these pieces of information may be known for a given social sphere, meaning that norms that form part of an abstraction may not be enforceable without additional information.

For example, when considering a post-secondary educational institution after business hours, there might exist the following norms about using the location of a student on campus for additional data processing:

- The approval of a parent or guardian must be obtained to access location information of a student under the age of majority.

- The exact whereabouts and identity of all other students in non-restricted areas may be accessed by authorized staff members for approved purposes.

- The exact whereabouts of students in areas restricted to the general public may be accessed by authorized staff members for approved purposes, except for students who are in residence buildings which they are registered to live in, in which case, only the information that they were in their building may be accessed.

- The exact location of students in residence buildings which they are registered to live in is only available for purposes of safety, security or law enforcement, or in circumstances

for which the university might otherwise be held liable.

- If students are in locations booked by specific student organizations for a meeting or event, their exact location may only be released to parties outside of the university with their consent, or in the event of an emergency. Otherwise, only the information that the student was on campus may be accessed.

- Specific students have consented to have some information about their locations released via social media, to only members of the university they have already connected with using pre-specified platforms.

5.1.2 Observation: Environment, Person, Practice

We envision the observation as the most tangible and measurable layer. Those who are most involved with this layer are those (likely technicians and developers) working with sensors (perhaps Internet of Things devices), recording data such as locations, IPs, and MAC addresses, and who handle requests for access to data (such as database queries).

Environment There is real-time data which computer scientists have traditionally considered the "context" for their systems: physical factors such as the location, time, and place, hardware characteristics such as the IP or MAC address of a given device, hostnames, information about the system such as the software (including operating systems) being installed. This framework considers such ambient data to form the *environment*.

Person The *person* in this framework is the individual performing an operation about data, or the subject of the data. At this level, we may have a collection of facts which can be resolved to identify an individual. This may include data that can authoritatively be linked to that individual, such as an authenticated system access, the use of a physical token such as an electronic key or identification, but may also include data which is less certain,

such as the MAC address of a mobile device, the license plates of a vehicle registered to an individual, or surveillance footage of a person entering a building.

Practice *Practices* are the requests the system receives for data: storing and managing data for data collectors, or processing and access data for data processors. Before they are received and processed by the system, some practices may not conform to the rules to be formed by the system at this point.

5.1.3 Operation: Situation, Actor, Rule

The operation layer serves to connect the real-time elements of the operation with the abstraction, whose elements may not be as concretely defined. It thus becomes a layer which can provide an instance of the abstraction as an *one-time* entity. We envision the operation as being the responsibility of developers and analysts who are responsible for privacy on an operational basis.

Situation A *situation* is a real-time instantiation of a social sphere. It represents the realization of a social sphere in a specific time and place (using data identified as part of the observation layer's environment) with identifiable individuals. Situations limit the number of possibilities inherent in a social sphere so that rules can be generated, if necessary, and processed in a realistic period of time.

Actor *Actors* link the abstraction of a role with an identifiable individual (resolved from the data gathered by the observation layer about a person). Actors may simply instantiate a given role, but may also inhabit several roles, or have additional data associated with them which may affect which rules apply to them, and how. For example, they may belong to sub-groups that require different privacy constraints compared to the general population of individuals inhabiting particular roles, or share relationships with other actors which may be considered in composing rules.

For example, a TA may be required to avoid grading activities for a given student, if that student is a close relative, even if ordinarily, they may grade the work of all of their students, and all other TAs may grade the work of that particular student. Professors may not be permitted to evaluate the research of another professor if they have recently collaborated on a project, although they may be permitted to evaluate the work of professors with whom they do not share such a conflict of interest.

Rule *Rules* are a real-time representation of a norm, with any parameters in the norm which were unknown or uncertain made specific and actionable. Rules may be more constrained than the norm they are based on, due to elements which were not part of the social sphere, but exist in the real-time situation. Consider the following norm:

> "Professors should not share information about any student with anybody who is not part of the university."

This means that rules based upon this norm may be:

A norm without modification Professors should not share information about any student with anybody who is not part of the university.

A norm with specifics A specific professor, such as **Prof. Zee** should not share **the grades** of **Student A** with an actor who is not part of the university.

A norm with obligations Professors should not share the grades of Student A with an actor who is not part of the university, and **Student A should be notified of such requests**.

A norm with conditions Professors should not share the grades of Student A with an actor who is not part of the university, **unless the actor represents an organization evaluating a scholarship application submitted by Student A, and Student A provides written consent by email**.

5.1.4 Putting the framework together

To demonstrate at a high level how these pieces work together, we return to the setting of the university after hours. In this setting, we offer some commentary about a pair of scenarios.

A student in their own dorm room, with a visiting classmate When at home, we would not expect any organization to monitor what we are doing, or where exactly in the house we were. We are, after all, *home*. This would not be an unreasonable expectation for a student in a residence building to share. Even though the institution has the capability to monitor the student however they choose, a student-resident should not expect the institution to be monitoring their every action in residence: sleeping, studying, socializing, doing laundry and other chores, or using the washroom. There may be agreed-upon limits to this freedom, such as a bandwidth cap to Internet use, which means that the volume of the resident's internet use might be monitored, or an online booking system for laundry facilities, to maximize the use of the facilities and minimize conflict with other residents. However, the student might reasonably expect to be free of surveillance while in their own assigned space.

The same time and location can comprise a different social sphere (and situation) for the non-resident student visiting their friend in residence. They are not at home here. For reasons of liability and to ensure all occupants of the building are safe, the institution might insist upon different rules for visitors. There might be a requirement for the visitor to check in with an attendant, to provide the name and room number of the student hosting them, to wear a visitor pass (presumably with some sort of tracking mechanism such as a RFID chip), or to observe specified time limits for their stay. There may be different types of rules for different visitors. The student's parents, as older adults who may need to stay for several days, may be offered more degrees of freedom than one of the student's friends, who is only planning on socializing for the evening.

Here, both resident and visitor are in the same physical space at the same time, but differing methods may be used to observe where they are. For the resident, an electronic

room key or their ID card may be used to register when they are in the building, but for the visitor, it is likely that some other means to determine their presence may be used, since not all visitors may be directly affiliated with the institution.

In this case, the abstraction offers the institution the most guidance as to what the appropriate rules and practices should be. Because the norms between the social sphere of "at home in residence" and "visiting somebody in residence" are quite different, the corresponding operations must differ as well.

A student works in a restricted lab, but leaves briefly for a coffee Now consider a research building and a student within who is performing lab work of a delicate nature using specialized equipment. In this case, it is reasonable for the institution to monitor where that student is, and at what time, for purposes of safety and liability, as long as they are engaged in work at that location. However, should that student leave the lab to buy a coffee somewhere on campus, it would also be reasonable for that student to expect more anonymity while on campus but outside the lab.

It should also be noted that the same token (the student's ID card) could be used to provide access to the lab but also used to purchase their coffee, as a debit or meal card, and to log both activities. However, there should be an expectation that even when the student uses the same physcial token for both activities, the student's research supervisor(s) should not be able to track the number of coffees bought with the card (or where, if it was a location close to the lab or further away), nor should the operators of the university's food kiosks be able to track where the student went with their coffee.

The lab and the university outside of that lab clearly comprise different environments, situations, and social spheres, even if they might rely on similar observations of the student's environment (the use of the ID card). This results in different rules around the practices of handling data based upon the ID card.

In the next section, we consider more in-depth examples, situated in settings which any

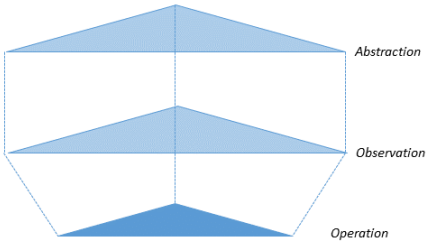


Figure 5.4: A reconfiguration of the framework to help visualize the flow of operations in this section's extended examples

member of the public might encounter. These will be used to illustrate how the framework would operate from the moment a system must determine whether data should be collected to the moment when data is collected and processed. We will return to the post-secondary setting in subsequent chapters, to delve more deeply into how context and privacy plays a role in this arena.

5.2 The Framework in Action

In this section, we use an extended example to illustrate how we might create an operational framework for contextual privacy. To help visualize the flow of operations in these examples, consider a reconfiguration of the framework, so that the abstraction and observation layers are overlaid against each other to constrain the operation layer. This differs from how we have presented the framework in previous sections but will help the reader to better visualize the flow of operations.

We will use three different scenarios to examine how the framework could be operationalized. As we proceed, we will track how the system might record each element of information in the framework.

The first example examines the framework's operation layer by using a research project tracking the proximity of individuals to sensors in a shared office, which is a semi-private environment that might use these sensors to pilot new security systems or to ensure that staff

are taking breaks mandated by company policy. The second example examines the observation layer, by using a research project which places similar sensors in a public environment, such as may be done in a "smart city" (and is being proposed in small portion by projects such as Sidewalk Labs' Toronto Waterfront development[1]) to monitor usage of public infrastructure, such as bicycle pathways, parks, or streetscapes. The third scenario examines the framework's abstraction layer, by using a bar that has a kiosk to scan identification to ensure that customers are legally able to be served alcohol.

5.2.1 Example 1: Andrea and Charles at work

Consider an office in which sensors have been placed strategically to monitor when employees are in proximity to them. Employees have been informed about these sensors, and have been given the option to opt in to the data collection or to opt out. Those who have opted into the data collection may have their ID badges registered to interact with sensors placed around their office or to use a wearable device such as a smartwatch to interact with the sensors instead.

Sensors located by a common area of the office (such as a lounge area) register two "pings" around 11 AM on a Monday. One is a proximity badge registered to Andrea Barton and the other is a wearable device registered to Charles Drake.

At this point, we can assign certain facts to the framework. We begin with an observation:

[1] `https://sidewalktoronto.ca/`

Observation

Environment	Radius of 1 m around Sensor X Radius of 1 m around Sensor Y 11 AM Monday March 11
Person	Proximity badge registered to Andrea Barton Wearable device registered to Charles Drake
Practice	{not yet known}

These pings can be resolved with high confidence to be related to specific individuals, because they have been pre-registered by the system, and social norms about the workplace dictate that employees do not misuse the credentials of their colleagues, and that the personal property of employees remains their property throughout the workday. We can begin to populate the operation with relevant information:

Operation

Situation	Shared work area (office lounge) at 11 AM on Monday March 11
Actor(s)	Andrea Barton Charles Drake (has specified time-related restrictions on data collection) Researchers (identity not specified)
Rule	{not yet known}

To finish populating the operation with relevant rules, we must determine how to generate them. To do this, we refer to the relevant abstraction:

Abstraction

Sphere	Shared work area
Role	Personnel, Researchers
Norm	Well-marked sensors will record the location of those sharing the office when they come within 5 m of each sensor, unless consent has been withdrawn by the employee in advance.

The study has been explained to all personnel who may be affected, and the opportunity to withdraw (by not registering personal wearable devices or proximity badges) has been provided in advance. Charles has chosen to register a personal device to the system but opted to only participate in the study between 10 AM and 3 PM. He has chosen these restrictions to avoid having researchers (and any others who subsequently may have access to the study data) make inferences about when he has chosen to arrive at the office late or leave the office early.

Generated rules are added to the operation:

Operation

Situation	Shared work area (office lounge) at 11 AM on Monday March 11
Actor(s)	Andrea Barton Charles Drake (has specified time-related restrictions on data collection) Researchers (identity not specified)
Rule	**Andrea may have data about her proximity to well-marked sensors collected** **Charles may have data about his proximity to well-marked sensors collected between the hours of 10 AM and 3 PM**

When the system registers events from Andrea and Charles, it must generate rules to determine whether to record these events. Once they have been generated (and depending

upon the exact implementation of the system), it would be possible for the system to store some form of these rules to avoid having to re-generate the rules (and the corresponding practices which are being generated and are shown below) at a later point in time. Once these rules have been generated and stored, the system would not need to refer to the abstraction again unless something else necessitated it.

The operation can be used to produce the appropriate practices for the observation:

Observation

Environment	Radius of 1 m around Sensor X Radius of 1 m around Sensor Y 11 AM Monday March 11
Person	Proximity badge registered to Andrea Barton Wearable device registered to Charles Drake
Practice	**Record sensor ping from proximity badge registered to Andrea Barton** **Record sensor ping from wearable device registered to Charles Drake**

5.2.2 Example 2: Yolanda and Zachary in the park

In the case where data collection is happening in public areas, privacy protection must begin at the point of data collection.

We modify the previous example so that now, the system is collecting information on visitor traffic inside a public park. Infrared sensors have registered the presence of two individuals, one larger and one smaller. Two visitors, Yolanda and her young son, Zachary, walk through the park past these sensors.

Observation

Environment	Park 11 AM Area Surrounding Sensor A
Person	Sensor detects "adult-sized" person Sensor detects "child-sized" person
Practice	{not yet known}

This example suggests one possible implementation for data collection in this scenario.

The sensors (or the device controlling them) may hold this information temporarily while another part of the system (nearby digital signage, such as a tablet or a kiosk) attempts to negotiate consent with the detected entities.

The system may determine that the smaller individual is too small (*i.e.* too young) to have information about them legally collected. Smaller individuals are often children, and it can be commonly accepted that children may not be able to provide consent for research, especially in an environment where they may not fully understand their participation in a study.

Sensor data cannot be resolved to identify specific individuals without further information, but it may be resolved to identify general classes of park users. Thus (unlike our previous example), the identified actors in this scenario are not specific individuals.

Operation

Situation	An adult and a child walk through the park at 11 AM on Monday
Actor(s)	Adult (no consent yet) Child (unable to consent)
Rule	{not yet known}

Since the range of possible situations in a public area may be quite large, it is more

likely for the system to generate rules in real time based upon agreed-upon norms, compared to the previous example where the system only needs to consider a very small number of individuals, all of whom are identifiable.

Abstraction

Sphere	Public Park
Role	Researchers (data collectors) Members of public (data providers)
Norm	Members of public must be notified and provide informed consent to have their data collected by an infrared sensor by researchers. Children are unable to provide consent and information about them must not be collected by automatic systems.

Operation

Situation	An adult and a child walk through the park at 11 AM on Monday
Actor(s)	Adult (no consent yet) Child (unable to consent)
Rule	**Collection of data about adults is permitted if they provide consent** **No data about children may be collected**

The interactive part of the system may attempt to negotiate consent with Yolanda herself, by displaying or playing a message about the study and providing some form of a call to action so that Yolanda actively consents (pressing a button, filling out a form, tapping her mobile device to trigger some app).

Once some threshold has been passed (the individuals pass the detection range of the sensors, or they exceed a time limit for Yolanda to respond to the call to action), the sensors must not store the fact of Yolanda or Zachary's presence in the park, or if Yolanda has

consented, then the sensors may pass on the fact of Yolanda's presence in the park to the system.

Zachary's presence in the park may not be recorded since Zachary is unable to provide informed consent, and there is no guarantee that Yolanda is a parent or legal guardian who is able to provide consent for Zachary without further data collection, which may not be made available.

Observation

Environment	Park 11 AM Area Surrounding Sensor A
Person	Sensor detects "adult-sized" person Sensor detects "child-sized" person
Practice	**Collect data about adult if adult has responded positively to the provided call to action to provide consent within the set time limit** **Do not collect data about adult if the adult has not responded positively to the provided call to action** **Do not collect data about child**

5.2.3 Example 3: Morris at a public establishment

We examine the case where environmentally, conditions are the same, but in which the social sphere may be significantly different.

Morris is scheduled to attend an event being held at a local bar and arrives early. At the door, he is asked to scan his identification at a kiosk so that the establishment can determine that Morris may be lawfully served alcohol. He is told that the kiosk will only indicate to the staff member present: whether Morris' appearance matches that of the identification's

photograph; whether Morris is old enough to be served alcohol or not; and the likelihood that the provided identification is authentic. He is also told that the kiosk will not record any data captured from Morris' identification.

Morris complies and can order a drink from the bar while he waits for his event to begin.

We might record Morris' entry as the following observation:

Observation

Environment	At the door of a bar, Wednesday, 6:45 PM
Person	A prospective patron at the kiosk presenting ID
Practice	Any photographs take of the prospective patron may not be stored; Do not store any details captured from the patron's identification

As a straightforward, non-exceptional interaction, the observation is combined with the following abstraction:

Abstraction

Sphere	Public Establishment
Role	Customers, Staff
Norm	Customers must be over the legal drinking age to be served alcohol

This creates the following operation:

Operation

Situation	Morris enters the bar
Actor(s)	Morris (verified via identification)
Rule	The system may show that Morris may be served alcohol if he shows valid identification, *i.e.* if: • The kiosk's photograph of person matches the photograph on the provided ID • The kiosk's scan of ID indicates that person of age to be served alcohol • The likelihood that provided identification is fraudulent is below a specified threshold

When Morris has finished his drink, he remembers that he left something in his vehicle and leaves to retrieve it.

When he returns, the event (a kickoff event for a political campaign) has begun. To his annoyance (having already done this once in the evening), the event organizers ask for his identification so they can scan it (using a similar system - or even the same system - as that used by the bar) to determine that he may be served alcohol and to record his name for their records. This time when he is permitted to enter, he is issued a wristband which indicates to staff at the bar that he may be served alcohol.

When he objects, because he has already produced his identification once in the evening, he is told that the event organizers have been asked to perform the task of checking identification because the bar does not wish to have any information about the identities of those attending a political event.

From the viewpoint of Morris and the kiosk, the observation is identical to what has

recently transpired on Morris' previous entry, with the only difference being the time of entry:

Observation

Environment	At the door of a bar, Wednesday, **7:05 PM**
Person	A prospective patron at the kiosk presenting ID
Practice	Any photographs take of the prospective patron may not be stored; Do not store any details captured from the patron's identification

The operation is again similar:

Operation

Situation	Morris enters the bar during a scheduled political event
Actor(s)	Morris (verified via identification)
Rule	The system may show that Morris may be served alcohol if he shows valid identification, *i.e.* if: • The kiosk's photograph of person matches the photograph on the provided ID • The kiosk's scan of ID indicates that person of age to be served alcohol • The likelihood that provided identification is fraudulent is below a specified threshold

However, the main difference actually lies in the abstraction, with the introduction of a new role (Organizers), and new norms surrounding that role. In addition, the staff at the pub have new norms, specifically that they do not deal with any information about the identity of attendees of the event.

Abstraction

Sphere	**Political Event at a Bar**
Role	Attendees, Organizers, Staff
Norm	Attendees must be over the legal drinking age to be served alcohol **Only event organizers may record the identity of attendees**

5.3 The framework in summary

This framework expands context into three separate layers: an *abstraction* which encompasses the social sphere and its supporting roles and norms, an *operation* which encompasses a situation and its supporting actors and rules, and an *observation* which encompasses an environment and its supporting persons and practices.

The design of this framework allows practitioners considerable flexibility in the actual implementation of a system which respects contextual privacy. The three in-depth examples in the previous section each demonstrate how a separate layer of the framework contributes to the enforcement of contextual privacy. Andrea and Charles have specified that the system observe different rules at the operation layer, sensors register observations of Yolanda and Zachary in very different ways; Morris finds himself navigating different social spheres by leaving the bar and returning to it some time later.

In each scenario, some aspect of contextual privacy is being considered or enforced, but in very different ways from the other scenarios. Furthermore, data can be collected and processed in different ways and at different levels of granularity (or not at all), using different kinds of hardware and servers.

The framework also presents a means to represent and handle dynamic norms and observations, as time progresses and conditions change. Change is an important element of contextual privacy since many contributing elements will differ over time and result in new contexts. This can be due to new information, a different set of individuals in the location,

the reaction of individuals to an organization's data handling practices, or changes in the beliefs of the individuals themselves. For example, in our first example, it is possible that Charles will ask to have the system's rules about him altered because he has been assigned a new shift, from 4 PM to midnight. A group of adults (or adult-sized humans) may also enter the park where Yolanda and Zachary are, none of whom would necessitate being automatically excluded from the project's data collection. Morris may become bored with the political event, and decide to return to the bar on the following night, when it is once again only a bar. The framework provides a means to deal with all of these changes without unwieldy modifications.

This framework expands the concept of a context into three related layers to handle a greater range of conditions which contribute to context. This provides the practitioner with a great deal of flexibility in implementation and operation, and the field with a richer approach to handling context.

Chapter 6

Privacy and North American Higher Education

The different contexts in which privacy becomes a major factor is critical for how individuals will live their future lives. A common vision is that of the "smart city," a modern development in which governments and other organizations (such as utility providers, major technology vendors, and other companies) are able to measure a myriad of facets of everyday life, and process that data nearly instantaneously so that the city can react to changing conditions in real time.

To understand privacy concerns in this broader domain, we seek to understand privacy concerns in post-secondary institutions, in which data collection is as comprehensive (if not more so) as in a smart city, but in which information flows to a single major organization. Because the majority of the data subjects who exist in this domain also are at a time of their lives in which behaviour, beliefs, and preferences can be highly dynamic, context-aware privacy is particularly applicable in this domain.

Consider the modern post-secondary institution, particularly a larger institution focused on research as well as teaching. Our interest here is in the wide range of activities supported on a daily basis by these institutions, and how data collection is an essential component of

many of these activities. While these institutions differ greatly in size, geographical setting, financial resources, and goals, there are a number of attributes which they often have in common.

We can consider these institutions in the frame of only academic or research-based activities, which themselves can generate large amounts of sensitive data. Students must be enrolled, registered, made identifiable, and awarded grades or other credentials while they complete their studies. Academics are employed, trained, and paid while they conduct their teaching or research-based activities. As categorized by Prinsloo[Pri17], the types of decision-making generated by the data gathered in the course of these activities is significant, especially as analytics practices in higher education become increasingly automated.

However, the modern post-secondary institution also collects data outside of their core missions of research, teaching, and learning which can capture a large portion of student life spent away from the classroom. There exist on modern campuses a considerable amount of physical infrastructure to contain educational or research activities, and the institution may be responsible for providing basic services such as utilities, internet access, and security for this infrastructure. For faculty and staff as much as students, the modern post-secondary institution may find it convenient to provide amenities such as food and retail services, libraries, parking and transportation services, fitness centres, sporting facilities and theatres, health services and others.

A subset of students (small or large depending on the institution in question) and perhaps even faculty or staff may also live in residence spaces operated by the institution, and as a result, may partake of an even greater number of services offered by the institution, including kitchen, cleaning and laundry facilities.

The information flows present in a modern post-secondary institution are considerable. To simplify the analysis of how data is used in this environment, we consider activities which collect data from students separately from those which then process this data.

In the first part of this chapter, we consider how data from and about typical under-

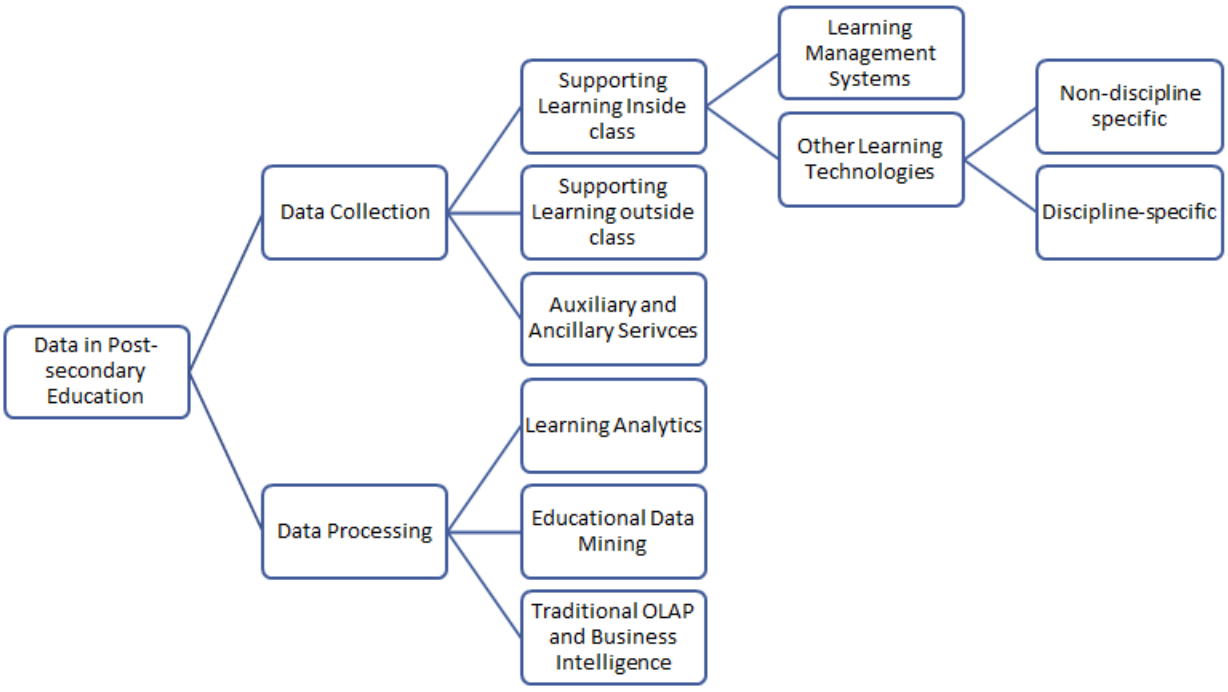


Figure 6.1: A taxonomy of data collection and use in post-secondary institutions, as presented in the chapter

graduate students might be collected. The second part of this chapter connects the collected data to methods (and motives) for data processing by a typical post-secondary institution.

6.1 Data collection

Considerable amounts of data are collected from students while they attend a post-secondary institution. To help in the analysis of this data collection, we consider the typical daily experience of a student, starting with the core activities which bring them to campus.

We might think about data being collected from students in terms of relationships between the student, educator and institution, in an administrative, traditional form as follows:

Information flows between the student and institution Institutions require a certain amount of information from students before students can attend class. So students can be registered, they must provide a name, email and physical mailing addresses so the institution

can contact them. To verify that students can be registered, other types of identification might be required. To establish that students are qualified to enter their chosen program, they must submit transcripts, copies of diplomas, or reference letters. To ensure that students can pay for the services they will receive while attending class, they must submit financial information to the institution.

Information flows between the institution and educator Once institutions have enrolled students, this information must be communicated to educators. Educators do not receive all the data the institution has about their students, but receive enough to be able to identify students for the purposes of teaching and assessment.

Information flows between the student and educator In the classroom, relatively little personal data flows between the student and educator (as well as support from others, such as teaching assistants, guest lecturers or administrative staff). Most of it is transmitted on an *ad hoc* basis (such as in class discussions) and is not recorded. There are some instances in which a student's personal information may be recorded. Students may indicate a preference for a name which was not officially recorded by the institution, or have to share personal details about themselves to their classmates. When students complete assessments such as assignments, tests, or exams, they must attach some information (usually a name, email address, or a student identification number assigned by the institution) to it to identify their work.

While the examination of these three relationships serve as a good starting point to think about student privacy in post-secondary institutions, data collection and processing involves a more complex ecosystem between many different relationships and multiple players. Each student has individual privacy preferences, stemming from their own identity, history, and beliefs. They work with multiple educators, each of whom brings a unique perspective to how they teach and relate to both students and the institution. While it is convenient to think of the institution as a single entity, in most cases, the institution is composed of a multitude

of smaller units and organizations with different data needs and differing philosophies which underlie how they view the student data that may be available for collection and processing.

Thus, in the modern post-secondary institution, there are many more aspects of the learning experience which are mediated by technology, which collects a significant amount of data about learners. Because technology is increasingly responsible for recording and collecting information about students on behalf of the institution, we will examine some of the common types of technology in this space.

These are of course, to learn, thereby to work towards the completion of their chosen program and earning a degree, diploma, or other credentials. However, students undertake other activities on campus which support their learning experiences, whether it is completing work assigned outside of class, or participating in co-curricular activities (volunteering, participating in student clubs or sports teams), or engaging in research.

There are also activities students undertake which may have very little to do with their learning experience. These can range from the mundane, such as purchasing food and sundries, to the exceptional, such as receiving counselling or health care services.

This section explores these three groups of activities in-depth so that we can see in what ways private data is collected from students. In each set of activities, we examine the kinds of data being collected from students and to whom it flows.

We consider first the activities which surround the primary reason students attend post-secondary institutions: to learn.

6.1.1 In the classroom and as learning experiences

Besides the kinds of data institutions collect directly from students for the purposes of administration, other kinds of data are also collected from students in the course of their learning.

Automating data collection: technology in the classroom

It may be a common assumption that technology plays a relatively minor role in these interactions, but the truth is that these interactions are increasingly mediated by technology. Many institutions have adopted the use of a *learning management system* or LMS, typically a web-based application package intended to help schools and academic staff organize the large volume of information which must be communicated to and from students. Most commonly, these systems serve as repositories for lecture notes, specifications and prompts for assignments and projects.

Other technologies are being used to mediate learning experiences for students. A common medium for students to communicate with faculty as well as staff is email. While many students continue to use accounts provided by external providers, post-secondary institutions typically provide students with one or more email accounts, and most encourage (if not require) the use of these accounts. Many students also use private email accounts, generally provided by an outside vendor, to communicate with faculty as well as classmates. While the expectation for students is that the messages within these accounts are not accessible to any except administrators and themselves, the content of email servers has been used for research in the past[MM92] and could be accessed by others, including the providers of the cloud-based email services used both on campus and off.

Increasingly common in the last decade, audience response systems allow faculty members to pose questions to students, and receive real-time data about student responses to the question. These systems originally required the use of specialized devices to register responses, necessitating their purchase by either the student or institution, making these systems more expensive and less accessible to use. Increasingly, vendors are taking advantage of student-provided devices such as smartphones, tablets, and laptop computers, and only require the use of an application that is installed on the student's device, or use the built-in mobile web browser to access specific webpages.

The use of the smartphone opens up the range of data which can be collected by these

technologies. Not only do such systems collect student responses, but they may also collect the number of incorrect guesses, the amount of time which it took students to respond, or even where students are physically located. For example, a common use of such systems is to track student attendance in class. With the use of smartphones, audience response applications can forgo student input in gathering data about attendance, instead relying directly upon location data from student-owned devices.

The largest collector of student data, however, is the learning management system, due to their prevalence and the degree to which many faculty members now rely upon them.

The Learning Management System

A primary area of focus for those seeking to leverage the data being collected by post-secondaries often focus on the LMS, and thus we consider them in this section.

In addition to serving as a repository for materials and grades, these systems may provide the ability for educators to interact with students. They may provide assessments such as quizzes (which can then be administered automatically by the system), infrastructure so that students can submit documents for assessment by educators, and provide discussion boards and other communication tools to encourage engagement with the course material. These systems may be integrated with application suites such as Microsoft Office or Google Drive for better ease of use. They are also used to help educators keep grades organized and updated, so that students can receive feedback quickly, and track their standing in their classes over the semester.

Very commonly, student access to these sorts of systems are tightly tracked by the application. Data can be collected about how long students spend engaging with course material in the LMS, when they are accessing such course material, and what materials they accessed (for example, whether they have viewed specific pieces of content, such as a set of lecture notes). Certainly, whenever the system is being used by the student as part of a learning experience (such as posting a message to the discussion board, attempting a quiz, or submit-

ting content), data is being collected by the LMS, and often offered back to the instructor or institution as part of an analytics component.

Other technologies as part of learning

Students may also be asked to use technology to complete learning assessments (such as labs, assignments, or papers) on their own. This may include publishing material to their own websites or blogs, developing an e-portfolio as part of a reflective process for a course, a semester, or for their program, or engaging with course material using social media or services such as Slack™(or other equivalents, such as Mattermost™)[1], which are intended to help groups of individuals collaborate more effectively. Students are also often required to work together in teams for part of their program, and often will collaborate using services hosted in the cloud, such as with Google Documents, Microsoft Office 365, or Dropbox. Often these services are provided to the students by the post-secondary institution, but where no appropriate services exist (or where students are accustomed to using other services), students often seek out these services on their own initiative.

Students will also be asked to master different technologies in the course of their studies, as part of their chosen discipline. For example, students in the technical disciplines may be asked to write code or develop larger software applications, which may be developed on the student's own devices, on workstations belonging to the university, and/or hosted on university servers (or those provided by an external vendor), especially when published to repositories such as github or gitlab. Students in other disciplines may be asked to analyze and visualize data, create and edit video or still graphics, or use other applications (as well as tangible technologies, such as modern fabrication equipment) as part of their studies.

As advancements in technology become more commonplace, we see proposals for other types of technology - for example, video capture and hosting systems so that lectures can be recorded and put online for those unable to physically access the classroom, or to pro-

[1]Slack is a trademark of Slack Inc. Mattermost is a trademark of Mattermost Inc.

vide content for online courses offered by some institutions. While these proposals have been more commonplace in the elementary and secondary school systems, there have also been suggestions that schools employ facial recognition or biometric technologies to better facilitate the recording of attendance, and to ensure that students are engaged in classes.

With each tool adopted by the institution, there is a risk that data about learners, their activity, and their learning is collected for use by vendors or other third parties. Often, this data collection may be negligible, a vendor may reassure the institution about their data use, or the institution may require that each vendor sign a contract with specific conditions about data processing and collection. However, once this data leaves the institution's immediate control, particularly if exported to servers located in other jurisdictions, then the institution is unable to guarantee that the privacy of their learners remains intact, or that the data is used in ways that are consistent with the values of the institution.

6.1.2 Supporting the student's learning experience

As we move away from the core interactions between the student and the institution, data collection becomes less transparent to the student. Much of the data about students that is being collected consists of "ambient" data which is collected from the student's environment, instead of directly from them.

To make sense of the other kinds of data collection which occur, we focus first on the institutional functions which exist to directly support learning. These are intended to help the student maximize their academic experience at the institution. These might include other services students use, such as those designed to help them build additional transferable skills via experiential learning or volunteer work, to strengthen their academic skills, or help students enhance their program through non-credit workshops or work experiences. There may be systems which help learners with their finances by administering awards, scholarships, loans, or other financial aid.

These services often use additional technologies (or may use the institution's learning

management system) to register and track student activity, but often they must also interface with other forms of data collection. For example, a potential employer or a host institution for an exchange program must also collect data about the learner, who is also their employee or student. Often, volunteer work requires criminal checks or registration with external organizations and systems. The kinds of extra-curricular work students often engage in are typically tracked as part of their record somewhere in the institution (for example, for awards), sometimes with micro-credentialling systems which can post evidence of their work to external sites (such as social media profiles).

There are also the functions which students may interact with to complete coursework while not actively sitting in lecture or other scheduled course times. While this section does not completely describe every way in which student-provided data is collected in the process of supporting learning experiences, we describe several of the most common forms.

It is now difficult to conceive of a post-secondary institution in which internet connectivity does not exist. The networks which support the connectivity of students, and predominantly via WiFi, now carry some of the most essential data about students: their web-browsing activities, searches, social media activity, and streaming behaviour.

While the majority of this data is restricted from common use by legislation and institutional policy, this data nonetheless is logged by the numerous collection points which exist in networks: routers, gateways, servers, and others. For knowledgeable practitioners, this data offers a gateway to understand exactly what activities students undertake online while on campus. Off-campus behaviour can also be monitored to a certain extent, as students interface with university-provided portals and sites to access coursework, plan their academic programs, and search for supporting material. Some campus-provided websites or systems may ask the student to download apps or accept cookies on their private devices, which can then be used to track browsing behaviour.

In addition to wireless access, most institutions provide other technical resources for students which can be accessed outside of course time. Despite the pervasiveness of smartphones

and laptops, computer labs continue to be a mainstay for many students. Here, either access to the lab itself, or logins to a specific workstation may be tracked.

One element of most post-secondary institutions where students often search for supporting material is in the library and the numerous resources it hosts. Students often access libraries for references for course work and research, and often access library materials and resources to satisfy their personal curiosity and interests. Many may rely on workstations provided by libraries to complete homework, assignments, and other types of assessments. Libraries also offer space and facilities which can be accessed or booked by students for periods of time so that they can study in private, work with a group of peers, or use specialized equipment such as projectors, recording studios, and others.

6.1.3 Auxiliary services

Most post-secondary institutions offer a wide range of services beyond those which support learning experiences - given the amount of time spent on campus by most students, institutions must provide for necessities beyond learning and the classroom. These provide a great deal of range in terms of collection of student data.

We first consider facilities (usually operated by the institution, but quite possibly contracted to private providers) such as fitness and recreation facilities. These facilities often track usage by requiring that students check-in in some way - either in person with staff members, or automatically via kiosks or turnstiles which generally require proof of enrollment (the student's ID card is a common requirement) for entry. Some facilities also impose additional requirements (such as specific training or proof of competence) for access - to meet such requirements, students may have to register online or using their ID cards for specific orientations, classes or workshops.

ID cards are used to access other resources on-campus. They are often required to access physical spaces, such as restricted offices, laboratories (especially outside of business hours), residence buildings, to sign into individual machines in computer labs or at libraries, and to

print or photocopy materials. Many institutions provide students with the ability to use their ID card like a debit card at on-campus locations (to pay for printouts, photocopies, vending machine purchases, or purchases at point-of-sale terminals at physical vendors for meals, coffee, or goods), and thus a record of their financial transactions, if not their purchases is easily recorded when students use this form of purchase.

Students (and all others moving through a post-secondary campus) also experience surveillance by cameras in a widespread manner. Smaller cameras with a more compact installation footprint, with higher-resolution capabilities mean that cameras can be reporting on a much more pervasive basis and in more detail than in the past. Many providers of camera surveillance also offer some degree of data processing, with facial recognition, and systems which can be trained to flag potentially problematic incidents.

Student residences are also a potentially rich source of data. Doors to buildings, to individual living spaces and shared amenities are secured using electronic locks, typically using cards or electronic keys (which can be very quickly re-keyed). In newer sustainable buildings, smart thermostats and occupancy sensors are used to save energy by only heating or cooling living spaces which are occupied. Of course, residents and guests are on campus longer than students living off-campus, and thus their Internet usage can be tracked far more completely than those who leave campus. Even within residences, access to amenities such as laundry machines or communal kitchens may be controlled by card or electronic locks, and thus can be tracked.

Those who have not been equipped with kitchen facilities generally buy their food on-campus, and institutions which are now cost-conscious track how food is bought and consumed as obsessively as any food service establishment off-campus. Those who have access to cooking facilities may be encouraged to buy their groceries on-campus, at a convenience or grocery store by the institution or one of their contractors.

For those not living on-campus, parking services are increasingly another point through which data can be gathered. For enforcement and business purposes, it is common to record

information about the vehicle (such as the license plate number, but potentially also information about the vehicle such as the make, model or colour) as well as financial information from those accessing these services. As these services become increasingly automated through the use of pay stations, smartphone apps, and cameras, the recording of such information becomes correspondingly commonplace.

Large universities offer a varied set of services beyond what this section describes. This can include highly sensitive services such as health clinics, counselling, and spiritual support. To enumerate every such service provided would be well outside the scope of this thesis. However, it should be clear that the services are myriad and the opportunities for data collection manifold.

6.1.4 A day in the life is not every day in the life

A discussion about the effects of data collection and use on undergraduate life is not complete without considering the interplay between the young adult (most typical of undergraduate students), universities and other post-secondary institutions, and the process of identity formation. As summarized by Klimstra *et al.*[Kli+10], researchers in psychology have posited many models for the identity formation of adolescents as they enter adulthood. For many in late adolescence, encountering significant educational experiences (such as university or college) can be a catalyst for the identity formation process for many students[ARK00; BK00].

From a privacy perspective, practitioners should consider how students undergo identity formation. Often, personal milestones can deeply affect individuals, to the extent that deeply engrained attitudes (such as those concerning privacy, or which influence how privately certain pieces of information should be kept) may be radically changed. These milestones may require a gradual adjustment of the individual's thought processes or attitudes, but they may also occur suddenly, as part of a personal crisis or epiphany. As a result, practitioners should certainly consider the possibility that the norms (especially implicit ones) which apply

to a particular student may be changed over their time at the institution.

Included in her own survey of influential scholarship and thought on privacy[Nis09] , Nissenbaum makes particular note of the link between individual privacy and identity formation, observing that identity formation cannot occur without a sense of privacy, allowing one the ability to self-govern how they identify to others and behave with them.

There is likewise an effect on privacy as a result of identity formation. Slade and Prinsloo[SP13] point out that students undergo a substantial amount of change during their academic careers, using the term "temporal dynamic construct" to describe the identity and performance of students. There is a temptation by those working with student data to view the cumulative record of all data about an individual student to be representative of that student, but only a subset of this data may be highly relevant to the student at any given point in time. Many undergraduate students are at a point in their lives when they transition from being adolescents to fully adult, and thus are accumulating life experiences which can significantly alter their point of view and identity. The nature of such change is also not a constant. Because students are experimenting with aspects of their personal identity, some changes will be lasting, and some are merely temporary.

A system designed to protect the privacy of undergraduates must take into account the changing identities of this constituency, and the effect this has on the norms which must be observed. A contextual approach which allows flexibility in terms of the preferences of the individual data provider (such as a student) is necessary in this domain.

6.2 Data Processing

Data, once collected, is not useful until it can be analyzed for insights which lead to actions. Despite having collected a mass of information, most post-secondary institutions do not yet use all the data they gather for analysis. Constraints exist in the form of privacy legislation, policies, institutional structure and practice, as well as the personal beliefs and norms by

staff, adminstrators and faculty who may be involved in such analysis.

There exist, broadly, two main divisions in the use of data by post-secondary institutions. One division is focused on the types of planning and decision-making required of enterprises of the size of the University, in which business intelligence and traditional OLAP (Online Analytical Processing) systems are used and are generally well-understood. These types of decisions are vital to the operations of the University. For instance, vending machines must be stocked, custodial and maintenance staff must have their shifts scheduled, and there must be sufficient numbers of classes and sections scheduled for the number of students enrolled and for the space available.

The other division is focused on improving the experience of the learner, with the goal of optimizing the number of students who successfully complete their program at the institution. Within this grouping, we can again divide the types of data processing which are being employed into two.

The term "learning analytics"[SB12] is ambiguous, because different groups of researchers and practitioners have applied this term to different types of analytics within the educational context. There is a general movement towards using this term to indicate analytics specifically directed towards improving the experience of individual students in the context of helping them to complete their programs successfully, by informing educators (and potentially students) so they may be able to appropriately act given the insights generated by these analytics.

The term "educational data mining" also exists in the literature[SB12]. Generally, this refers to the use of data mining techniques extant throughout the field of data management to approach the same kind of data as used by learning analytics, but with the goal of institutional improvement rather than improving the experience of individual students. Initiatives which might benefit from this approach could include the establishment of student-focused programs, course or curriculum redesign, or adjustments to how the institution is staffed.

It is important to note that there is significant overlap in the types of activities de-

scribed by each term, and that for many in the literature[SB12], the terms are often used interchangeably, depending on the specific background of the researcher.

We discuss some ways in which data is being applied to solve institutional problems in the following section.

6.2.1 Improving a student's learning experiences

Increasingly, LMS vendors are aware of the potential benefits the student data they have gathered can bring to their clients. Many of the leading products now feature support for reporting and visualization of the data housed within the LMS.

Using LMS data: early-alert systems

One of the immediate uses of data within the LMS is to provide "early-alert" systems which signal when students are experiencing difficulties in a course. Different implementations of such systems rely on different sources of data, but the most common include data drawn directly from the LMS, such as student grades in the course, either from individual assessments, or interim grades, but may also include behavourial data such as how often a student accesses course material within the LMS, and how long the student engages with such material.

The objective of early-alert systems is to provide a sense of student performance in the course before other traditional indicators (such as final grades) would indicate that a student is having difficulty completing a course. The hope is that appropriate interventions (for example, more individualized assistance from the professor/instructor, placing the student in additional workshops, providing extra tutoring or counselling), can be put into place to assist students before they drop a course, or fail to complete the semester, or even to help them improve their performance in a course.

The appeal of these systems is that the data being used to indicate when students are experiencing difficulties in their coursework is already being collected by the institution, but

is processed somewhat minimally to provide services that traditionally would be difficult to put in place in a timely fashion for students. While this data is usually available to faculty, they rarely have sufficient time to scan through the relevant data for each of their students and determine on a case by case basis which students require additional resources.

Course Signals, a program developed at Purdue University to identify at-risk students[AP12], is one of the best known examples in which data from different on-campus sources is used to develop a simple predictive model about student performance in selected courses. The data used includes the student's academic history, demographic data such as age and residency, their current standing in their current courses, as well as the time spent in a LMS working on the materials posted for the course. The results of this model are integrated into a simple dashboard inside the LMS (in this case, Blackboard) to warn students about impending issues in their academic career, and to help faculty identify students who may require further (and timely) intervention to maintain an adequate standing in the course. Faculty members can then enact an appropriate intervention for the student (which might include notifying the student of the algorithm's results, contacting or meeting with them, or referring them to other on-campus resources) in the hopes of improving the student's in-class performance.

The program was piloted with great fanfare in the educational technology community and has since been commercialized, with similar dashboards being developed by major LMS vendors. However, there continues to be some doubt that the program has any impact on student retention, and particularly that the statistical models used by Course Signals to predict student performance can adequately capture the full nuance of students experiencing difficulties in a course. The discussion surrounding the efficacy of Course Signals is summarized by Ferguson and Clow[FC17].

Course Signals is not the only program making use of data already collected by the LMS to affect student outcomes. For example, at the University of Calgary, the Thrive Priority Support network[2] uses data from the University's LMS to determine whether students may

[2] `http://www.ucalgary.ca/ssc/faculty/thrive` - accessed August 18 2018

be experiencing difficulties with their semester so that at-risk students can be contacted and offered appropriate supports. While no additional data other than that which can be accessed from the LMS is used, faculty members are also invited to submit reports directly to the system, on a purely voluntary basis.

Learning Analytics

Large vendors in this space are increasingly promoting "learning analytics" as a means of providing additional value to their customer. This encourages institutions to treat the data stored by learning management systems as a type of data warehouse, with primary motivations typically being to aid student performance and retention.

In addition to the early-alert systems, learning analytics have been applied to helping students select courses. Future growth may see the incorporation of biometric technologies, such as facial tracking and the use of Internet of Things devices such as fitness trackers, or smart speakers deployed in residence rooms[3].

The adoption of learning analytics by the vendors of learning management systems and the post-secondary institutions who patronize them is the early phase of a future in which more post-secondary learning happens online or mostly online (what those in the field refer to as "blended" learning). Ultimately, the goal is to automate the use of learning analytics so that each learner is presented with an individualized learning experience even when enrolled in the same course as dozens (if not hundreds) of others. This is commonly referred to as "personalized learning."

As interest grows in the application of analytics to a wide range of domains, many researchers and practitioners will also turn to learning analytics as a means by which institutions can improve student outcomes. Learning analytics have the potential to deeply affect how students experience education at a post-secondary level. However, the adoption of ethical approaches, in particular, a circumspect approach to student privacy, lags well

[3] `https://www.insidehighered.com/news/2018/08/22/meet-new-kid-campus-alexa`

behind the pursuit of new technologies. A comprehensive review of the field by Viberg *et al.*[Vib+18] observes that only 18% of the 252 papers included in the review, all of which were published between 2012 and 2018, mention ethics or privacy.

6.2.2 Fulfilling institutional goals

Post-secondary institutions, like other organizations and enterprises, must also attend to the business side of their operations. To do this, most rely upon business intelligence and reporting systems which have been commercially viable for decades. These systems are used for budgeting, decision-making, and evaluating the progress and success of specific initiatives and programs.

6.2.3 Post-secondary institutions in a Big Data Future

The state of the art in processing data at post-secondary institutions is at this point relatively benign, especially in comparison to the types of analytics in use outside of post-secondary campuses. Student privacy will be more strongly at risk as data in the two categories which have been broadly defined in this section begin to be merged with each other. The major risk is that the contexts in which individual pieces of data have been collected and processed lose their integrity - when improving the educational experience of an individual student becomes difficult to distinguish from the purpose of improving the institution.

Data which is traditionally considered to come from the business side of the university, such as data about food purchases, does not impinge on individual student privacy until it begins to be applied to individual students, perhaps to encourage more "acceptable" lifestyle choices. Data recorded about a student's individual performance threatens the privacy of that student when the institution sees the retention of students and adherence to rigid standards for student performance as a higher priority than the educational experience and priorities of the individual student.

Post-secondary institutions will also be approached by vendors who see their accumula-

tion of student data as a vast resource to be mined. Vendors will propose that data about student registration and retention be combined with demographics to produce recommendation systems for not just one, but many institutions, so that students find the course selection less daunting. Other vendors may seek out samples of student work to populate systems to detect plagiarism. Other vendors may seek information about internet usage and occupancy to best place the products they offer for sale on campus.

Vendors are already working with insitutions to change how students access institutional information by collecting even more student data so machine learning algorithms can better predict which informational sources students are seeking. Amazon has contracted with Saint Louis University to place an Echo Dot, one of their proprietary smart speakers, in each student-occupied residence room on campus[Sai18]. This has been done with the stated purpose of bringing students current information about student life and campus events. While posted privacy statements are consistent with Amazon's privacy policies, and clearly state that school administrators cannot access queries made to the devices[Sai], it is unclear to what use the vendor would put the data generated by these devices in each residence room, which might be any voice recording registered by an internet-connected listening device. Any privacy concerns which would apply to these devices in general would also apply to this specific deployment. In another pilot project, IBM and York University are collecting student input to develop a virtual assistant which can direct students to appropriate information sources[Yor19].

We already see early-alert programs such as Course Signals being commercialized and offered as a generic LMS add-on. While many of these proposed innovations may improve student experience in some way, they generally propose taking student data out of the context in which it was meant to be used and processing it in a vastly different way. Even before the question of whether any of these proposed innovations can deliver on their promise of an easier, better future for students, the question of how the use of student data affects student privacy should be a critical consideration.

Some universities may also consider approaching vendors to acquire data about student behaviour off-campus for incorporation into their own analytics efforts. For example, social media feeds are easily available, as are tools to derive insights from them. Literature about student engagement, notably by Tinto[Tin87] (as summarized by Pistilli *et al.*[PWC14]) often highlights the importance of both formal (such as clubs, workshops, or other extra-curricular programs) and informal opportunities for students to interact with their peers. In general, the more interaction a student has in these social spheres, the greater the level of engagement they experience with their institution, and the more likely they are to successfully complete their chosen program. Monitoring which groups and individuals a student interacts with on social media would be a measurable indicator of these forms of student engagement.

Mining social media for student information also has other potential uses. As demonstrated by Chen *et al.*[CVM14], social media may also offer some insight about student attitudes towards their post-secondary experience, although much work would remain to determine whether students felt institutional use of the content they generate and post to social media was appropriate, and whether any insights gathered using social media as a data source could be meaningful and actionable.

Chapter 7

Case studies in Data Collection and Use in Post-Secondary Institutions

This chapter studies selected use cases to understand how we might use the contextual framework presented in Chapter 4 to better understand data collection and use in post-secondary institutions. The goal of these studies is not to immediately generate implementable or operational rules, but rather to explore how even seemingly simple tasks might be re-evaluated using the framework. This may allow an organization to not only meet the standard set by policy or legislation, but work with each data provider to provide individualized privacy protection.

The previous chapter gives many examples of data collection and use at post-secondary institutions. We select two categories, one which is not usually exposed to learners and educators, and another with which learners and educators frequently interact.

We begin with an examination of typical applications of business intelligence at post-secondary institutions. Although this is a broad category, we have identified three different domains, with different priorities, interactions with students and other community members, and outcomes. We follow this with a more in-depth examination of use cases typical of students and educators using a learning management system to perform commonplace tasks,

such as accessing course materials.

7.1 Business Intelligence

Analytics play a key role in the operations of post-secondary institutions, as with large organizations in every field. As with many complex organizations, analytics have numerous consumers at post-secondary institutions, with a wide range of goals and motivations. Post-secondary institutions use the insights generated by analytics to demonstrate that they are effectively fulfilling their missions in research, teaching, and learning to numerous stakeholder groups which includes funding and granting bodies such as governments and research foundations, current and prospective donors, and prospective students.

Analytics also help to inform the day-to-day operations of the institution, such as capacity planning for information technologies infrastructure, and supply chain management for ancillary services such as dining and residences.

As with many large organizations, post-secondary institutions face many issues in the thoughtful and intentional application of analytics techniques to facilitate their operations. Williamson[Wil18] describes the significant amount of work being done globally to formalize the use of big data and analytics in the post-secondary sector, whether it is building physical infrastructure, standardizing architectures, defining standards so that different analytics platforms can be made interoperable between institutions, or sharing data sets between institutions at the national or international level.

Williamson and others[BT16] observe a number of common issues amongst post-secondary institutions in their ability to leverage data for business intelligence and other analytics. A typical challenge is being able to identify and locate all credible sources of data within the institution. Once this is done, it is equally challenging to ensure that data sources can be made semantically equivalent. Another issue is in determining what the data can tell the institution.

In this section, we examine specific uses of business intelligence at a post-secondary institution, captured within the framework of contextual privacy. For each specific use, we focus our analysis on a separate layer of the framework. We examine alumni relations using the abstraction layer, institutional analysis using the operation layer, and fitness and recreational facilities using the observation layer.

7.1.1 Alumni relations

Most institutions view developing relationships with their alumni to be a priority. Alumni are an important part of the campus community, as potential donors, connections for current students, guest speakers for events, volunteers, and as stewards of the institution's reputation. In short, alumni present a valuable resource for the institution if relationships with individual alum can be built and carefully maintained.

However, alumni relations are also fraught with privacy issues, as former students grow apart in time, distance, and engagement with the institutions which provided them with their education. With the growth of electronic communications, it is easier for the institution to maintain contact with alumni via email (and indeed, many institutions provide newly-minted alumni with long-term email addresses as a benefit for both the graduating student and to facilitate future communications), but this does not necessarily imply that the institution can maintain a relationship with each alumnus. An alumnus may consider contact from the institutution in the far future to be an inappropriate use of their private data. This may be exacerbated especially if the institution cannot guarantee that this data is up-to-date, and sends correspondence to an old mailing address, fails to keep up with new titles (consider the acquisition of an advanced degree) or name changes (as in the case of marriages or gender transitions).

Using the contextual privacy framework, the institution might see their use of contact information from an alumnus as follows, in the abstraction layer:

Abstraction (institution)

Social Sphere	Fund development
Role(s)	staff member, alumnus
Norms	Staff may contact the alumnus about making a donation using the provided information, if the contact information meets specific quality standards

Quality standards would be introduced in this scenario in the operation layer. For each alumnus, the institution may choose whether to contact them depending on a variety of criteria, including whether the campaign would be relevant to the alumnus, how recently the alumnus had already made a donation, or engaged with the institution in some way (for example, registering and attending an event). Other data such as whether the alumnus had ever replied to (or received) previous communications may also be taken into account. The institution might integrate external sources such as publicly available phone directories to verify the accuracy of their information.

However, for some alumni, they may see receiving contact from their university or college in a different way. For them, the abstraction layer might look more like the following:

Abstraction (alumnus)

Social Sphere	Marketing
Role(s)	staff member, alumnus
Norms	The institution may contact the alumnus about relevant initiatives

In the operation layer, for those alumni, what they view as relevant initiatives may differ from the institution's view. They may view these initiatives only as those involving their field of employment (which may differ from their chosen program of study as a student), involving their close friends while they were students at the institution, or where they may receive a direct benefit.

What begins as a small divergence as viewed by each participant at the abstraction layer rapidly becomes a greater difference at the operation layer, unless those working in alumni relations can achieve better alignment between these two views of the abstraction layer. To maintain the sense that the post-secondary institution is respectful of the privacy of their alumni, it is necessary to capture the norms which are representative of their alumni and appropriately accommodate those norms. For example, they may ask the alumni to voluntarily provide updates about their current employment and location via a regular survey, or ask alumni with whom they maintain a good relationship to broker an introduction with alumni who have maintained more distance with the institution.

7.1.2 Institutional analysis

Institutional analysis is a standard form of analytics at universities and colleges. Those who perform institutional analysis are responsible for delivering the metrics demanded of the institution by the bodies responsible for the governance and funding of the institution. These include governments which provide funding and mandates for the institution, bodies which provide accreditation to the institution as a whole, or to specific programs, as well as to the public at large (in the case of publicly institutions).

While this function (which can be housed in varying parts of an institution, depending on its organization and mission) typically handles student data in the aggregate, sourced from other units such as the registrar (for information about enrolment and grades), human resources (for information about staff), and planning or operations (for information about the use of spaces on campus such as classrooms or conference rooms), it also deals with data *about* students.

We might think of a typical document published by such a unit. Consider a report that releases a distribution of student grades[1], showing across all faculties of the institution, how

[1]similar to the data published at this page, last accessed on December 20 2019: `https://public.tableau.com/views/Factbook-SummaryALL/Introduction?:display_count=y&:origin=viz_share_link`

many students received which grade.

Abstraction

Social Sphere	Institutional analysis
Role(s)	Staff, Students
Norms	Data may be published and released to the public provided individuals cannot be re-identified based upon the output

For grade distributions for a hypothetical Program X which has a large enrolment, this may not be problematic, since dozens or hundreds of students have grades, and thus it is relatively difficult to link any specific student with a given grade. Operationally, they may set up the publication of such a report as follows:

Operation

Situation	For institutional analysis, publish the grade distributions of Program X
Actor(s)	Staff, Students presently registered in courses offered by Program X
Rules	Grade distributions may be published for institutional analysis for each grade, **provided that at least N% of students received that grade**

However, for grade distributions for another hypothetical Program (Program Y) with only a few students, this is more problematic, since it is likely that one could guess the identity of one of a handful of students, and further, be able to link the identity of such an individual to the published grades with high probability. Furthermore, it is likely that the most sensitive grades (such as an A+ or an F) are also the ones which are least commonly assigned to students, and thus the difficulty of guessing the identity of those students receiving such a sensitive grade in these programs would be extremely low.

Operation

Situation	For institutional analysis, publish the grade distributions of Program Y
Actor(s)	Staff, Students presently registered in courses offered by Program Y
Rules	Grade distributions may only be published publicly for institutional analysis for each grade provided the course had an enrolment of more than M students, and at least N% of students received that grade. Otherwise, grade distributions may only be released to individuals authorized by specific bodies (such as those who provide accreditation for the program).

7.1.3 Fitness and Recreational Facilities

Many institutions provide recreational facilities for the use of students and other members of the campus community. Typically, because memberships, program fees, and facility and equipment rentals provide additional revenue to the institution, these facilities must collect, store, and process significant amounts of data themselves.

Increasingly, the importance of these facilities for overall student wellness and health is being recognized, and institutions are introducing programs designed to promote the use of these facilities. For example, the Level Up program at the University of Calgary[2] invites participants to register activities they attend by using a check-in code provided at the activities (which may include workshops, sporting events, and other approved activities), and also counting how many times on-campus fitness facilities are accessed with the participants' campus cards. In return, points are offered which can be redeemed for a variety of rewards, including entries to raffle-style contests to win more valuable prizes.

In this case, there is considerable alignment between participants and the operators of the program with regards to norms about what should happen with the participants' private

[2] `https://www.ucalgary.ca/levelup`, last accessed August 11 2019

data. However, the operators of the program also have access to information about how facilities are used by all campus cardholders. To determine the appropriate practice, they must ask participants to opt-in to the program.

In this case, we can start at the operational layer:

Operation

Situation	Record participation of Student V in the program
Actor(s)	Staff members administering the program, Student V
Rules	Record participation of Student V (by counting points earned, and sufficient details for how those points were earned for accountability) if Student V has opted into the program Do not record any data for Student V if Student V has not opted into the program

An observation occurs when Student V opts into the program:

Observation

Environment	Student V submits an online form consenting to participate to the program
Person(s)	Somebody using Student V's credentials
Practices	Begin recording points and related data for Student V[3]

Another observation occurs when Student V visits the swimming pool after opting into the program:

Observation

Environment	Student V, Swimming pool, program
Person(s)	Student V (according to campus card issued to Student V)
Practices	{not yet known}

Now that Student V has opted into the program, the following operation layer would follow:

Operation

Situation	Record participation of Student V in the program
Actor(s)	Staff administering the program, Student V
Rules	Record participation of Student V (by counting points earned, and sufficient details for how those points were earned for accountability)

The rules would permit the appropriate data handling practice, as shown in the following observation:

Observation

Environment	Student V, Swimming pool
Person(s)	Student V (according to campus card issued to Student V)
Practices	**Record points earned, date, location for Student V**

Because this system relies on the student's opting-in (and therefore their informed consent), less reference to the abstraction layer is required for the day-to-day operation of the program. Instead, rules can be generated for each student opting-in to this data collection to allow the use of data pertaining to their access of recreational facilities.

7.2 Using Data in the Learning Management System

As described in the previous chapters, Learning Management Systems (LMSs) are a class of applications used to centralize interactions between learners, educators, and other parties. In traditional learning environments, the LMS is used to supplement face-to-face learning experiences, such as those happening during a lecture. LMSs are often used to facilitate other kinds of learning environments, such as those which are primarily online and may

involve educators and learners who are geographically distributed, or may be interacting in an asynchronous manner, without a face-to-face component.

These systems are typically used to:

- Provide a central repository for course materials such as lecture notes, videos, and other learning resources
- Provide a means for students to submit material for assessments such as assignments
- Administer assessments such as quizzes
- Provide a means for learners and educators to interact by providing one-to-one communications (such as messaging, or email-style correspondence) as well as one-to-many communications (such as discussion boards)
- Administer grades
- Provide tools to help learners and educators measure performance and engagement in the course

Access to data within the LMS is subject to a complex interplay of legislation, institutional policy, vendor practices, pedagogical decisions, and learner participation.

7.2.1 Course content

The majority of course content held within a LMS is not generally considered private, but access to this data may be restricted in deference to other considerations including academic integrity and the protection of intellectual property rights (such as copyright). For instance, by law in the province of Alberta[Alb00], teaching materials are explicitly not subject to Freedom of Information requests as other records held by a public institution may be. The following excerpt is from a statement posted to the University of Calgary's LMS to which all users must agree on an annual basis, and illustrates the expectation that posted materials

observe other laws and policies, and specifically with regard to creating additional copies and sharing content:

> "Anything posted to D2L must be compliant with Canadian Copyright Law and university policies and agreement. Please consult the copyright web pages or the Copyright Office prior to making materials available... Single copies of materials posted to D2L may be downloaded for personal use. Copyright-protected materials (including course notes, assignments, quizzes and presentations) available on D2L may only be shared provided it is permissible to do so under Canadian Copyright Law, university agreements and permission from the copyright holder."[4]

In the LMS, additional conditions for access may be placed upon course content already held by the system for access by users, but this is done primarily for pedagogical reasons, or to protect intellectual property, rather than for privacy. Educators may decide to withhold advanced material for learners until later, or to only allow other educators (such as co-instructors, teaching assistants, or accessibility experts) to access other portions of course content.

Broadly, the abstraction of a given course using a LMS to organize and distribute course content might look as follows:

Abstraction

Social Sphere	Learning Management System (Education)
Role(s)	Educator(s), learners
Norms	Educators and learners may use freely; content may not be copied unless educator consents

Each time a learner attempts to access course material, a specific situation arises. This is how this might look, beginning with the conditions detected by the system:

[4] `https://ucalgary.ca/provost/copyright_update`, accessed May 23, 2019

Observation

Environment	A laptop not belonging to the institution; Connected via the institution's wireless network; Wednesday, 4 PM
Person(s)	Authenticated user (using the credentials of somebody registered in the course)
Practices	{not yet known}

Operation

Situation	Identified learner accesses course material posted in the LMS by an educator
Actor(s)	Identified learner registered in the course
Rules	Access must observe any restrictions specified by the educator in the system (for instance, lecture slides may only be accessed beginning on the day before the lecture)

Once rules have been produced, the system can generate the practices which should be in place for the particular environment.

Observation

Environment	A laptop not belonging to the institution; Connected via the institution's wi-fi; Wednesday, 4 PM
Person(s)	Authenticated user (using the credentials of somebody registered in the course)
Practices	**Provide the authenticated user with the content they have requested and log the access**

7.2.2 Student work

In some cases, students may be invited to upload their own work to the LMS, generally to be evaluated in some way by educators. The privacy of such work is not as straightforward as course materials from the educator, which are clearly intended to be shared with students. Instead, it is more dependent upon the implicit norms established for a specific course (and assessment).

For example, a personal piece of writing, such as a reflection, should be kept confidential between learner and educator. Students who have made a presentation (and have in effect, taught a portion of the course to their fellow learners) might expect their presentation materials to be shared with the rest of the course, for later reference. Students who have engaged in some form of community-based learning might actually expect their work to be shared publicly and broadly outside the course or even institution.

We illustrate how the LMS would deal with student submitted work, taking contextual privacy into account. We begin by establishing what we know in the abstract:

Abstraction

Social Sphere	Learning Management System (Education)
Role(s)	Educator(s), learners
Norms	Student work should only be shared appropriately

Observation

Environment	A laptop belonging to the institution; Connected to the institution's wi-fi; Thursday 9 AM
Person(s)	Authenticated user (using the credentials of an educator registered in the course)
Practices	**Provide the authenticated user with the content they have requested and log the access**

The LMS provides a protected area for educators and learners to share course content and learner-created artifacts. Such artifacts may simply be byproducts of the learning process (such as questions asked on a discussion board), but may also be work which is shared or submitted for assessment. Such assessments are typically kept private between the educator(s) and the learner(s), (as is often required by law, such as in the province of Alberta[Alb00], for records which form the educational history of an individual), but the confidentiality of the work itself is a more complex issue.

This may be handled by the elaboration of relevant norms:

Abstraction

Social Sphere	Learning Management System (Education)
Role(s)	Educator(s), learners
Norms	Student work should only be shared appropriately; **Group work is available to all group members;** **Work students have been requested to post publicly (such as blog or social media posts) should remain public;** **Student work may be published with consent and co-operation from the student (for example, to disseminate research or to use submitted work as exemplars in the future)**

However, some situations in which the educator proposes to share student work may require the elaboration of relevant rules. For example, in the case of one of the norms listed above ("Group work is available to all group members"), we might require more detail, as follows:

Operation

Situation	Educator of the course decides whether to share student work
Actor(s)	Educators, Students
Rules	Group work is available to all group members, **except for peer evaluations, which should only be shared anonymously and if the educator decides that evaluations have been conducted fairly and constructively**; Group work is available to all group members, **except for members of Team C, for whom peer evaluations will not be shared, due to a possible conflict of interest within the team (two of the three team members are closely related)**

7.2.3 Ambient data

Students often leave traces in the system which can be logged and used by educators, system administrators, other staff members of the institution, and potentially even by the vendor who has licensed the LMS for use by the institution. These traces include session information which records details of a student's environment (including time, duration, device, location, IP addresses) when they access the system, and clickstream data which indicates when students access specific course content and for how long. The LMS may be integrated with other parts of the institution's technical infrastructure such as authentication tools, productivity suites, library resources, or students information portals. Interactions with this infrastructure can also be logged and analyzed. The LMS may also be integrated with other system components (including those created by third-party vendors) which are used for enhancing course content, administering student assessments (for example, restricting a student's browser use when they are completing a quiz within the LMS), or managing communications.

For example, to prepare for meeting with a student who has questions about their course, an educator may check the LMS to see which pieces of course content the student has accessed, how recently they were accessed, and for how long.

The observation of the data collection might include the following:

Observation

Environment	A laptop belonging to the institution; Connected to the institution's wi-fi; Monday 10 AM
Person(s)	Authenticated user (using the credentials of an educator registered in the course)
Practices	**Provide the authenticated user with the content they have requested and log the access**

It is more difficult to collect information about the abstraction (and consequently, the operation). Explicit norms can be gathered without difficulty from relevant legislation, the institution's privacy policy, the contract between the vendor of the LMS and the institution, and possibly by advice from experts such as the institution's legal advisors.

However, implicit norms are more difficult to set from this "top-down" perspective. They may depend upon the particular culture created by the educators and learners brought together by a particular course. The course's outline may explicitly set forth policies to inform students how the LMS will be used to administer their learning, but it is very likely that this information is not present.

Individual students will also have a different comfort level with the data gathered by the LMS. In a study intended to establish privacy boundaries around learning analytics systems, Ifenthaler and Schumacher[IS16] show that students are able to distinguish between data intended for use in different contexts. This includes data intended for educational use, such as test scores and course enrolment data, and data for use in contexts outside of educational use, such as medical data or information about the students' income. Ifenthaler and Schumacher

discuss a third class of data that includes such ambient data as we have discussed in this section, specifically data about when students were online, what they accessed, and what they downloaded. Their findings indicate that students were more reluctant to permit the use of this data compared to data they felt was intended for educational use, but more willing to permit the use of this data compared to data they felt was not intended for educational use.

These findings are also supported by qualitative results by Roberts *et al.*[Rob+16] in which students exhibit unease with pervasive data collection, such as tracking their use of a LMS. Both studies, as well as others (including Jones *et al.*[Jon+19]) also conclude that the willingness of students to permit the use of their data was influenced by the degree of control students felt they had over how their data would be used. Further, Roberts *et al.* observe that individual students exhibit specific privacy preferences dependent upon factors such as their field of study, their own history and experiences, and individual beliefs.

Roberts *et al.* note the "diffculty in developing uniform policies concerning the uniform application of rules and processes that can also allow for autonomous and personalized decision-making and action by each individual student." It is this gap which the framework in this thesis aims to fill, by providing a means to generate rules which can be customized according to specific situations.

However, further understanding of the privacy attitudes and behaviours of each actor in a given social sphere is still required to provide realistic norms upon which rules can be individually built. The study presented in the next chapter presents a first step towards understanding the norms which may apply in a post-secondary educational setting.

Chapter 8

Studying the privacy behaviours and attitudes of undergraduate students

Previous chapters in this thesis describe a possible framework for contextual privacy, and lay out a map of how data is collected and used in post-secondary education. This chapter seeks to unify this by interrogating the privacy attitudes and behaviours of post-secondary students to discover the degree to which students are conscious of privacy issues in their immediate environment, and how they already vary their privacy behaviours based upon contextual cues.

Appendix A contains the text of the questions created for the study described in this chapter. Readers may refer to Appendix B for the details of the study, including questions asked in the survey and full results.

8.1 Motivation and Background

Comprehensive studies of consumer privacy behaviors already exist in the literature. Notable examples include studies undertaken by Westin[Wes91; SOW95] and Pew Research[MR15] in the United States. Other studies have focused on other parts of the world, such as a two-year mixed methods study in New Zealand in 2013 and 2014 published by Lips and Eppel[LE17].

This study does not focus specifically on students at post-secondary institutions, although Lips and Eppel did include in their study individuals aged 18-24. The study by Lips and Eppel identified several segments of their population (including their younger individuals) which had more careful and nuanced approaches to privacy than the population at large.

In general, there is a suggestion in all of these studies that context matters to data providers. As indicated by the Pew Research study on Americans' Attitudes About Privacy, Security, and Surveillance[MR15], data providers have varying amounts of confidence in the ability of different companies and service providers to keep their data private and secure, and also vary their expectations with regards to data retention depending on the company or service provider. In general, there is more trust in government and financial services providers, and less trust with Internet-based companies such as search providers or advertisers. Lips and Eppel[LE17] also had similar findings in New Zealand, but with more of a focus on privacy behaviours, finding again that participants tended to trust government and financial services with their data more than with other services and businesses online.

There should be a more careful focus on post-secondary students, as several factors suggest that they may approach privacy in a different way than other kinds of consumers.

Students live highly contextual lives. In the course of an average day, they negotiate a wide range of social settings, including being in class (with different courses potentially occupying distinctly separate disciplinary cultures, and which may involve large lectures, more focused labs, or more intimate tutorials or seminars), pursuing extra-curricular pursuits such as participating in student groups, volunteer work, or intramural sports. They also experience life in less formally organized settings such as studying in the library, patronizing campus coffee shops, restaurants, or bars, or accessing recreational facilities such as the fitness centre or gym. They may socialize with a few close friends and roommates, or more distant acquaintances they met in class or through other activities. Some may live on campus, and thus, while in residence, also experience aspects of home life while on campus. They may also be part of other social spheres, such as while working at a part-time job, receiving care

from health or social services providers, being in contact with family members, or socializing with friends from off-campus.

Undergraduate students are a well-studied population in many aspects of life, because they are readily accessible to academic researchers. Many disciplines have engaged in privacy studies involving students, focusing on social media, and Facebook in particular. These studies began as soon as social networking services (especially Facebook) entered the public consciousness[AG06; SK10], and studies on this topic continue to the present day. Readily accessible data from Facebook about its users (including their privacy settings) provides researchers with a platform from which the privacy practices and attitudes of their subjects can be readily studied. However, these studies do not provide much insight with regards to how subjects approach privacy issues outside of how they interface with social media.

In recent years, additional studies have investigated user privacy in the context of mobile devices and applications[LK17; Nae+17]. These indicate that users have privacy concerns that can be connected to a variety of factors, including many which are contextual (such as where and the reasons for which data is being collected). One of these studies, by Lee and Kobsa[LK17], was conducted at a university campus asking students about privacy preferences. These preferences were then used to separate students into four clusters based on their responses which could then be used to predict their preferences. Some distinguishing factors that affected student preferences were noted, including the medium by which data was collected (video, photo, eye-tracking), the party performing the collection (the institution, government, unknown), or the general purpose for which data was being collected, and the frequency with which students were being monitored (continuous monitoring *vs.* one-time data collection).

While Lee and Kobsa are very thorough in asking students about their privacy preferences for hypothetical scenarios, the focus of these scenarios were all on-campus. There is relatively little data about how students feel about the entire gamut of activities that collect data from them on an everyday basis, which includes not only activities based on campus, but may

also involve online and off-campus interactions.

One of the most comprehensive series of studies of undergraduate technology use (and occasionally privacy) is the ECAR Study of Undergraduate Students and Information Technology, published annually by EDUCAUSE, a professional organization focused on information technologies practitioners in post-secondary institutions. The 2015 edition of the study[Dah+15] asks students about the kinds of data that they are comfortable having their institution use. The responses to this particular question anticipates the research which this chapter undertakes as students indicate comfort with institutions using data that are more tightly linked to learning activities, and increasing amounts of discomfort with institutions using data which is less tightly linked to these.

However, there is a gap in the research with regards to undergraduate privacy concerns in a comprehensive way, which incorporates both their lives on-campus and off-campus, offline and online, and considers the ways in which data is both actively and passively collected about them. The study presented in this chapter is a first step to filling that gap.

8.2 Research Questions

We seek to gain a better understanding of the privacy attitudes and behaviours of undergraduates enrolled at the University of Calgary. Passive data collection from students is as important as data students actively provide, so it is important to investigate a broad snapshot of how students interact with technology, both off-campus as well as on-campus. This includes the devices, services, and activities they engage with. It is also important to learn to what extent students feel that their privacy is protected when they are on campus. We also want to understand whether (and how) students distinguish between the appropriateness of different activities which may affect their privacy, from a range of institutions.

Specifically:

- How do students use technology, on and off campus?

- To what extent are students aware of the privacy policies which govern their data?
- To what extent do students believe their privacy is protected while on campus?
- What are students' privacy beliefs when engaging with their school, businesses off-campus, and government organizations?
- Do students vary their attitudes depending on the context of the specific interaction which requires the release of their personal data?

8.3 Methodology

As a first study into this important topic, we conducted a quantitative study which touches on the research questions identified above. Passive data collection limits the range of data we can collect and how we conduct the analysis. For instance, the Likert questions we use, for instance, must be treated as categorical data and should not have statistical means, such as averages, applied to the data generated from them. However, passive data collection does allow us to ask students about a broad range of topics. Therefore, the study is intended to touch on all of the research questions to identify areas for future, more in-depth investigations.

We recruited undergraduate students to take a survey we developed and hosted on Surveymonkey[1]. The recruitment occurred in two stages, first soliciting respondents by flyer (posted in prominent locations at the University of Calgary's main campus) in Winter 2018 (in March and April of that year), and then by reaching out to student organizations on campus to forward our call directly to their membership via email or announcements posted to non-credit D2L courses in Fall 2018 (in September and October of that year).

We received 79 responses, of which 58 completed the survey instrument in full. All results discussed only use responses provided by the 58 respondents who completed the entire survey instrument. A full set of results is provided in Appendix B.

[1] `http://www.surveymonkey.com`

Table 8.1: Questions asked about the frequency of device use and online behaviours

Frequency of device use	Laptop, Desktop, Smartphone, Tablet, Wearable, Smart Home Devices
Frequency of social media use	Facebook, Twitter, Instagram, Google+, Snapchat, Whatsapp, Signal, WeChat, Weibo
Frequency of behaviours	Streamed content, Searched for content, Patronized online retailers, Patronized traditional retailers online, Performed online banking, Used government services online

Table 8.2: Questions asked about the recent use of campus services

Campus wi-fi
Computer labs (such as those located in the library)
Food services (such as the Dining Centre, or those located in food courts)
Loaded money onto a student card
Accessed a campus space using a student card
Accessed a campus printer using a student card
Used campus recreational facilities (requiring sign-in with a student card)
Used the library (either online or physically accessing resources)

Technology Use Questions in this section are intended to determine a baseline for how respondents ordinarily experience technology. We ask respondents about how frequently they use selected devices, social media services and engage in specific technology-mediated activities such as streaming content, online banking and shopping. Table 8.1 lists the specific devices, platforms, and behaviours mentioned in these questions.

Behaviour and Technology Use On Campus Questions in this section are intended to establish a baseline for how respondents experience technology while on campus.

In this section, respondents are asked about their behaviour and technology use on campus. They are asked about their device use (which may be different than when respondents are off-campus), which campus services they use, as well as what University-supported technologies they access. To understand which services were used, we asked students which of the services listed in Table 8.2 had been used in the six months prior to taking the survey.

Pre-existing privacy attitudes and the University In this section, respondents are asked questions about how they feel the University handles their private data. They are

Table 8.3: Questions asked about the University's privacy policy and how the university protects personal data

The University's privacy policy is easy to find
The University's privacy policy protects my privacy
I remember consenting to the University's privacy policy
My data is well-protected by the University
The University is careful about how it reuses my data
The University should be allowed to use my personal data if it helps me with my degree
The University should be allowed to use my personal data if it helps to improve my well-being
The University should be allowed to use my personal data if it helps to improve the quality of my classes
The University should be allowed to use my data if it helps to improve my experience at the U of C

Table 8.4: Questions asked about who students trust with their personal data

I trust student advisors with my personal data
I trust my professors with my personal data
I trust my TAs with my personal data
I trust other University staff with my personal data
I trust student organizations that I belong to with my personal data
I trust my classmates with my personal data

asked how strongly they:

- trust the university's privacy policy and in general how it protects personal data (see Table 8.3 for a full listing of statements that respondents were asked to rate)
- trust the University-affiliated groups (such as professors, student advisors, student groups) (see Table 8.4 for a listing of statements that respondents were asked to rate)
- would trust the university to reuse data given a variety of circumstances

Pre-existing privacy attitudes off-campus In this section, respondents are asked questions about:

- their general attitudes towards privacy (how their personal data is reused or treated in contrast to published privacy policies)

- the degree to which they trust different categories of off-campus entities (search providers, social media services, business they interact with)

Students and context In this section, respondents are asked questions about how they would feel about different parties (such as the university, government agencies, physical "bricks-and-mortar" retailers, and online retailers) handling their data in specific ways:

- Sending information to a private email address provided by the respondent
- Contacting the respondent via social media to continue an interaction initiated by the respondent
- Initiating an interaction via social media
- Providing the respondent's contact information to a third party

Demographic Data For analysis of their data, respondents were asked to provide demographic data including their age range, gender, time in program, Faculty, and whether they lived on or off campus. The full set of demographic data we collected is included with our results in in Appendix B.

Respondents were primarily young adults, with thirty (51.72%) respondents reporting that they were between the ages of 18 and 20, and another fourteen (24.14%) respondents reporting that they were between the ages of 21 and 23. Thirty-three (33) of the respondents were female, twenty-five (25) were male. Each year of program had representation, with most (twenty, or 35.09%) being in third year. Respondents represented six different faculties, with forty-one (41) or 70.69% of the respondents being from either the Haskayne School of Business or the Faculty of Nursing. This is an artifact of the recruitment processing, which saw some student groups respond more enthusiastically to the survey than others. Thirty-six (36) of the respondents reported that they lived off-campus, with family. The other living situations with large groups of respondents were those who lived on or off-campus with roommates, with nine (9) and six (6) respondents respectively.

Table 8.5: Frequency of usage for selected devices

	Every day	Several times a week	Weekly	Monthly	Never Use
Laptop	47	7	1	0	3
Desktop	6	7	7	22	16
Smartphone	57	0	0	0	1
Tablet	6	6	9	10	28
Wearable	6	1	1	1	50
Smart home device	3	2	3	3	57

8.4 Results

In this section, we present some of the most relevant findings. Due to the number of respondents, we were unable to further segment the data based upon the demographic data with which we were provided.

8.4.1 Technology use by undergraduates on and off-campus

Understanding the kinds of technology in use on campus by students also lets us understand how data is being accumulated from students on campus, and also what other organizations may have access to some of that data.

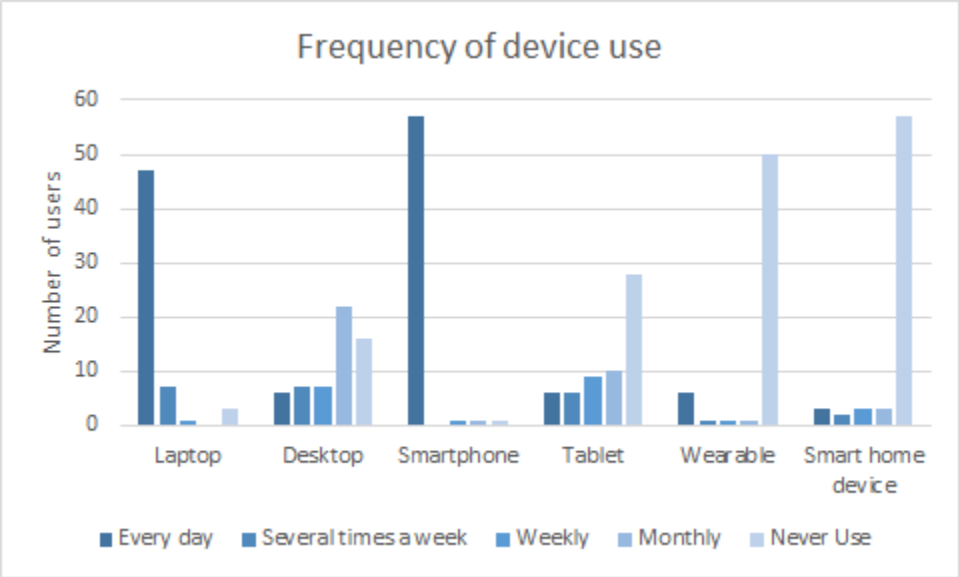
Devices Students were asked about how frequently they use a variety of devices, the results of which are reported in Table 8.5, and illustrated in Figure 8.1. Smartphone use is prevalent, as nearly all (57 of the 58) respondents use a smartphone on a daily basis. The remaining respondent reported that they never used a smartphone.

The other device used frequently is a laptop computer, for which most respondents also reported daily use. Desktop computers and tablets also see frequent use, although few respondents use either device on a daily basis.

Wearables and "smart home" devices do not see frequent use by undergraduates. There are many possible explanations for this, including the cost of the devices, and the perceived utility provided by them.

Students were also asked about the devices they used on campus. Laptop computers and

Figure 8.1: Frequency of usage for selected devices



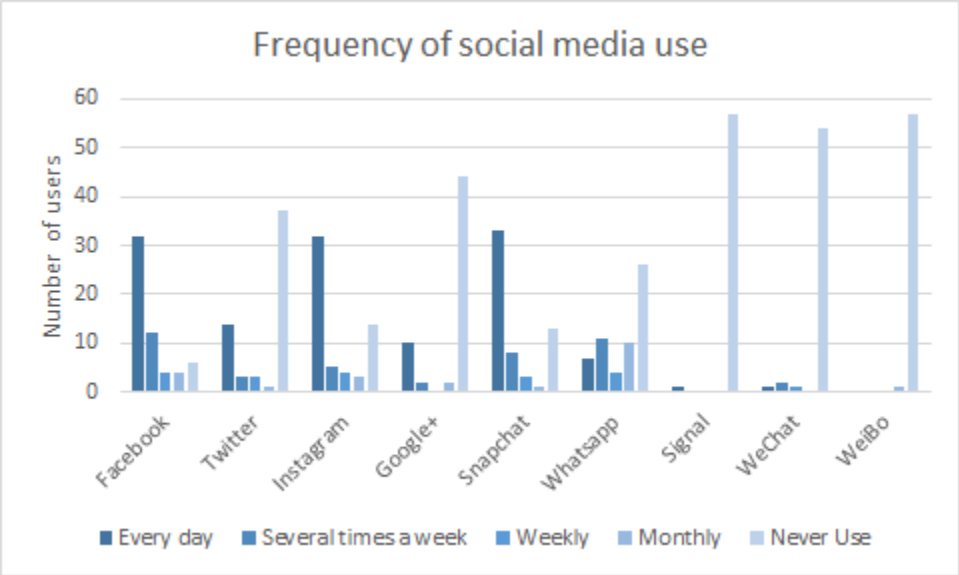
smartphones were used by nearly all of the respondents. Similar to the patterns observed for overall use, tablets and desktop computers also see some use, although not nearly as frequently.

Services We surveyed respondents on the social media platforms they commonly used, with the results illustrated in Figure 8.2. Facebook, Instagram, and Snapchat were the services most commonly used, with over half the respondents reporting daily use. Twitter and Google+ (which had not yet begun to be decommissioned by Google at the time of the survey) had a smaller group of regular users, whereas approximately half the respondents used Whatsapp, but rarely on a daily basis.

Other platforms such as Signal, WeChat, and Weibo have not been widely adopted by the respondents. Because respondents were also provided with an open-ended prompt about other social media services they used, a handful also reported that they used related services such as Telegram, YouTube, and Tinder, but the study did not focus on these services as the goal of this question was to acquire a broad snapshot of the social media platforms popularly used by students at the University of Calgary.

Respondents were asked how frequently they engaged in selected online activities. Fig-

Figure 8.2: Frequency of social media usage

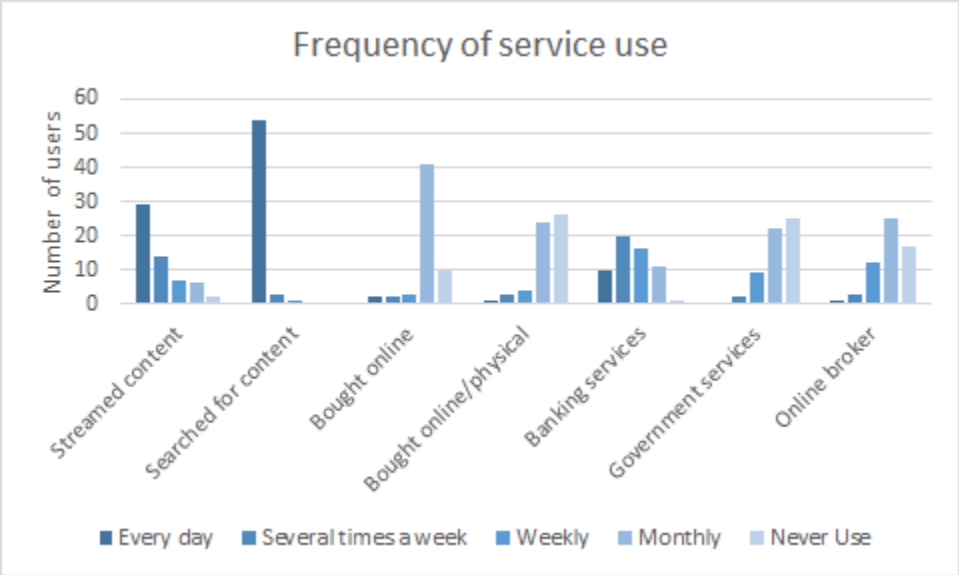


ure 8.3 shows the most common activities were searching for content online and streaming content from digital providers. Other activities, such as buying items from online stores or accessing government services are not activities which students perform daily, probably due to the lack of necessity or funds, and thus the frequency of their use is significantly less. However, a sizeable number of students do access banking services online on a regular (daily to weekly) basis.

Respondents were also asked about whether they had used selected campus services in the last twelve months. Nearly all had used the school's wireless network or one of the computer labs available for their use. Nearly half had used their ID card as a debit card (requiring students to load money onto the card at selected locations on campus), accessed a campus space using the ID card, or used the University's fitness centre (again, requiring students to check-in using their ID card).

Respondents were also asked which technologies supported by the university they had used. These include services such as the university's learning management system, email services, and productivity suites. In the twelve months previous to their taking the survey, all but one respondent had used the university's learning management system. Nearly 90

Figure 8.3: Frequency of online services usage



percent of the respondents had accessed their university-provided email (using the provided webmail interface).

8.4.2 Privacy polices on-campus

Students were asked to rate a number of statements about how the University handles their privacy. While most did not express an opinion as to how well the University protected their privacy, or handled their data, there was a more obvious pattern to student responses when asked about their engagement with the University's privacy policy, to which all students implicitly consent to upon enrolment[2].

Twenty-two (22) students, or 37.9% of the respondents, responded either "Strongly disagree" or "disagree" to the statement "The University's privacy policy is easy to find." However, most of these students (17 of them, or 77.2% of these students) did not espouse a strong opinion *i.e.* answering "disagree") with respect to the statement.

[2]The University's Calendar[Cal19b] lays out the responsibility of members of the University community (including students) "to familiarize themselves with the Statement on Principles of Conduct" contained within the calendar, which commits them to follow numerous policies, amongst them, the University's Code of Conduct[Cal19a], in which stakeholders, including students, are enjoined to "fulfil their University responsibilities in compliance with applicable laws, and applicable University policies and procedures..." which includes the University's Privacy Policy[Cal17].

By contrast, forty-two (42) students, or 72.4% of the respondents answered either "Strongly disagree" or "disagree" to the statement "I remember consenting to the University's privacy policy." Nearly half of these students (20 of the 42, or 47.6%) did express a strong opinion, answering "strongly disagree" to the statement, which is in strong contrast to the responses to the previous statement.

The combination of the responses to these statements are a strong expression of a lack of engagement with the University's privacy policy by respondents. While they may assume that there are no concerns with how the University protects their data, students also do not know how exactly their data is protected according to policy, nor do they recall specifically providing consent to these policies.

8.4.3 The privacy beliefs of undergraduate students

Beliefs do not necessarily imply behaviour, or vice versa. However, we can ask students to provide information about their beliefs even if the parameters of this study leave us unable to fully capture the behaviour of students.

On-campus When asked about what uses they would like to see their data put towards, respondents generally do not express strong opinions either in support or in opposition of such uses, as shown in Table 8.6. There is a slight tendency to disagree more with statements which suggests more general uses of data, such as "The University should be allowed to use my personal data if it helps to improve my well-being" and "The University should be allowed to use my personal data if it helps to improve my experience at the U of C."

Off-campus Respondents were also asked to rate statements about the organizations they trusted, and statements about their private data in general. The results are shown in Table 8.7.

The data summarized by Table 8.7 shows a number of interesting results.

Table 8.6: Student attitudes towards data use by their university

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
The University should be allowed to use my personal data if it helps me with my degree	3.45%	27.59%	20.69%	39.66%	8.62%
The University should be allowed to use my personal data if it helps to improve my well-being	12.07%	20.69%	24.14%	36.21%	6.90%
The University should be allowed to use my personal data if it helps to improve the quality of my classes	5.17%	20.69%	27.59%	39.66%	6.90%
The University should be allowed to use my personal data if it helps to improve my experience at the U of C	10.34%	20.69%	22.41%	39.66%	6.90%

Off campus, students exhibited similar trust of institutions such as the government, healthcare providers, and financial institutions. Respondents show a particularly high degree of trust in healthcare providers, with a marked increase to ratings of "Agree" and "Strongly Agree" compared to statements about other institutions such as the government and financial institutions. This may be expected due to their age, since young adults may have more familiarity with their healthcare providers than they would with governments or financial institutions. This may also explain the high "Strongly Disagree" rate received to the statement "I trust the government with my personal data" from 22.41% of respondents, which is higher than equivalent statements for other institutions and businesses (online and offline).

As we expected, respondents did not trust businesses as much as they did institutions. Only around 5% of respondents gave a rating of "Strongly Agree" to the statement "I trust companies with whom I do business offline with my personal data," compared to percentages of 12.06%, 17.24%, and 27.59% of respondents giving the rating of "Strongly Agree" to

Table 8.7: Student attitudes towards data use by organizations and privacy in general

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I trust the government with my personal data	22.41%	13.79%	18.97%	32.76%	12.06%
I trust financial institutions with my personal data	10.34%	13.79%	20.69%	37.93%	17.24%
I trust my healthcare providers with my personal data	3.45%	5.17%	15.52%	48.28%	27.59%
I trust companies with whom I do business offline with my personal data	12.28%	21.05%	22.81%	38.60%	5.26%
I trust companies with whom I do business online with my personal data	18.97%	25.86%	25.86%	25.86%	3.45%
I trust search providers that I use to search the web with my personal data	37.93%	24.14%	17.24%	15.52%	5.17%
I trust my social media services with my personal data	37.93%	20.69%	17.24%	22.41%	1.72%
I consider my personal data private no matter which organization uses it	15.52%	15.52%	20.69%	17.24%	31.03%
I am aware that some companies use my private data for more than one thing	0.00%	3.45%	8.62%	34.48%	53.45%
I don't mind if companies use my private data for more than one thing, as long as it's clearly stated in their privacy policy.	12.07%	27.59%	15.52%	29.31%	15.52%
It disturbs me when a company uses my private data for more than one thing	6.90%	22.41%	20.69%	24.14%	25.86%
Sometimes I am okay with trusting one company with my private data, but not another	3.45%	8.62%	12.07%	51.72%	24.14%

equivalent statements about the government, financial institutions, and healthcare providers respectively. Online businesses were trusted even less than offline businesses, with more respondents providing ratings of "Strongly disagree" and "Disagree" to the statement "I trust companies with whom I do business online with my personal data" compared to the equivalent statement about offline businesses. Respondents were the least trusting of search providers and social media services, which both had 37.93% of respondents provide a rating of "Strongly Disagree" when asked if they trusted these organizations with their personal data.

Students are, on the whole, aware of the realities of data protection in the modern world. Most are aware that their data is often reused by organizations (87.83% of respondents agreed with this statement in some way), with exactly 50% reporting that they are not at ease with this practice. Three-quarters (75.86%) of the respondents indicated that in most cases, they would trust some companies with their data in comparison with others.

8.4.4 Students and contextual beliefs about privacy

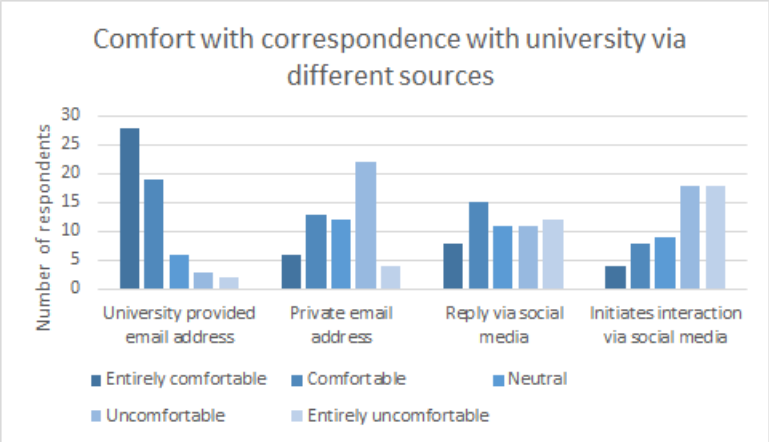
We also sought to understand how contextual changes influence how students view the privacy of their data. We tested this in the study in numerous ways.

A single organization: the University First, we asked students how comfortable they were trusting different groups of people (faculty, staff or students) at the university with their data.

We asked students about their comfort level with various forms of correspondence originating from the University:

- from the institution to a university-provided email address
- from the institution to a non-university provided email address
- from the institution to continue a social media interaction

Figure 8.4: Comfort with university correspondence via different sources



- from the institution to initiate a social media interaction
- from a department or unit they had previously interacted with
- from a department or unit they had not previously interacted with
- from a third party who had obviously received information about the respondent from the institution

There are many differences made evident by context. Consider the differences illustrated in Figure 8.4 in how students view correspondence sent to a university-provided email address, a non-university provided email address, and via social media (both at the student's initiation, and at the institution's).

We then asked students about three kinds of data, pertaining to their academic performance, non-academic activities on-campus, and activities off-campus. We asked about their comfort level should different units at the university (classified as academic and non-academic) contact them with knowledge about these three types of data.

Here, there are also evident differences between how comfortable students report they are with academic units handling their data, and with non-academic units. When asked to rate their comfort level (as entirely comfortable, comfortable, neutral, uncomfortable, entirely uncomfortable) in response to the statement "A (academic/non-academic) unit con-

tacts you with knowledge about your (academic performance/non-academic activities on-campus/activities off-campus)," there are discernible differences in the responses based upon the type of unit, and the type of data. Table 8.8 shows the percentages of responses collected in this way.

We expect that most students are comfortable with academic units who use information about their academic performance, and 52.63% of the respondents rated this statement with a comfort level of "entirely comfortable" or "comfortable." Only 5.26% of respondents (or 3 individuals) rated this statement with a comfort level of "entirely uncomfortable."

We note that the respondents are significantly more uncomfortable with a non-academic unit handling data about their academic performance, compared with an academic unit. Only 15.79% of the respondents chose comfort levels of either "entirely comfortable" or "comfortable," whereas 75.44% of the respondents chose comfort levels of either "uncomfortable" or "entirely uncomfortable."

Students are somewhat less comfortable with academic units using data gathered about their non-academic activities, even on campus. 38.59% of the respondents chose comfort levels of either "comfortable" or "entirely comfortable," a percentage change of 14.04% (meaning there are 8 fewer respondents who chose these comfort levels).

Once we ask students about data about off-campus activities, students demonstrate a striking amount of concern. Two-thirds of students chose a comfort level of "uncomfortable" or "entirely uncomfortable" when asked about academic units using such data, while a similar number (57.9%) chose the same comfort levels when asked about non-academic units using such data.

Multiple organizations: Universities, Governments, Businesses Students were also asked to compare how comfortable they would be with a variety of data-handling scenarios. The only element which changed in the scenarios presented to them was the kind of organization performing the data collection or processing. We find similar levels of comfort

Table 8.8: Use of academic, non-academic and off-campus bodies data by different on-campus bodies

Type of data	Type of Unit									
	Academic					Non-academic				
	EC	C	N	U	EU	EC	C	N	U	EU
Academic	15.79%	36.84%	24.56%	17.54%	5.26%	7.02%	8.77%	8.77%	36.84%	38.60%
Non-academic	8.77%	29.82%	24.56%	14.04%	22.81%	5.26%	15.79%	12.28%	26.32%	40.35%
Off-campus	5.26%	14.04%	14.04%	24.56%	42.11%	7.02%	21.05%	14.04%	26.32%	31.58%

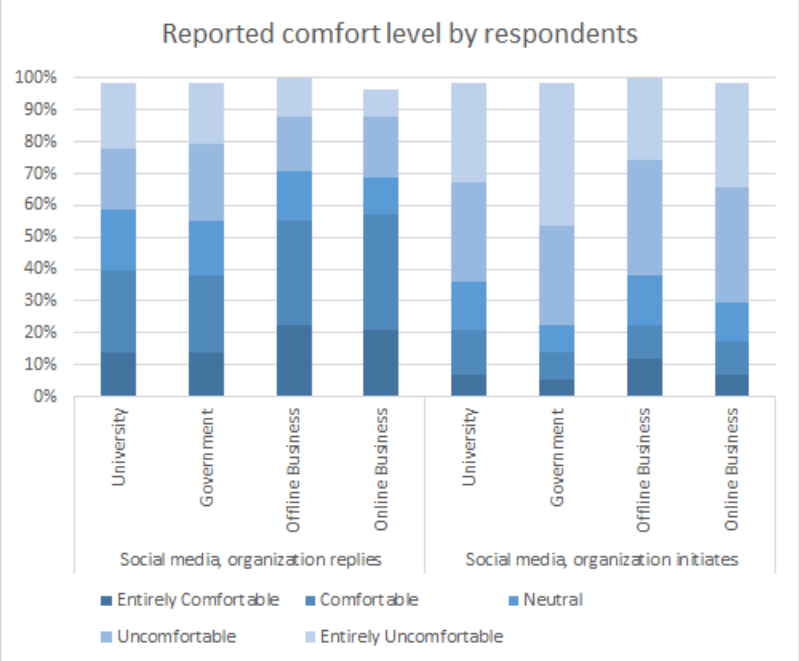
Students were surveyed on their comfort level on the uses of different types of data by different university units, and asked to rate it as EC (Extremely Comfortable), C (Comfortable), N (Neutral), U (Uncomfortable), EU (Extremely Uncomfortable)

indicated by students when asked about receiving correspondence from universities, government agencies, offline businesses (defined as businesses which have a physical storefront), and online businesses (defined as businesses with whom they only interact with online) from a provided email address.

Respondents are more uncomfortable when asked about organizations using a private email address for correspondence which was not provided to them by the respondent, compared to when those organizations use a private email address which was provided to them by the respondent.

The most marked difference in these sets of questions comes when respondents are asked about being contacted (using a hypothetical scenario) via social media. Two scenarios are presented. The first is when an organization replies using social media to an interaction initiated by the student, and the second is when the respondent receives an interaction via social media, initiated by the organization. Results are illustrated in Figure 8.5.

Figure 8.5: Reported comfort level of social media interactions by respondents



Context influences results in at least two ways in this figure. Respondents display obviously different comfort levels depending on whether they initiated the interaction via social

media, or whether the organization initiated the interaction. No matter the organization, respondents are distinctly less comfortable with organizations initiating interactions via social media. We also see obviously different comfort levels depending on the organization with whom the respondents are interacting. Respondents are quite comfortable interacting with businesses via social media, and less so with universities or governments, especially when the organization is replying to respondents via social media.

In the first scenario, students are decidedly more comfortable interacting with businesses in this way. In the second scenario, students have similar comfort levels (or discomfort levels) no matter the organization, but in all cases, respondents are significantly less comfortable.

Although this result may be counter-intuitive to some readers (who may assume that interactions with government or institutions should always inspire more trust than with private entities such as businesses), a number of possible explanations are possible.

The first is that this may be an instance of the "privacy paradox" documented in the literature beginning with Acquisti *et al.*[AG05; Kok17]. This is a well-studied aspect of human behaviour in which the short-term privacy practices of end users do not align with their longer-term privacy beliefs.

Another explanation may be that the survey respondents are typically young adults living at home with family. Since they have often had family members assist or act as intermediaries in dealing with government services such as registering a vehicle or filing taxes, they may not have enough familiarity with different government agencies to be comfortable with them.

Another explanation is grounded in uses and gratifications theory, which emerges from communications studies. This theory suggests that end users are active consumers of media, and have specific uses for the media which they consume, and specific gratifications which they hope to derive from the media[Blu79].

Applied to our scenario, this theory would suggest that our respondents have specific uses and gratifications for engaging with social media (perhaps, to project a specific self-image or identity, to connect with family or friends, or to please themselves or a close associate),

and the reasons our respondents interact with businesses align more closely with these uses and gratifications than the reasons they have for interacting with government agencies or institutions. Research in marketing suggests that businesses have taken advantage of this insight (deliberately or not) to a greater degree than government agencies or institutions as a whole have[Dol+16] and therefore inspire more trust from our respondents when interacting with them via social media. Studies coordinated by Ifinedo[Ifi16] provide additional support for this explanation by finding that students use social media mainly for entertainment and social connection rather than for gathering information. Governmental agencies and institutions tend to use social media more for distributing information than for providing entertainment or social connection, compared to business or personal social media accounts, which may explain why undergraduates trust them less. Additional research with more in-depth questioning would be required to fully explain this finding, which is somewhat peripheral to our broader study of privacy in an undergraduate environment.

8.5 Students and Context

In this section, we link our findings back to the contextual framework presented earlier in the thesis. Consider a student who is a typical respondent of our survey. They are on campus at the University of Calgary for a limited amount of time each day, as they live off-campus with family members. They regularly use their phone (on which they use Instagram and Snapchat), and a laptop, which is used for studying (for example, accessing their university email or D2L) or streaming content.

Now consider a scenario where the University contacted this student about a new initiative related to their program of study. To determine the most appropriate course of action, the contextual framework could be used to frame the particular situation to understand the factors contributing to the student deciding whether or not their privacy had been infringed upon. To demonstrate this, we describe three cases. In the first, the messaging will not

contain information targeting any particular students. In the second, the messaging will contain information targeting students who volunteer for a initiative at the University. In the third, the messaging will contain information targeting students who volunteer for an organization based off-campus.

They begin with a set of potential operations :

Operation (no Targeted Messaging)

Situation	Email Student, no messaging
Actor(s)	Institution's staff, Student V
Rules	{Not yet known}

Operation (on-campus messaging)

Situation	Email Student, Messaging with on-campus activities
Actor(s)	Institution's staff, Student V
Rules	{Not yet known}

Operation (off-campus messaging)

Situation	Email Student, Messaging with off-campus activities
Actor(s)	Institution's staff, Student V
Rules	{Not yet known}

Each of these operations can refer to the same abstraction to generate appropriate rules. The norms in this abstraction have been populated by relevant policy and legislation, and also informed by our survey. Before the survey, we have the abstraction below:

Abstraction

Social Sphere	Promote initiative to students
Role(s)	Staff, Students
Norms	the University may contact students using a university-provided email address for business reasons Students may receive correspondence from the University to their university-provided email address

The survey prompts the addition of added implicit norms (see Table 8.8)

Abstraction

Social Sphere	Promote initiative to students
Role(s)	Staff, Students
Norms	the University may contact students using a university-provided email address for business reasons Students may receive correspondence from the University to their university-provided email address **Students may receive messaging about their on-campus, non-academic activities** **Students should not receiving messaging about their off-campus activities**

Now we return to our operations, with the following rules:

Operation (no Targeted Messaging)

Situation	Email Student, no messaging
Actor(s)	Institution's staff, Student V
Rules	**Email may be sent to student without targeting messaging**

Operation (on-campus messaging)

Situation	Email Student, Messaging with on-campus activities
Actor(s)	Institution's staff, Student V
Rules	**Email may be sent to student with messaging, but must contain endorsement or referral from the student's other University-related volunteer position.**

Operation (off-campus messaging)

Situation	Email Student, messaging with off-campus activities
Actor(s)	Institution's staff, Student V
Rules	**Student may receive messaging about the initiative, but email should not be sent with targeted messaging**

We note that in these examples, the framework shows that considerably more needs to be known about the privacy preferences of the individual. For example, the rule generated in the second case (which targets messaging to students volunteering on-campus) contains a condition for which we have no data. The intuition is that the student would find such messaging more reassuring if there is traceability, and they know how they were targeted for the new initiative, but we would need more empirical evidence before putting this into practice.

Likewise, in the final case, there may be a subset of students who are not concerned that the University has knowledge of their off-campus activities. They may assume this was information offered to the University via the admissions process or through an application to a program or scholarship. Considerably more work needs to be performed on an individual basis to be able to form implicit norms which are representative of students in general, and also of individual students.

8.6 Limitations and Broader Implications

Due to the two phases of subject recruitment used, it is unclear how results may translate to other cohorts. While some demographics (gender, living situation) of the response group are likely representative of the broader population of undergraduate students at the University of Calgary, there are others which are markedly not (year in program, faculty). The University of Calgary also is only one example of a post-secondary institution in North America, being a research-intensive institution located in a large Canadian city, with the main campus in a suburban setting. There may be differences amongst the undergraduate populations of different institutions, for example, those which are more isolated, have a primarily residential population, or are situated in settings with few nearby amenities (which would then assume that students must stay near campus throughout the entire day).

However, this study has implications which should be kept in mind to improve the student experience in terms of their privacy, particularly at North American post-secondary institutions.

The most obvious result is that of student engagement with University's privacy policy. This study surfaced the fact that relatively few respondents understood when (or if) they had provided consent, and if they had, did not have much knowledge about how to find the policy and see what protections they were entitled to. Is it important for students to know what policies they have consented to, and what their rights and obligations according to these policies will entail? Can students properly advocate for their own privacy rights under the *status quo*? On a broader note, one may be curious as to other University policies which play an important part in the on-campus lives of students with which students have relatively little knowledge and engagement.

We also see from the study's results that technical systems intended to safeguard privacy must take context into account, as our respondents did throughout the survey. Our respondents weigh their privacy decisions on a case-by-case basis, which suggests that systems designed to protect their privacy should also mirror this case-by-case approach. The exercise

we performed in the previous section shows one way in which this can be accomplished.

We also find that our results raise additional questions that are best answered through additional follow-up studies. Given the uneven recruitment of students from across faculties, data collection using the same (or a lightly modified) instrument might be more valuable with a more intentional recruitment strategy.

For example, why do students trust the government with their personal data less than other institutions? Why are they more comfortable on social media with businesses than their university? Additional studies may also contribute in surfacing a broader set of implicit norms, but more importantly, provide a method for eliciting implicit norms from a population, a task which is beyond the limits of computability and decidability, but might be more reliably generated by human specialists provided with the right methodology.

There are other reasons to continue the study of undergraduate privacy behaviours and attitudes. It is clear that there are numerous clusters of students which might share different implicit privacy norms than their peers. For example, mature students likely have a different relationship to technology use and view how data is collected in everyday life differently than their younger peers. International students may have longer histories in societies and cultures with differing privacy values and have been exposed to different ways of engaging with technology. We could also consider students of diverse backgrounds, differently abled individuals or those who ascribe to cultural, sexual, gender or other identities and are often viewed as different. These are the groups who may be the most dependent on how we consider context for privacy protection, and thus it is vital to better understand their privacy attitudes and behaviours to realistically implement a system for contextual privacy protection.

Chapter 9

Conclusion

The modern university serves many more roles than as merely an institution for research, learning, and teaching. As students pursue their lives within the complex of facilities and services which are often invisible to them, we have a responsibility to consider how to safeguard their data and their privacy. This is not only for the good of students, but as modern societies contemplate a way of life which mirrors the degree of data collection employed at most modern universities, understanding this problem at such a small scale will pave the way towards meaningful privacy protection for each member of the wider society.

9.1 Contributions

The major contribution of this thesis is a novel framework for contextual privacy (shown by Figure 9.1 which expands the concept of a context into three layers, and builds relationships between the three. The framework introduces the *abstraction*, *operation* and *observation*, and elements representing the ambient, transference and participatory aspects for each. This blends together the need for ways to elaborate context in a social sense as well as a physical sense. We explore the robustness and expressivity of the framework through a series of high level and low level examples.

The next contribution is an in-depth survey of data collection and processing practices

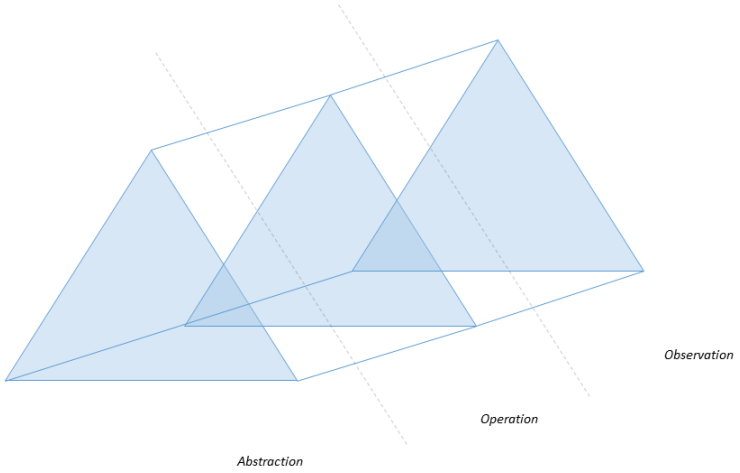


Figure 9.1: An overview of the framework

in North American post-secondary institutions. We start with data collection in the classroom, moving to activities which support teaching and learning but may not directly involve teaching and learning, and then to ancillary services. There are numerous types of data processing in North American post-secondary institutions, and this thesis discusses traditional business intelligence, educational data mining, and learning analytics as uses which are representative of the analytics landscape in this area. Unlike previous work in these areas, we link together data processing practices which are used solely for teaching and learning, with activity which takes place on campus which use students as data providers but outside of the learning environment.

We apply this framework to privacy in post-secondary education, considering both data collection and processing practices. In this, we consider how institutional analysis, alumni relations, and everyday operations such as fitness and recreational facilities or a learning management system collect and process data. We integrate what we know with other research in the field to build a case for a more in-depth study of privacy attitudes and behaviours of undergraduates.

We also establish the need for more information about undergraduate privacy behaviours and attitudes, and present a quantitative study as an initial step towards the discovery of relevant data. Results indicate that there is a contextual element to some privacy attitudes

and behaviours, and also that there is considerable work to be done in helping undergraduates to understand the privacy norms which govern what the university does with their privacy.

9.2 Future work

There remains an ever-growing amount of scholarship which can be pursued in support of the contributions of this thesis. In this section, I will highlight some of the most interesting threads to trace.

9.2.1 Using the framework to operationalize contextual privacy

A conceptual framework is not a completed system, although it is an important step towards building systems which understand context and use it to protect the privacy of individuals. There remains a significant opportunity for researchers to build upon the framework to create operational systems by integrating pre-existing techniques from usable privacy, ontologies, and reasoning, and to create new techniques to support what has not yet already been built, if necessary.

There is a gap in understanding how to derive implicit norms from user preferences. By creating the vocabulary which allows us to understand how norms fit in context, we lay the groundwork for future work in this area. Of particular interest is determining how to handle implicit norms which change over time, a phenomenon which has been observed repeatedly over the course of this thesis, but which has lain outside the scope of my work.

9.2.2 Managing privacy issues in post-secondary institutions

This thesis provides an overview of how student data is collected and used by post-secondary institutions, but this overview is by no means exhaustive. With data processing techniques continuing to develop and improve, and with the pressure on both public and private institutions to accomplish more with fewer resources, the collection and use of student data by

institutions will also grow in volume and sophistication.

There also needs to be an increased focus on the role of third parties in the educational landscape, such as software vendors (who provide productivity suites, learning management systems, and other tools), academic publishers, who increasingly are concentrating lines of business dependent on data from post-secondary institutions, and other interested organizations, all of whom view the data housed in post-secondary institutions as a valuable resource. There is room for deeper study, to interrogate the ethics of this sort of data use, and to have a conversation about the implications of this data use for our students and our society as a whole.

9.2.3 Gaining a better understanding of the role of privacy in undergraduate life

The study described in Chapter 8 was a first step in better understanding how the contemporary undergraduate at the University of Calgary views their privacy, both on and off-campus. There is room for considerable further study. It would be useful to gather more accurate quantitative data which could be segmented in numerous ways from a larger body of respondents. A missing part of this study is the lack of qualitative data, which could be gathered via interviews and focus groups, and help researchers identify the privacy issues which most concern undergraduates.

Studying students at a wider range of institutions, especially those with a different situation than observed at the University of Calgary (a large Canadian research-intensive institution, situated in a primarily suburban setting with only a small proportion of students living in residence) would also help to expand our understanding of what privacy should mean for students. Especially notable are those at institutions with a significant residential population, in different countries (where the cultural norms with respect to privacy may differ), or at institutions with a teaching rather than a research focus.

9.3 Closing thoughts

Students are at a unique point in their lives when they arrive at a university to study, and in 2019, they are at a unique point in the development of the technologies with which they interact on a daily basis. Understanding how their privacy is at risk allows us to extrapolate how privacy will be at risk for the wider community in the coming years.

Bibliography

[06] *First nations conceptual frameworks and applied models on ethics, privacy, and consent in health research and information summary report.* eng. Ottawa, Ont.: National Aboriginal Health Organization (NAHO) = Organisation nationale de la santé autochtone (ONSA) = Kanatami Nunaqaakaaqsimajunut Aanniarnanginnilirinirmut Katujjiqatigiit, 2006.

[AG05] A. Acquisti and J. Grossklags. "Privacy and rationality in individual decision making". In: *IEEE security & privacy* 3.1 (2005), pp. 26–33.

[AG06] A. Acquisti and R. Gross. "Imagined communities: Awareness, information sharing, and privacy on the Facebook". In: *International workshop on privacy enhancing technologies*. Springer. 2006, pp. 36–58.

[Agr+02] R. Agrawal et al. "Hippocratic databases". In: *VLDB '02: Proceedings of the 28th international conference on Very Large Data Bases*. Hong Kong, China: VLDB Endowment, 2002, pp. 143–154.

[Alb00] G. of Alberta. *Freedom of Information and Protection of Privacy Act*. 2000.

[AP12] K. E. Arnold and M. D. Pistilli. "Course signals at Purdue: Using learning analytics to increase student success". In: *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM. 2012, pp. 267–270.

[ARK00] G. R. Adams, B. A. Ryan, and L. Keating. "Family relationships, academic environments, and psychosocial development during the university experience: A longitudinal investigation". In: *Journal of Adolescent Research* 15.1 (2000), pp. 99–122.

[Ass48] U. G. Assembly. "Universal declaration of human rights". In: *UN General Assembly* (1948).

[Ban+11] M. Banerjee et al. "Quantifying privacy violations". In: *Proceedings of the 8th VLDB international conference on Secure data management*. SDM'11. Seattle, WA: Springer-Verlag, 2011, pp. 1–17. ISBN: 978-3-642-23555-9.

[Bar+06] A. Barth et al. "Privacy and contextual integrity: framework and applications". In: *Security and Privacy, 2006 IEEE Symposium on*. May 2006, pp. 184–198. DOI: 10.1109/SP.2006.32.

[Bar+09] K. Barker et al. “A Data Privacy Taxonomy”. In: *Dataspace: The Final Frontier*. Ed. by A. Sexton. Vol. 5588. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009, pp. 42–54. ISBN: 978-3-642-02842-7. URL: http://dx.doi.org/10.1007/978-3-642-02843-4_7.

[Bar12] L. Barkhuus. “The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2012, pp. 367–376.

[Bet+02] C. Bettini et al. “Provisions and obligations in policy management and security applications”. In: *Proceedings of the 28th international conference on Very Large Data Bases*. VLDB ’02. Hong Kong, China: VLDB Endowment, 2002, pp. 502–513. URL: http://dl.acm.org/citation.cfm?id=1287369.1287413.

[BGN17] S. Benthall, S. Gürses, and H. Nissenbaum. “Contextual Integrity through the Lens of Computer Science”. In: *Foundations and Trends in Privacy and Security* 2.1 (2017), pp. 1–69.

[BK00] M. D. Berzonsky and L. S. Kuk. “Identity status, identity processing style, and the transition to university”. In: *Journal of adolescent research* 15.1 (2000), pp. 81–98.

[Blu79] J. G. Blumler. “The role of theory in uses and gratifications studies”. In: *Communication research* 6.1 (1979), pp. 9–36.

[BT16] D. C. Brooks and T.-L. B. Thayer. “Institutional analytics in higher education”. In: *Research report*. ECAR, 2016.

[Cal17] U. of Calgary. *Privacy Policy*. https://www.ucalgary.ca/policies/files/policies/privacy-policy.pdf. Accessed September 27 2019. 2017.

[Cal19a] U. of Calgary. *Code of Conduct*. https://www.ucalgary.ca/policies/files/policies/code-of-conduct.pdf. Accessed September 27 2019. 2019.

[Cal19b] U. of Calgary. *University of Calgary: K. Statement on Principles of Conduct*. https://www.ucalgary.ca/pubs/calendar/current/k.html. Accessed September 27 2019. 2019.

[Can00] G. of Canada. *Personal Information Protection and Electronic Documents Act*. http://laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html. Accessed July 26 2019. 2000.

[Can19] Canadian Broadcasting Corporation. *MPs warn Facebook’s Zuckerberg and Sandberg could be found in contempt of Parliament for no-show — CBC News*. Accessed December 19 2019. May 2019. URL: https://www.cbc.ca/news/politics/facebook-contempt-parliament-1.5145347.

[Cra+02] L. Cranor et al. *The Platform for Privacy Preferences 1.0 (P3P 1.0) Specification*. http://www.w3.org/TR/P3P/. Accessed 2010-07-10. 2002.

[CVM14] X. Chen, M. Vorvoreanu, and K. Madhavan. “Mining social media data for understanding students’ learning experiences”. In: *IEEE Transactions on Learning Technologies* 7.3 (2014), pp. 246–259.

[Dah+15] E. Dahlstrom et al. *ECAR Study of Undergraduate Students and Information Technology, 2015.* Tech. rep. ECAR, 2015.

[DAS01] A. K. Dey, G. D. Abowd, and D. Salber. "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications". In: *Human–Computer Interaction* 16.2-4 (2001), pp. 97–166.

[Dat+11] A. Datta et al. "Understanding and protecting privacy: Formal semantics and principled audit mechanisms". In: *International Conference on Information Systems Security*. Springer. 2011, pp. 1–27.

[Dol+16] R. Dolan et al. "Social media engagement behaviour: a uses and gratifications perspective". In: *Journal of Strategic Marketing* 24.3-4 (2016), pp. 261–277.

[Dwo+06] C. Dwork et al. "Calibrating noise to sensitivity in private data analysis". In: *Theory of cryptography conference*. Springer. 2006, pp. 265–284.

[Dwo06] C. Dwork. "Differential Privacy". In: *Automata, Languages and Programming*. Ed. by M. Bugliesi et al. Vol. 4052. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, pp. 1–12. ISBN: 978-3-540-35907-4. URL: http://dx.doi.org/10.1007/11787006_1.

[Fai05] P. Fairfield. *Public/private*. eng. Lanham, Md.: Rowman & Littlefield Publishers, Inc., 2005. ISBN: 0742549577.

[FC17] R. Ferguson and D. Clow. "Where is the evidence?: a call to action for learning analytics". In: *Proceedings of the seventh international conference on learning analytics & knowledge*. ACM. 2017, pp. 56–65.

[FCK95] D. Ferraiolo, J. Cugini, and D. R. Kuhn. "Role-based access control (RBAC): Features and motivations". In: *Proceedings of 11th annual computer security application conference*. 1995, pp. 241–48.

[Hor81] M. J. Horwitz. "History of the public/private distinction". In: *U. Pa. L. Rev.* 130 (1981), p. 1423.

[Hul+04] R. Hull et al. "Enabling context-aware and privacy-conscious user data sharing". In: *Mobile Data Management, 2004. Proceedings. 2004 IEEE International Conference on*. IEEE. 2004, pp. 187–198.

[Ifi16] P. Ifinedo. "Applying uses and gratifications theory and social influence processes to understand students' pervasive adoption of social networking sites: Perspectives from the Americas". In: *International Journal of Information Management* 36.2 (2016), pp. 192–206.

[IS16] D. Ifenthaler and C. Schumacher. "Student perceptions of privacy principles for learning analytics". In: *Educational Technology Research and Development* 64.5 (2016), pp. 923–938.

[Jaf+11] M. Jafari et al. "Towards defining semantic foundations for purpose-based privacy policies". In: *Proceedings of the first ACM conference on Data and application security and privacy*. CODASPY '11. San Antonio, TX, USA: ACM, 2011, pp. 213–224. ISBN: 978-1-4503-0466-5. DOI: 10.1145/1943513.1943541. URL: http://doi.acm.org/10.1145/1943513.1943541.

[JL02] X. Jiang and J. A. Landay. “Modeling privacy control in context-aware systems”. In: *IEEE Pervasive computing* 1.3 (2002), pp. 59–63.

[Jon+19] K. M. L. Jones et al. “In Their Own Words: Students Perspectives on Privacy ad Library Participation in Learning Analytics Initiatives”. In: *Recasting the Narrative: The Proceedings of the ACRL 2019 Conference*. American Library Assocation. 2019.

[Kli+10] T. A. Klimstra et al. “Identity formation in adolescence: Change or stability?” In: *Journal of Youth and Adolescence* 39.2 (2010), pp. 150–162.

[Kok17] S. Kokolakis. “Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon”. In: *Computers & security* 64 (2017), pp. 122–134.

[LE17] A. M. B. Lips and E. A. Eppel. “Understanding and explaining online personal information-sharing behaviours of New Zealanders: a new taxonomy”. In: *Information, Communication & Society* 20.3 (2017), pp. 428–443.

[LeF+04] K. LeFevre et al. “Limiting Disclosure in Hippocratic Databases”. In: *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*. VLDB ’04. Toronto, Canada: VLDB Endowment, 2004, pp. 108–119. ISBN: 0-12-088469-0. URL: http://dl.acm.org/citation.cfm?id=1316689.1316701.

[LK17] H. Lee and A. Kobsa. “Privacy preference modeling and prediction in a simulated campuswide IoT environment”. In: *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE. 2017, pp. 276–285.

[Lom19] N. Lomas. *Most EU cookie ’consent’ notices are meaningless or manipulative, study finds*. Aug. 2019. URL: https://techcrunch.com/2019/08/10/most-eu-cookie-consent-notices-are-meaningless-or-manipulative-study-finds/.

[Mac+07] A. Machanavajjhala et al. “L-diversity : Privacy beyond k-anonymity”. In: *ACM Trans. Knowl. Discov. Data* 1.1 (Mar. 2007). ISSN: 1556-4681. DOI: 10.1145/1217299.1217302. URL: http://doi.acm.org/10.1145/1217299.1217302.

[MM92] N. B. McCormick and J. W. McCormick. “Computer friends and foes: Content of undergraduates’ electronic mail”. In: *Computers in Human Behavior* 8.4 (1992), pp. 379–405.

[MR15] M. Madden and L. Rainie. *Americans’ attitudes about privacy, security and surveillance*. Pew Research Center, 2015.

[Nae+17] P. E. Naeini et al. “Privacy expectations and preferences in an IoT world”. In: *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. 2017, pp. 399–412.

[Ni+07] Q. Ni et al. “Privacy-aware role based access control”. In: *Proceedings of the 12th ACM symposium on Access control models and technologies*. SACMAT ’07. Sophia Antipolis, France: ACM, 2007, pp. 41–50. ISBN: 978-1-59593-745-2. DOI: 10.1145/1266840.1266848. URL: http://doi.acm.org/10.1145/1266840.1266848.

[Nis04] H. Nissenbaum. “Privacy as Contextual Integrity”. In: *Washington Law Review* 79.1 (2004).

[Nis09] H. Nissembaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto, CA, USA: Stanford University Press, 2009.

[OEC80] OECD. *Guidelines on the Protection of Privacy and Transborder Flow of Personal Data*. https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm. Accessed March 1 2017. 1980.

[Pad+09] J. Padma et al. “Hippocratic PostgreSQL”. In: *Data Engineering, 2009. ICDE’09. IEEE 25th International Conference on*. IEEE. 2009, pp. 1555–1558.

[Pad19] D. Paddon. *Canada’s telecom industry faces challenge from deluge of spam phone calls*. Accessed December 19 2019. Nov. 2019. URL: https://www.ctvnews.ca/business/canada-s-telecom-industry-faces-challenge-from-deluge-of-spam-phone-calls-1.4673577.

[Pen03] S.-y. Peng. “Privacy and the construction of legal meaning in Taiwan”. In: *Int’l L.* Vol. 37. HeinOnline. 2003, p. 1037.

[Pri17] P. Prinsloo. “Fleeing from Frankenstein’s monster and meeting Kafka on the way: Algorithmic decision-making in higher education”. In: *E-Learning and Digital Media* 14.3 (2017), pp. 138–163. DOI: 10.1177/2042753017731355.

[PWC14] M. D. Pistilli, J. E. Willis, and J. P. Campbell. “Analytics through an institutional lens: Definition, theory, design, and impact”. In: *Learning analytics*. Springer, 2014, pp. 79–102.

[Reg16] G. D. P. Regulation. “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46”. In: *Official Journal of the European Union (OJ)* 59.1-88 (2016), p. 294.

[Rob+16] L. D. Roberts et al. “Student attitudes toward learning analytics in higher education: “the fitbit version of the learning world”””. In: *Frontiers in psychology* 7 (2016), p. 1959.

[Sai] Saint Louis University. *Alexa and Saint Louis University*. https://www.slu.edu/alexa. Accessed April 24 2019.

[Sai18] Saint Louis University. *SLU Alexa Project*. https://www.slu.edu/news/2018/august/slu-alexa-project.php. Accessed April 24 2019. Aug. 2018.

[San+96] R. S. Sandhu et al. “Role-based access control models”. In: *Computer* 29.2 (1996), pp. 38–47.

[SB12] G. Siemens and R. S. d Baker. "Learning analytics and educational data mining: towards communication and collaboration". In: *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM. 2012, pp. 252–254.

[SK10] F. Stutzman and J. Kramer-Duffield. "Friends only: examining a privacy-enhancing behavior in facebook". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2010, pp. 1553–1562.

[SOW95] J. M. Sever, A. M. O'Grady, and A. F. Westin. *Equifax-Harris mid-decade consumer privacy survey, 1995*. Louis Harris and Associates, 1995.

[SP13] S. Slade and P. Prinsloo. "Learning analytics: Ethical issues and dilemmas". In: *American Behavioral Scientist* 57.10 (2013), pp. 1510–1529.

[Swe02] L. Sweeney. "k-Anonymity: A Model for Protecting Privacy." In: *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems* 10.5 (2002), p. 557. ISSN: 02184885.

[The12] The White House. *Consumer Privacy Bill of Rights*. 2012.

[Tin87] V. Tinto. *Leaving college: Rethinking the causes and cures of student attrition*. ERIC, 1987.

[Vib+18] O. Viberg et al. "The Current Landscape of Learning Analytics in Higher Education". In: *Computers in Human Behavior* (2018).

[WB90] S. D. Warren and L. D. Brandeis. "The right to privacy". In: *Harvard Law Review* 5 (1890).

[Wes91] A. F. Westin. "Harris-Equifax consumer privacy survey 1991". In: *Atlanta, GA: Equifax Inc* (1991).

[Wij+15] P. Wijesekera et al. "Android Permissions Remystified: A Field Study on Contextual Integrity." In: *USENIX Security Symposium*. 2015, pp. 499–514.

[Wil18] B. Williamson. "The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'". In: *International Journal of Educational Technology in Higher Education* 15.1 (Mar. 2018), p. 12. ISSN: 2365-9440. DOI: 10.1186/s41239-018-0094-1. URL: https://doi.org/10.1186/s41239-018-0094-1.

[Yor19] York University. *York U & IBM launch AI-powered student support pilot*. https://news.yorku.ca/2019/03/21/york-ibm-launch-ai-powered-pilot/. Accessed April 24 2019. Mar. 2019.

Appendix A

Text of Survey Questions

SURVEY QUESTIONS

1. Technology use

Do you use the following (never use/monthly/weekly/several times a week/everyday):

- A desktop computer
- A laptop computer
- Smartphone
- Tablet
- Wearable such as a fitness tracker or smartwatch
- “Smart home” devices such as Google Home or other “smart speakers”, internet-connected appliances or devices like smart lightbulbs, doorbells, etc.

Do you use any of the following at school:

- Desktop computer
- Laptop computer
- Smartphone
- Tablet
- Wearable

Do you use the following (never use/monthly/weekly/several times a week/everyday):

- Facebook
- Twitter
- Instagram
- Google+
- Snapchat
- Whatsapp
- Signal
- WeChat
- WeiBo
- Other, please specify: _____

Have you (never use/monthly/weekly/several times a week/everyday):

- Streamed content from digital providers such as Netflix or Spotify
- Searched for content online from search engines such as Google or Microsoft Bing
- Bought items from online commerce sites such as Amazon or eBay
- Bought anything online from retailers with both a physical “bricks and mortars” store and an online presence?
- Accessed a bank via their website or a mobile app

- Accessed government services via their website or a mobile app (for example, Calgary's 311 service, Calgary ParkPlus, Alberta registry services, or filing taxes online?)
- Accessed an online broker for services, such as Uber, SkipTheDishes, or AirBnB?

Which of the following campus services have you used in the last twelve months (Yes/No/ N/A):

- Active Living
- Library (online or physical)
- Campus wi-fi
- Computer labs (such as the Learning Commons in TFDL)
- Food services (Dining Centre, Tim Hortons, other outlets)
- Loaded money onto your Unicard
- Accessed a campus space using your Unicard
- Accessed a campus printer using your Unicard

Which of the following University-provided technologies have you used in the last twelve months:

- D2L
- Office 365 (including webmail)
- Top Hat
- Other (fill in)

Students and privacy

To what extent do you agree or disagree with the following statements? Please use the provided scale, where 1 means strongly disagree, and 7 means strongly agree. You may indicate where you are unsure or don't wish to answer.

- The University's privacy policy is easy to find
- The University's privacy policy protects my privacy
- I remember consenting to the University's privacy policy
- My data is well protected by the University
- The University is careful about how it reuses my data
- I trust student advisors with my personal data
- I trust my professors with my personal data
- I trust my TAs with my personal data
- I trust other University staff with my personal data
- I trust student organizations that I belong to with my personal data
- I trust my classmates with my personal data
- The University should be allowed to use my personal data if it helps me with my degree
- The University should be allowed to use my personal data if it helps to improve my well-being
- The University should be allowed to use my personal data if it helps to improve the quality of my classes

- The University should be allowed to use my personal data if it helps to improve my experience at the U of C

Privacy and the “real world”

To what extent do you agree or disagree with the following statements? Please use the provided scale, where 1 means strongly disagree, and 7 means strongly agree. You may indicate where you are unsure or don’t wish to answer.

- I trust the government with my personal data
- I trust financial institutions with my personal data
- I trust my healthcare providers with my personal data
- I trust companies with whom I do business offline with my personal data
- I trust companies with whom I do business online with my personal data
- I trust search providers that I use to search the web with my personal data
- I trust my social media services with my personal data
- I consider my personal data private no matter which organization uses it
- I am aware that some companies use my private data for more than one thing
- I don’t mind if companies use my private data for more than one thing, as long as it’s clearly stated in their privacy policy.
- It disturbs me when a company uses my private data for more than one thing
- Sometimes I am okay with trusting one company with my private data, but not another.

Demographics:

I am:

- 18 to 20
- 20 to 22
- 22 to 25
- 25 to 30
- 30 to 35
- Older than 35
- Don’t wish to specify

I am:

- Female
- Male
- Other
- Don’t wish to specify

In years of study, I am in my:

- First year

- Second year
- Third year
- Fourth year
- Fifth or later
- Don’t wish to specify

My program belongs to:

- (list of faculties here)
- Don’t wish to specify
-

I live:

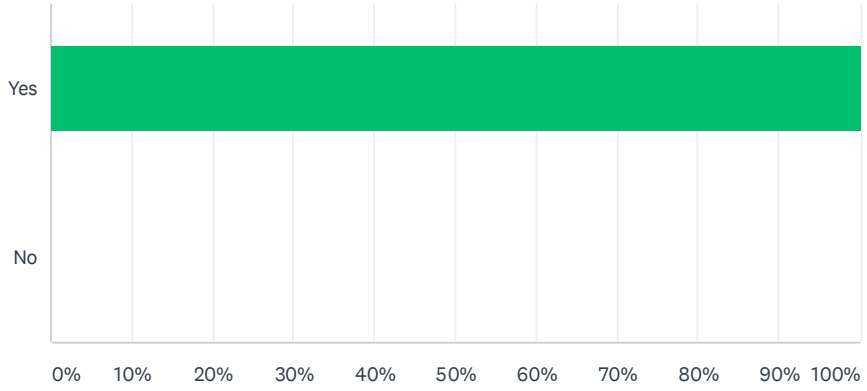
- Off-campus, with family
- Off-campus, with roommates
- On-campus
- Other or don’t wish to specify

Appendix B

Full data and results for privacy study

Q1 Do you consent to take this survey?

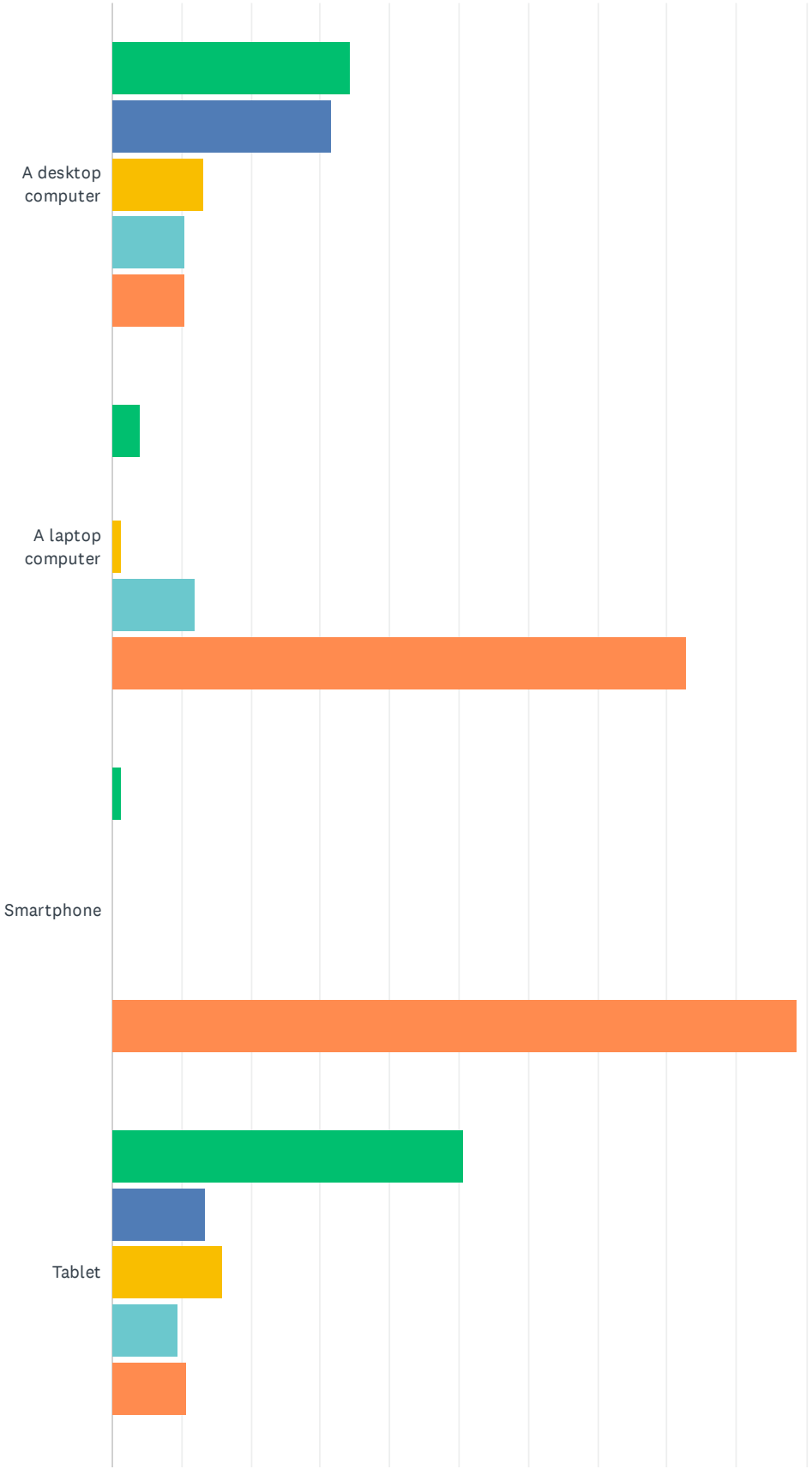
Answered: 80 Skipped: 0

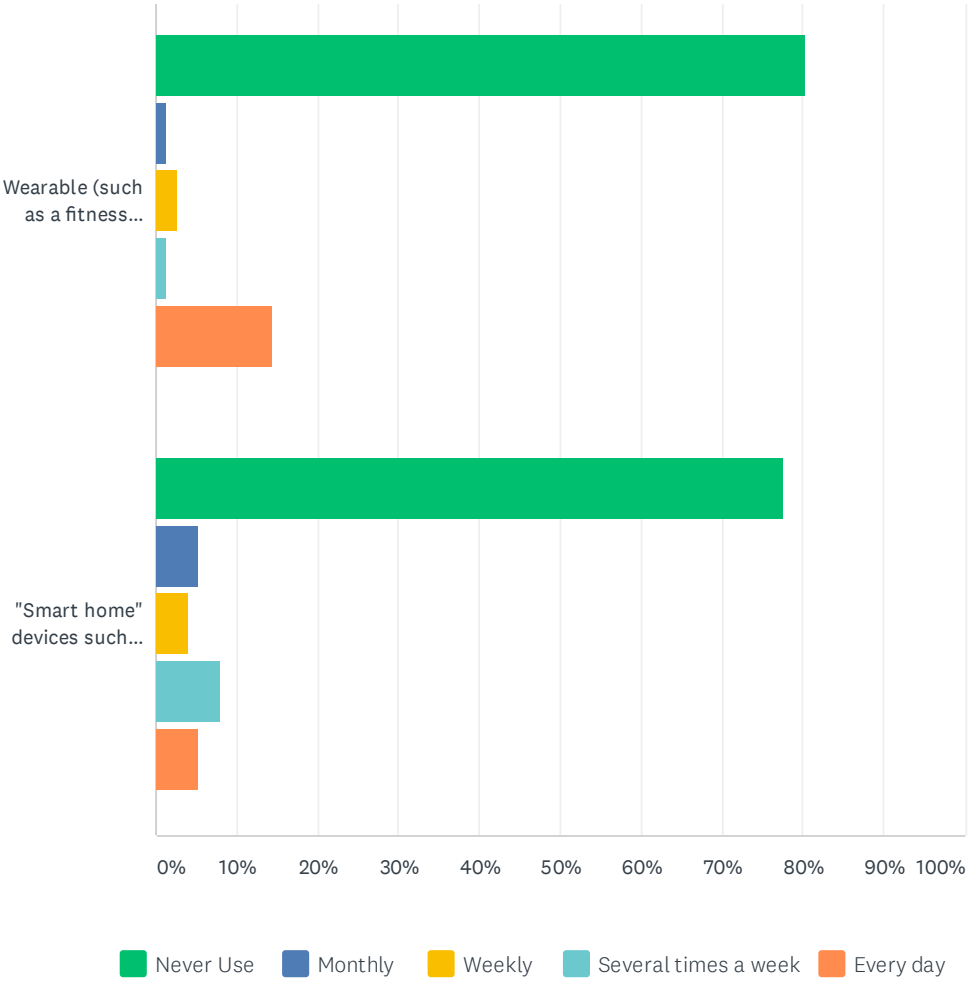


ANSWER CHOICES	RESPONSES	
Yes	100.00%	80
No	0.00%	0
TOTAL		80

Q2 How often do you use the following?

Answered: 76 Skipped: 4

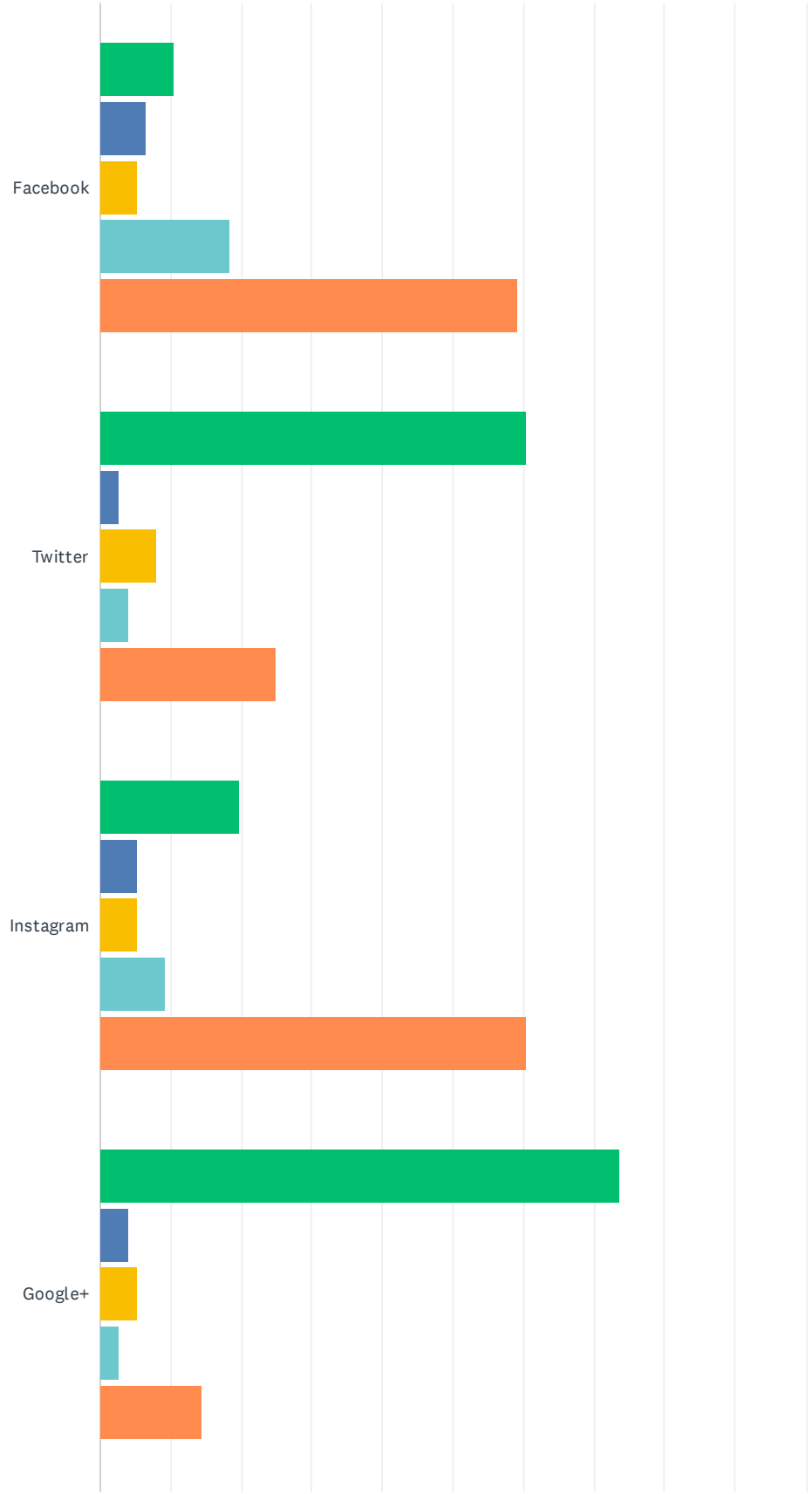


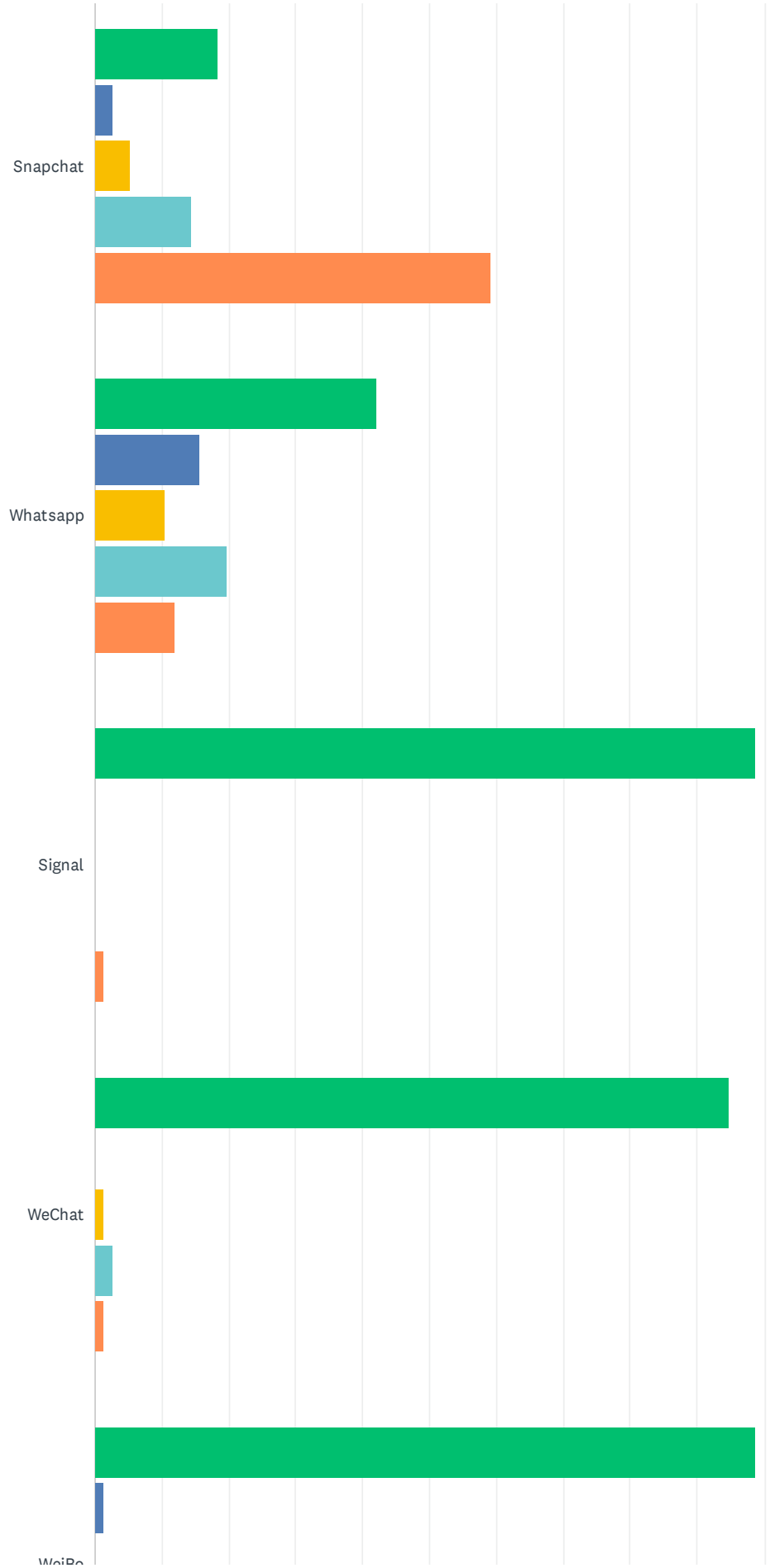


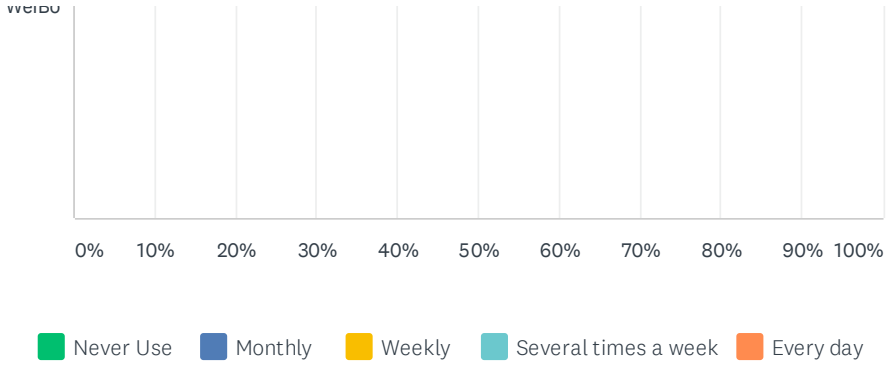
	NEVER USE	MONTHLY	WEEKLY	SEVERAL TIMES A WEEK	EVERY DAY	TOTAL
A desktop computer	34.21% 26	31.58% 24	13.16% 10	10.53% 8	10.53% 8	76
A laptop computer	3.95% 3	0.00% 0	1.32% 1	11.84% 9	82.89% 63	76
Smartphone	1.32% 1	0.00% 0	0.00% 0	0.00% 0	98.68% 75	76
Tablet	50.67% 38	13.33% 10	16.00% 12	9.33% 7	10.67% 8	75
Wearable (such as a fitness tracker or a smartwatch)	80.26% 61	1.32% 1	2.63% 2	1.32% 1	14.47% 11	76
"Smart home" devices such as Google Home, Amazon Echo, or other "smart speakers," internet connected appliances or devices such as smart lightbulbs, doorbells, etc	77.63% 59	5.26% 4	3.95% 3	7.89% 6	5.26% 4	76

Q3 Do you use the following

Answered: 76 Skipped: 4



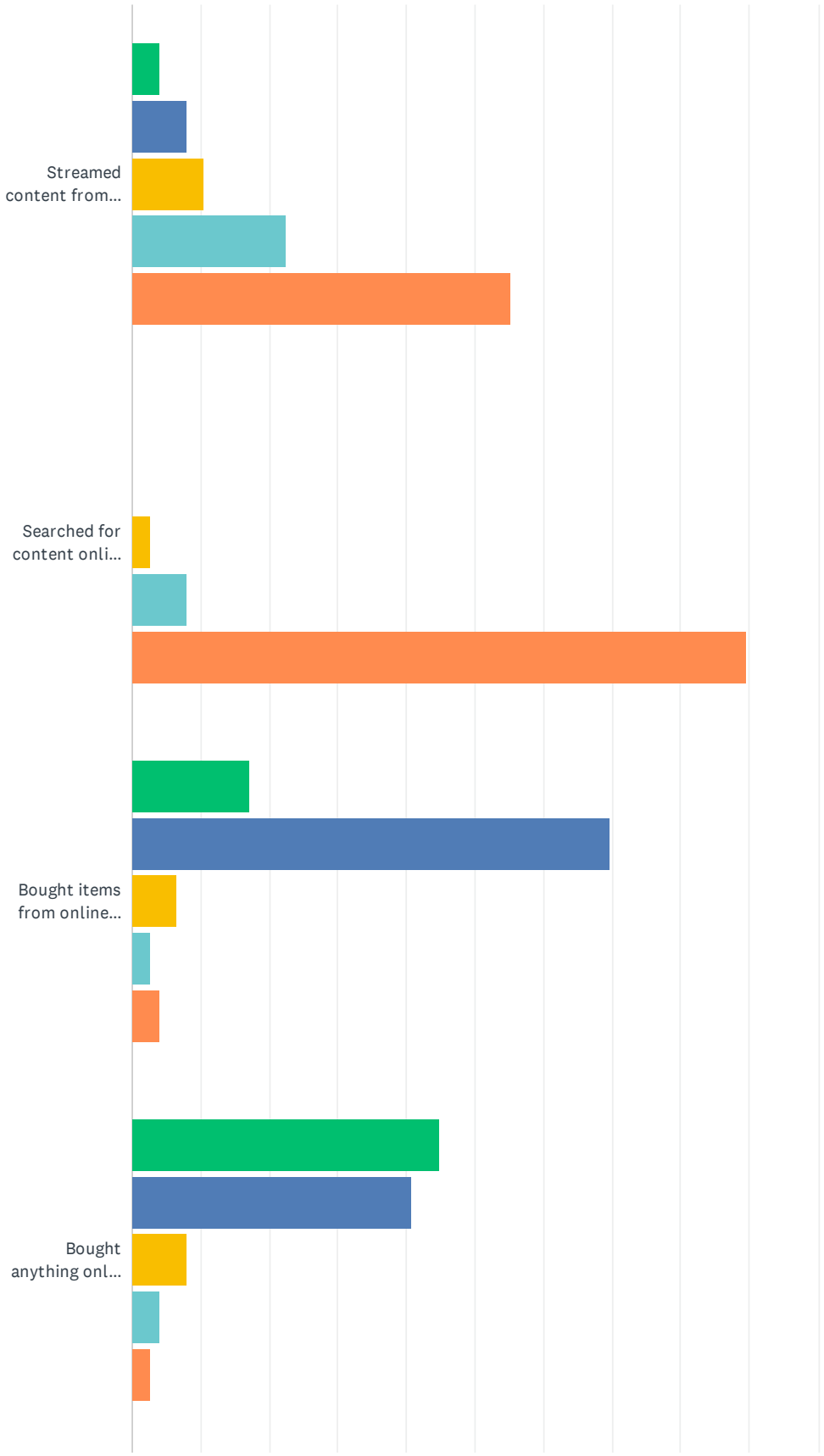

Snapchat
Whatsapp
Signal
WeChat
WeiBo

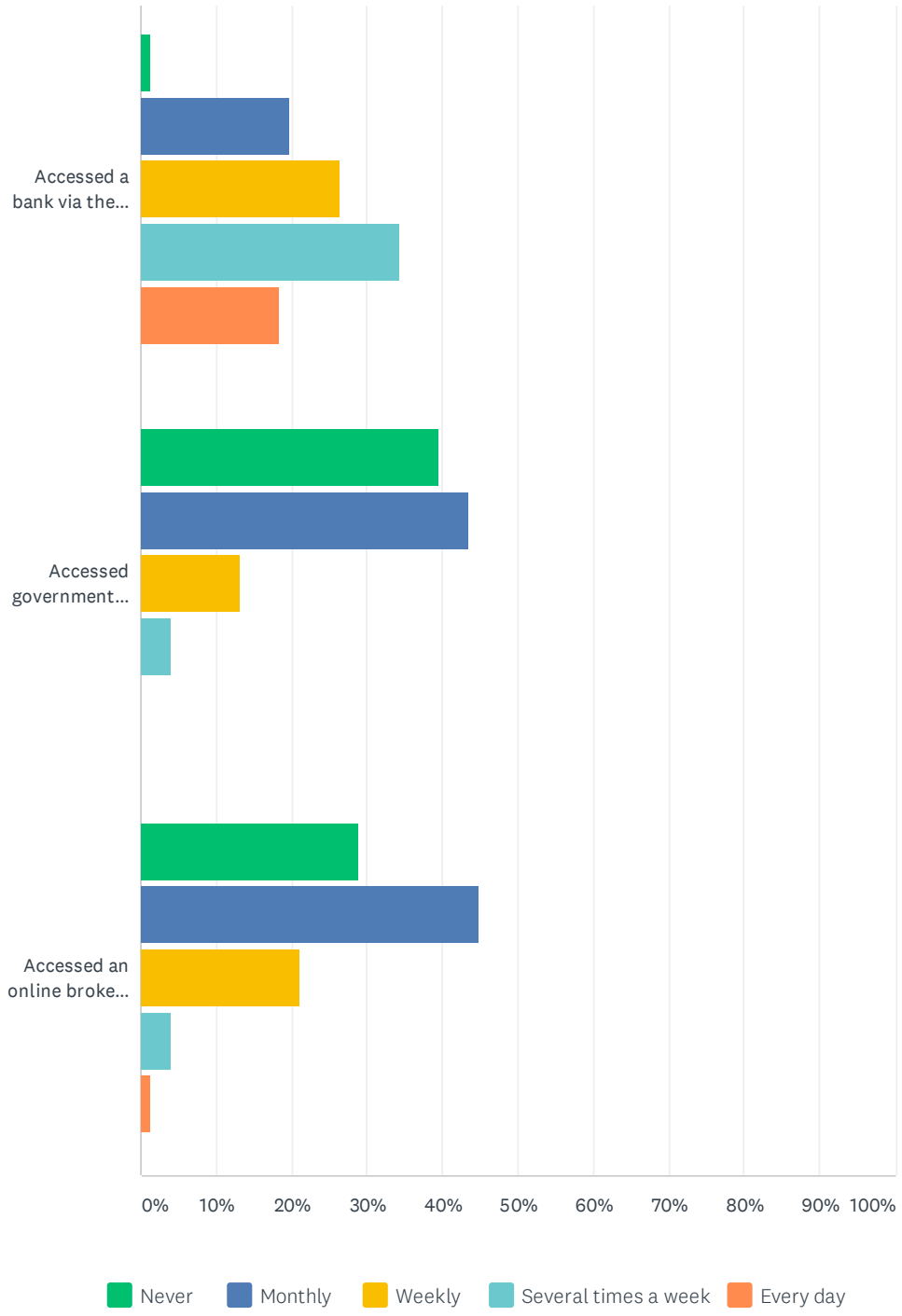


	NEVER USE	MONTHLY	WEEKLY	SEVERAL TIMES A WEEK	EVERY DAY	TOTAL
Facebook	10.53% 8	6.58% 5	5.26% 4	18.42% 14	59.21% 45	76
Twitter	60.53% 46	2.63% 2	7.89% 6	3.95% 3	25.00% 19	76
Instagram	19.74% 15	5.26% 4	5.26% 4	9.21% 7	60.53% 46	76
Google+	73.68% 56	3.95% 3	5.26% 4	2.63% 2	14.47% 11	76
Snapchat	18.42% 14	2.63% 2	5.26% 4	14.47% 11	59.21% 45	76
Whatsapp	42.11% 32	15.79% 12	10.53% 8	19.74% 15	11.84% 9	76
Signal	98.68% 75	0.00% 0	0.00% 0	0.00% 0	1.32% 1	76
WeChat	94.74% 72	0.00% 0	1.32% 1	2.63% 2	1.32% 1	76
WeiBo	98.68% 75	1.32% 1	0.00% 0	0.00% 0	0.00% 0	76

Q4 How frequently have you

Answered: 76 Skipped: 4

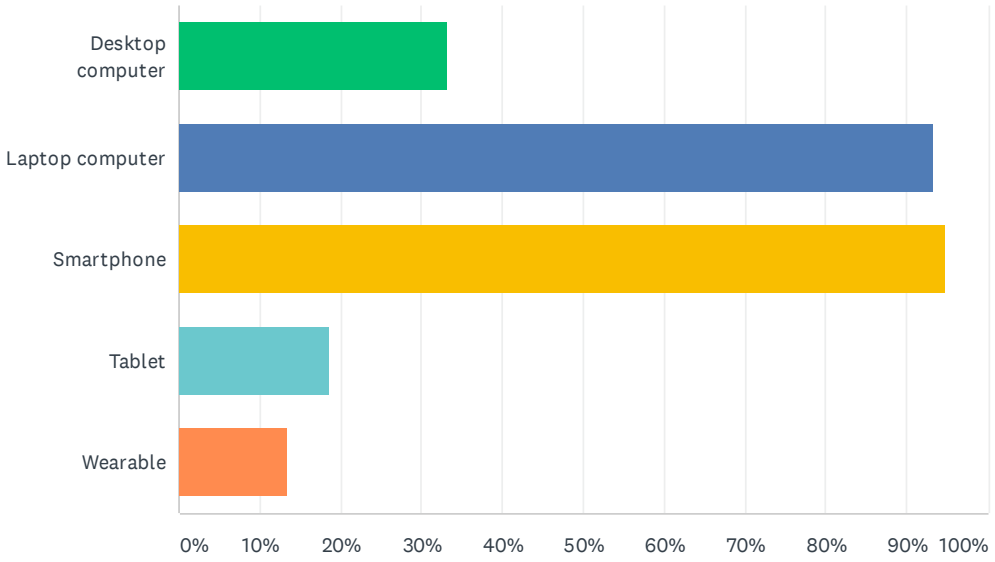



Accessed a bank via the...
Accessed government...
Accessed an online broke...
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Never
Monthly
Weekly
Several times a week
Every day

	NEVER	MONTHLY	WEEKLY	SEVERAL TIMES A WEEK	EVERY DAY	TOTAL
Streamed content from digital providers such as Netflix or Spotify	3.95% 3	7.89% 6	10.53% 8	22.37% 17	55.26% 42	76
Searched for content online from search engines such as Google or Microsoft Bing	0.00% 0	0.00% 0	2.63% 2	7.89% 6	89.47% 68	76
Bought items from online commerce sites such as Amazon or eBay	17.11% 13	69.74% 53	6.58% 5	2.63% 2	3.95% 3	76
Bought anything online from retailers with both a physical “bricks and mortars” store and an online presence?	44.74% 34	40.79% 31	7.89% 6	3.95% 3	2.63% 2	76
Accessed a bank via their website or a mobile app	1.32% 1	19.74% 15	26.32% 20	34.21% 26	18.42% 14	76
Accessed government services via their website or a mobile app (for example, Calgary’s 311 service, Calgary ParkPlus, Alberta registry services, or filing taxes online?)	39.47% 30	43.42% 33	13.16% 10	3.95% 3	0.00% 0	76
Accessed an online broker for services, such as Uber, SkipTheDishes, or AirBnB?	28.95% 22	44.74% 34	21.05% 16	3.95% 3	1.32% 1	76

Q5 Do you use any of the following at school?

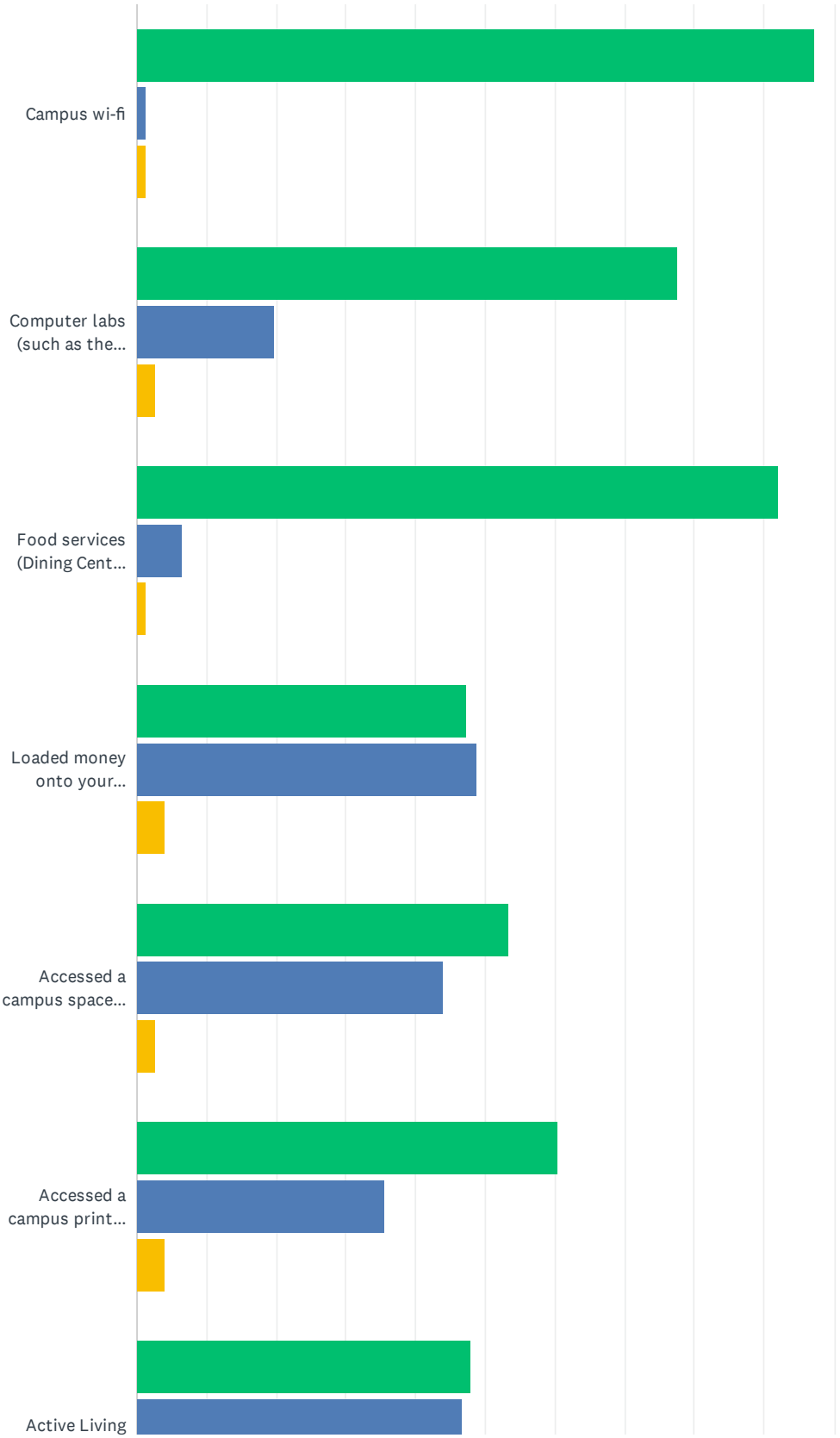
Answered: 75 Skipped: 5

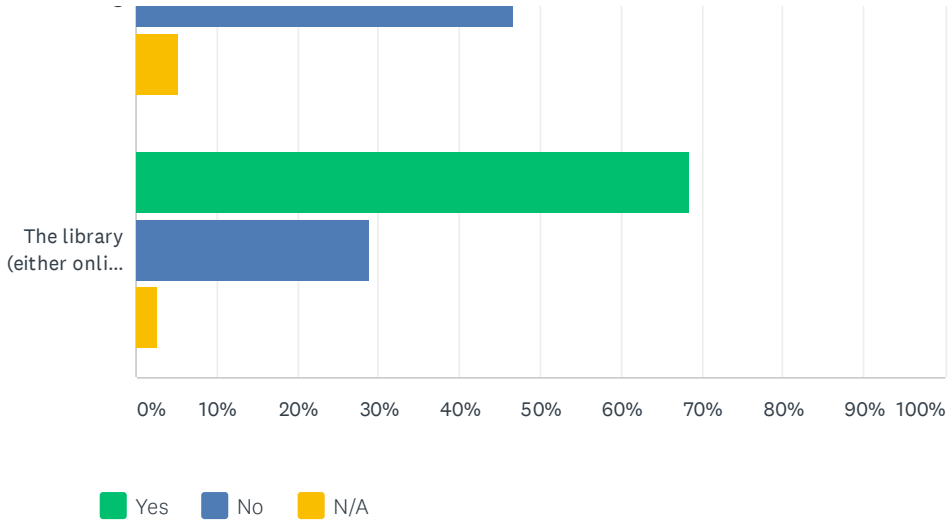


ANSWER CHOICES	RESPONSES	
Desktop computer	33.33%	25
Laptop computer	93.33%	70
Smartphone	94.67%	71
Tablet	18.67%	14
Wearable	13.33%	10
Total Respondents: 75		

Q6 Which of the following campus services have you used in the last twelve months?

Answered: 76 Skipped: 4

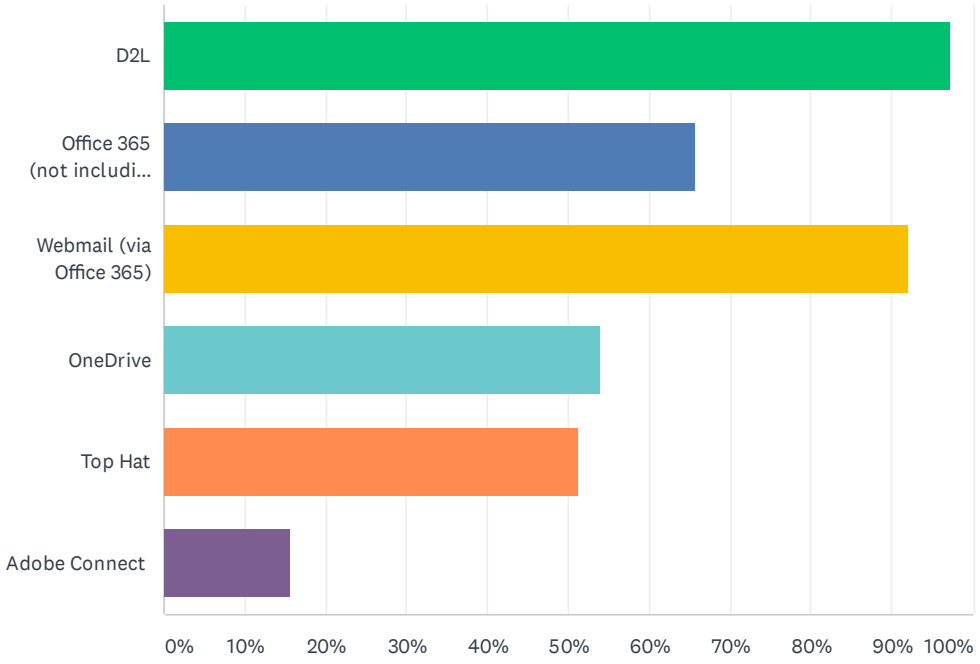




	YES	NO	N/A	TOTAL
Campus wi-fi	97.37% 74	1.32% 1	1.32% 1	76
Computer labs (such as the Learning Commons in TFDL)	77.63% 59	19.74% 15	2.63% 2	76
Food services (Dining Centre, Tim Hortons, other outlets)	92.11% 70	6.58% 5	1.32% 1	76
Loaded money onto your Unicard	47.37% 36	48.68% 37	3.95% 3	76
Accessed a campus space using your Unicard	53.33% 40	44.00% 33	2.67% 2	75
Accessed a campus printer using your Unicard	60.53% 46	35.53% 27	3.95% 3	76
Active Living	48.00% 36	46.67% 35	5.33% 4	75
The library (either online or physically accessing resources)	68.42% 52	28.95% 22	2.63% 2	76

Q7 Which of the following University-provided technologies have you used in the last twelve months:

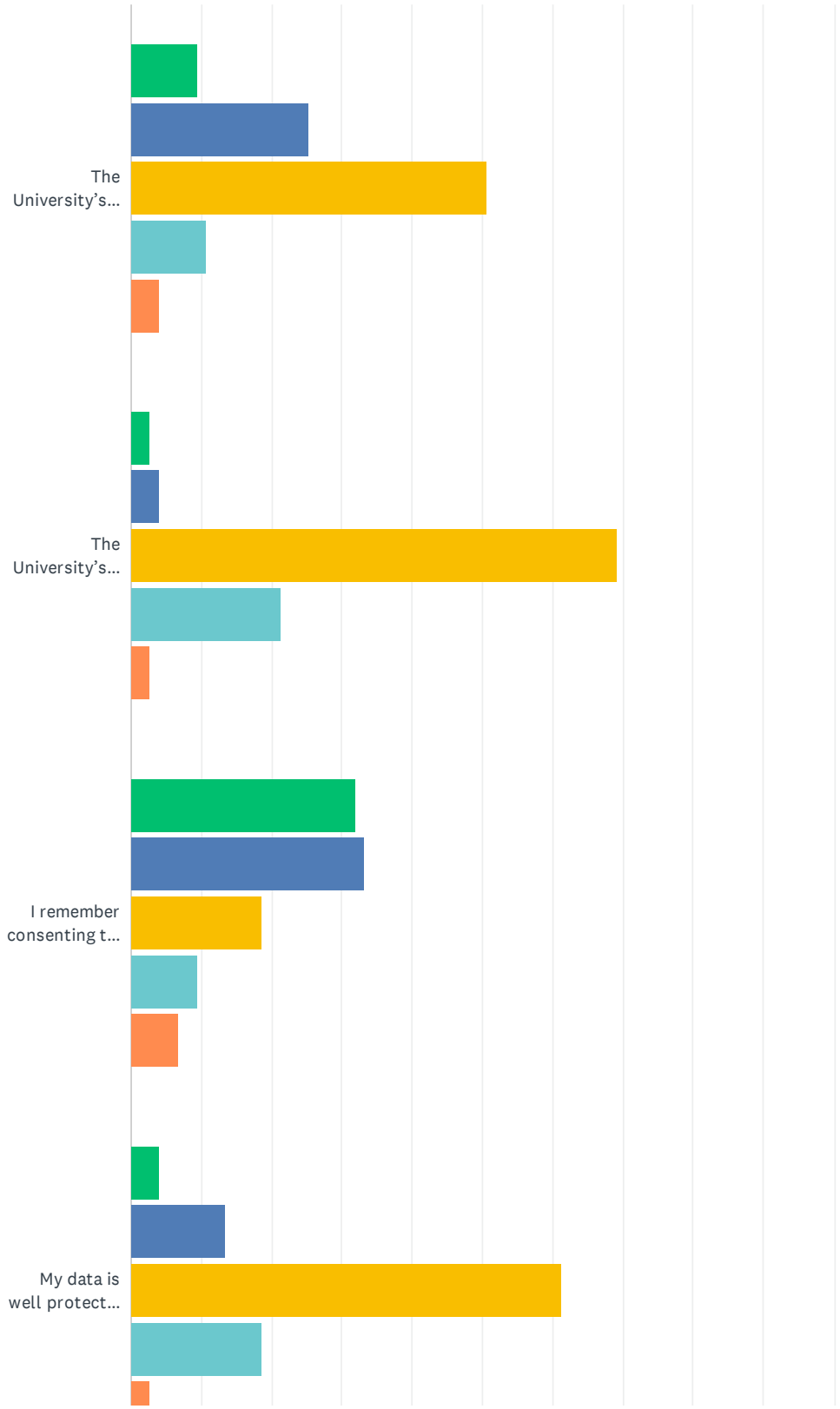
Answered: 76 Skipped: 4

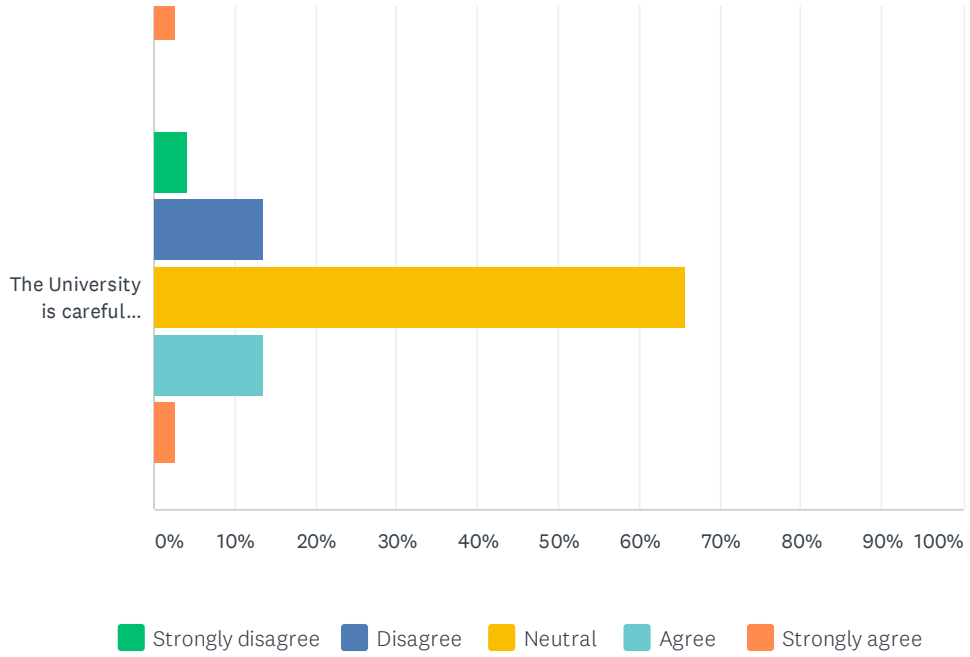


ANSWER CHOICES	RESPONSES	
D2L	97.37%	74
Office 365 (not including webmail or OneDrive)	65.79%	50
Webmail (via Office 365)	92.11%	70
OneDrive	53.95%	41
Top Hat	51.32%	39
Adobe Connect	15.79%	12
Total Respondents: 76		

Q8 To what extent do you agree or disagree with the following statements? Please use the provided scale, where 1 means strongly disagree, and 5 means strongly agree.

Answered: 75 Skipped: 5

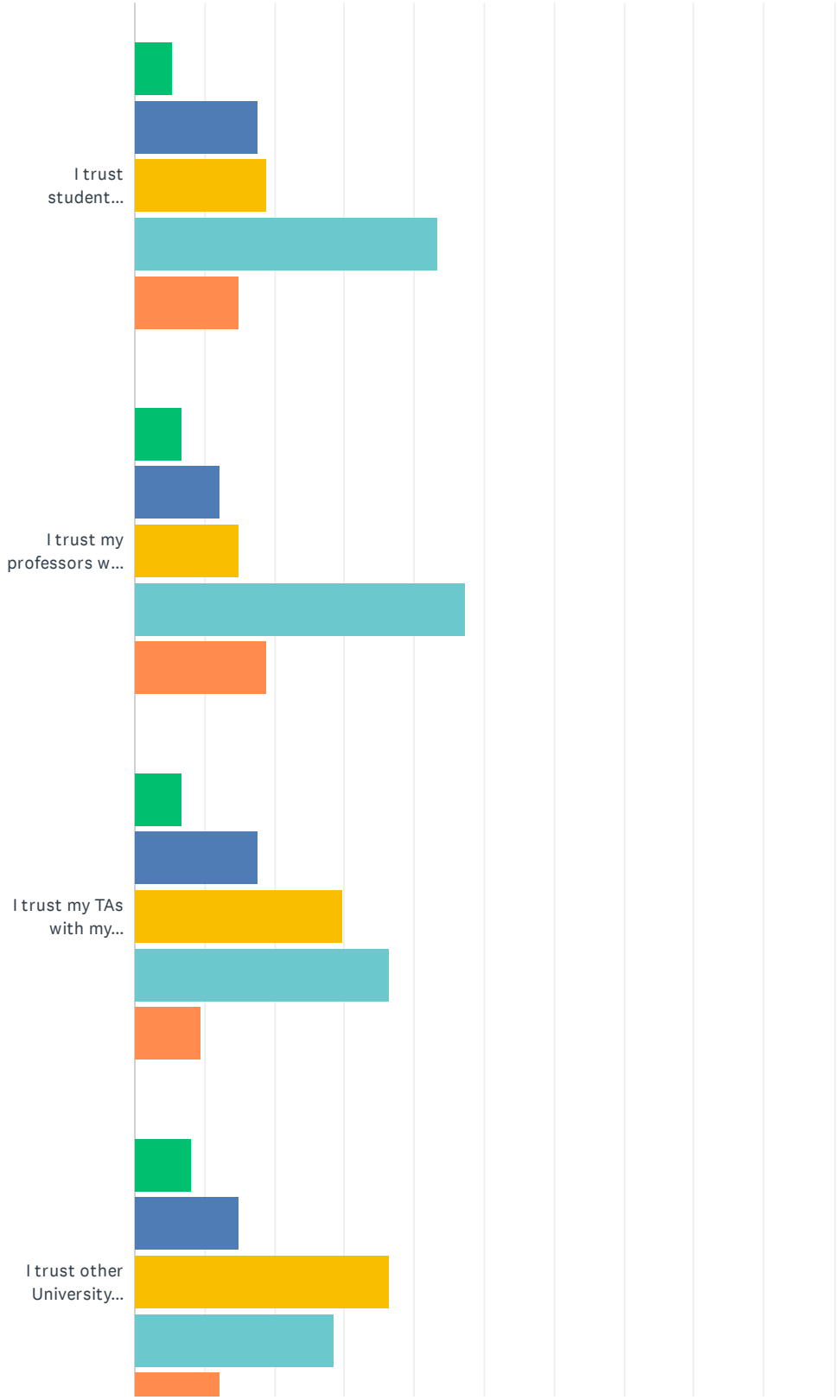


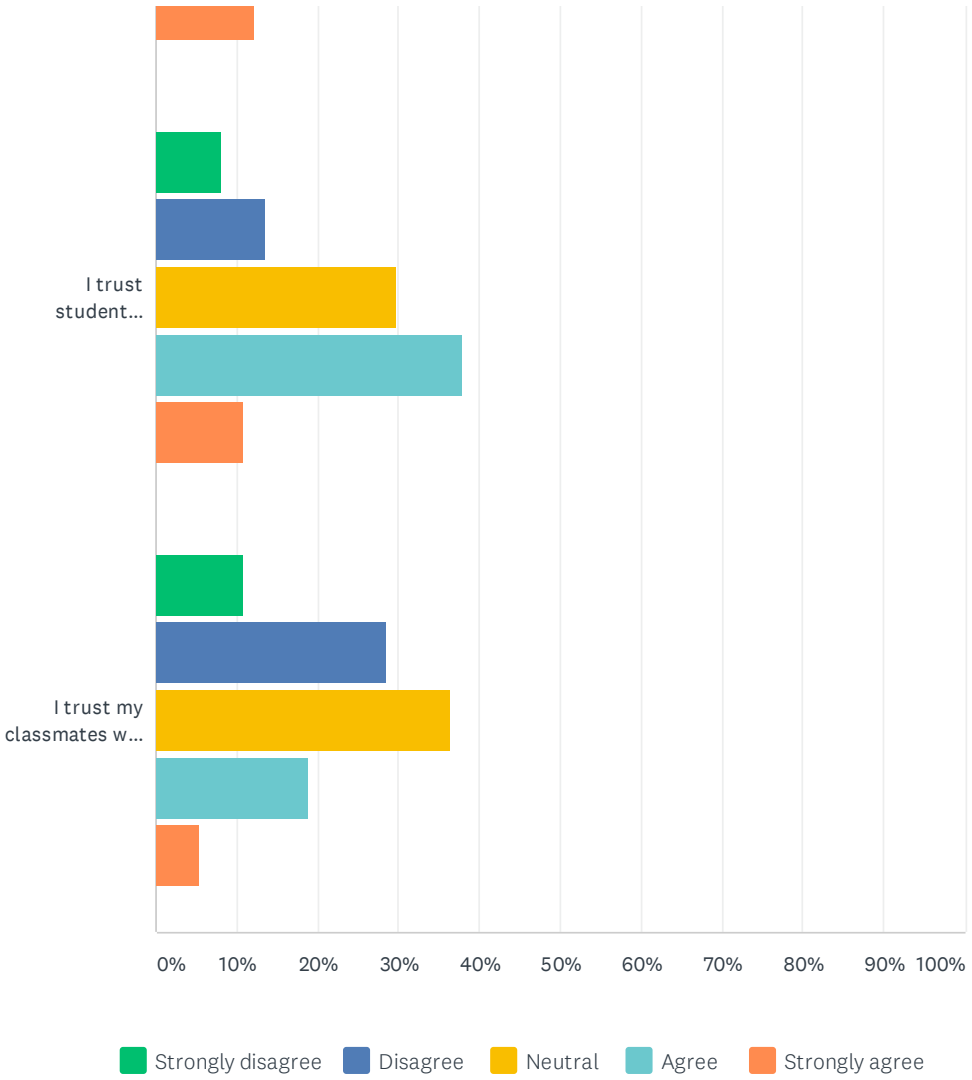


	STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE	TOTAL	WEIGHTED AVERAGE
The University's privacy policy is easy to find	9.33% 7	25.33% 19	50.67% 38	10.67% 8	4.00% 3	75	2.75
The University's privacy policy protects my privacy	2.67% 2	4.00% 3	69.33% 52	21.33% 16	2.67% 2	75	3.17
I remember consenting to the University's privacy policy	32.00% 24	33.33% 25	18.67% 14	9.33% 7	6.67% 5	75	2.25
My data is well protected by the University	4.00% 3	13.33% 10	61.33% 46	18.67% 14	2.67% 2	75	3.03
The University is careful about how it reuses my data	4.11% 3	13.70% 10	65.75% 48	13.70% 10	2.74% 2	73	2.97

Q9 To what extent do you agree or disagree with the following statements? Please use the provided scale, where 1 means strongly disagree, and 5 means strongly agree.

Answered: 74 Skipped: 6

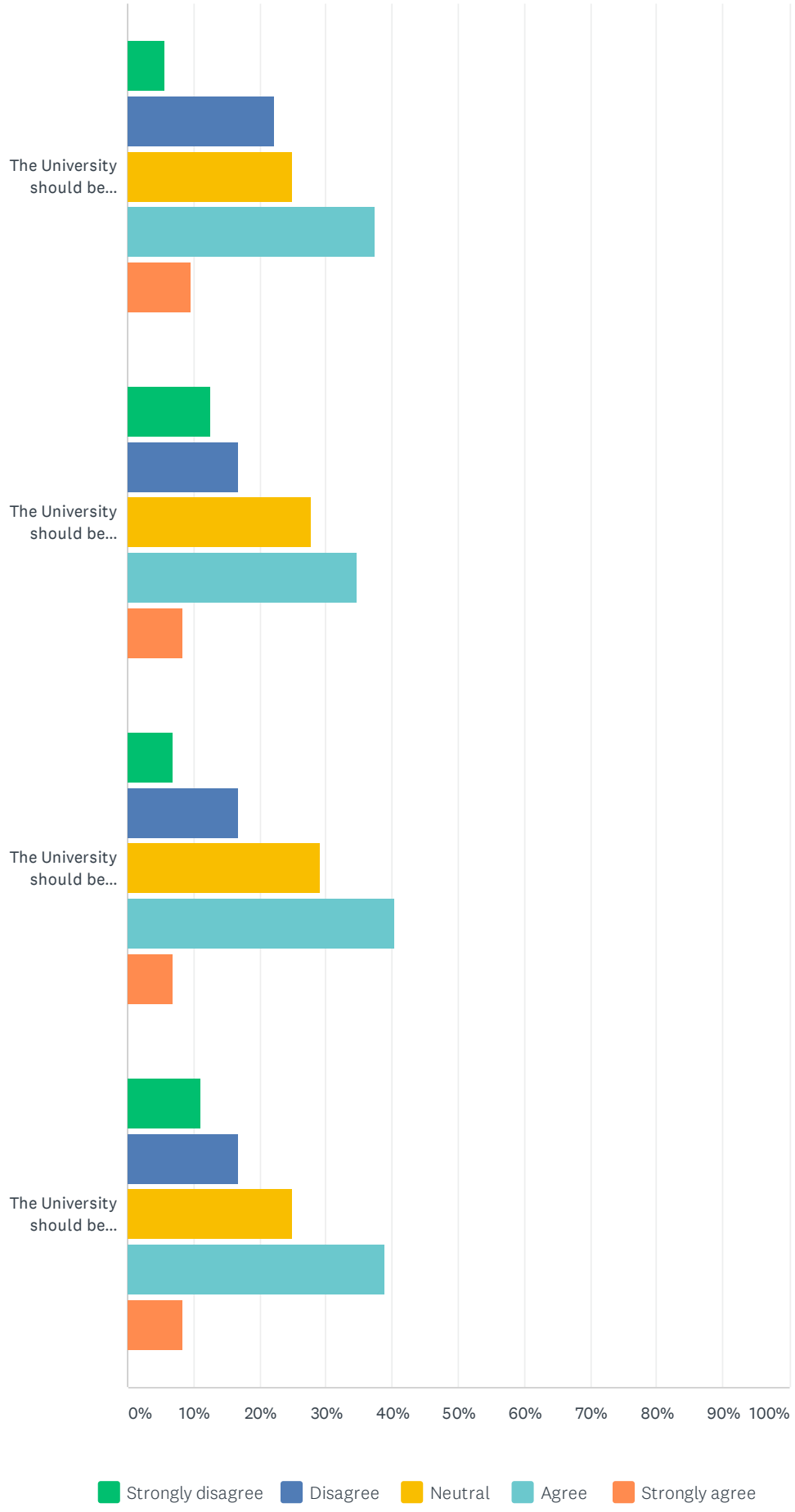




	STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE	TOTAL	WEIGHTED AVERAGE
I trust student advisors with my personal data	5.41% 4	17.57% 13	18.92% 14	43.24% 32	14.86% 11	74	3.45
I trust my professors with my personal data	6.76% 5	12.16% 9	14.86% 11	47.30% 35	18.92% 14	74	3.59
I trust my TAs with my personal data	6.76% 5	17.57% 13	29.73% 22	36.49% 27	9.46% 7	74	3.24
I trust other University staff with my personal data	8.11% 6	14.86% 11	36.49% 27	28.38% 21	12.16% 9	74	3.22
I trust student organizations that I belong to with my personal data	8.11% 6	13.51% 10	29.73% 22	37.84% 28	10.81% 8	74	3.30
I trust my classmates with my personal data	10.81% 8	28.38% 21	36.49% 27	18.92% 14	5.41% 4	74	2.80

Q10 To what extent do you agree or disagree with the following statements? Please use the provided scale, where 1 means strongly disagree, and 5 means strongly agree.

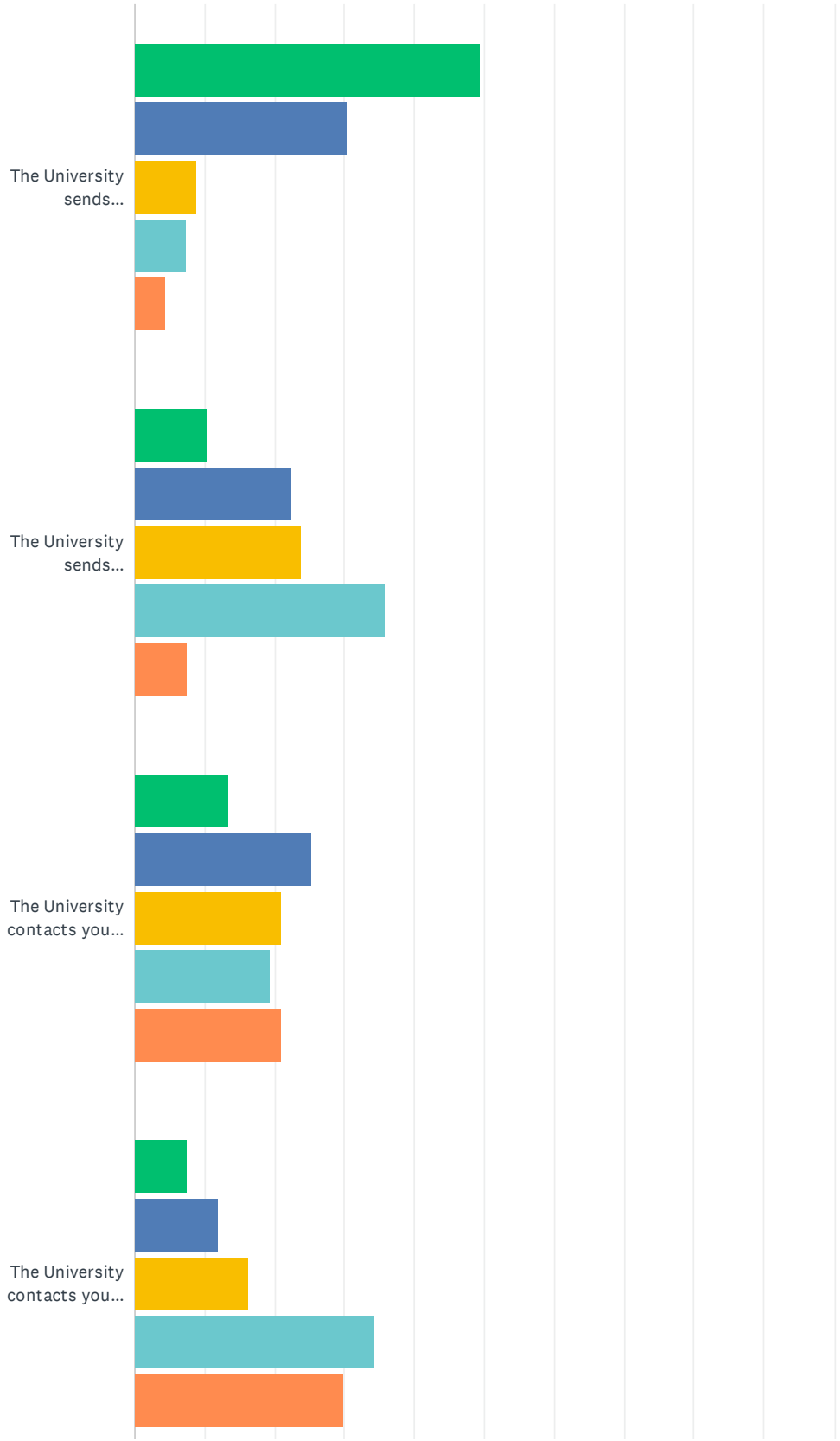
Answered: 72 Skipped: 8

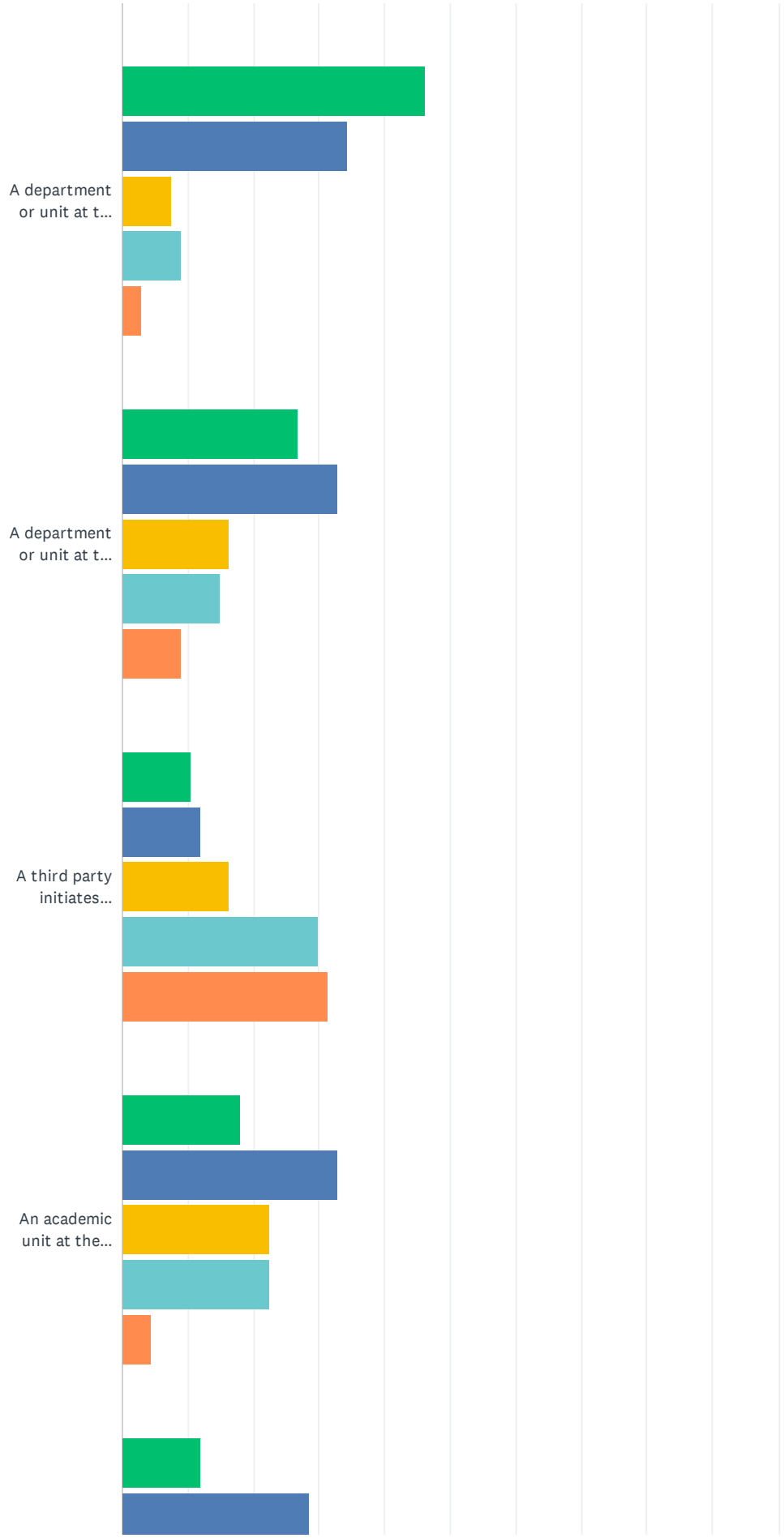

The University should be...
The University should be...
The University should be...
The University should be...
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Strongly disagree
Disagree
Neutral
Agree
Strongly agree

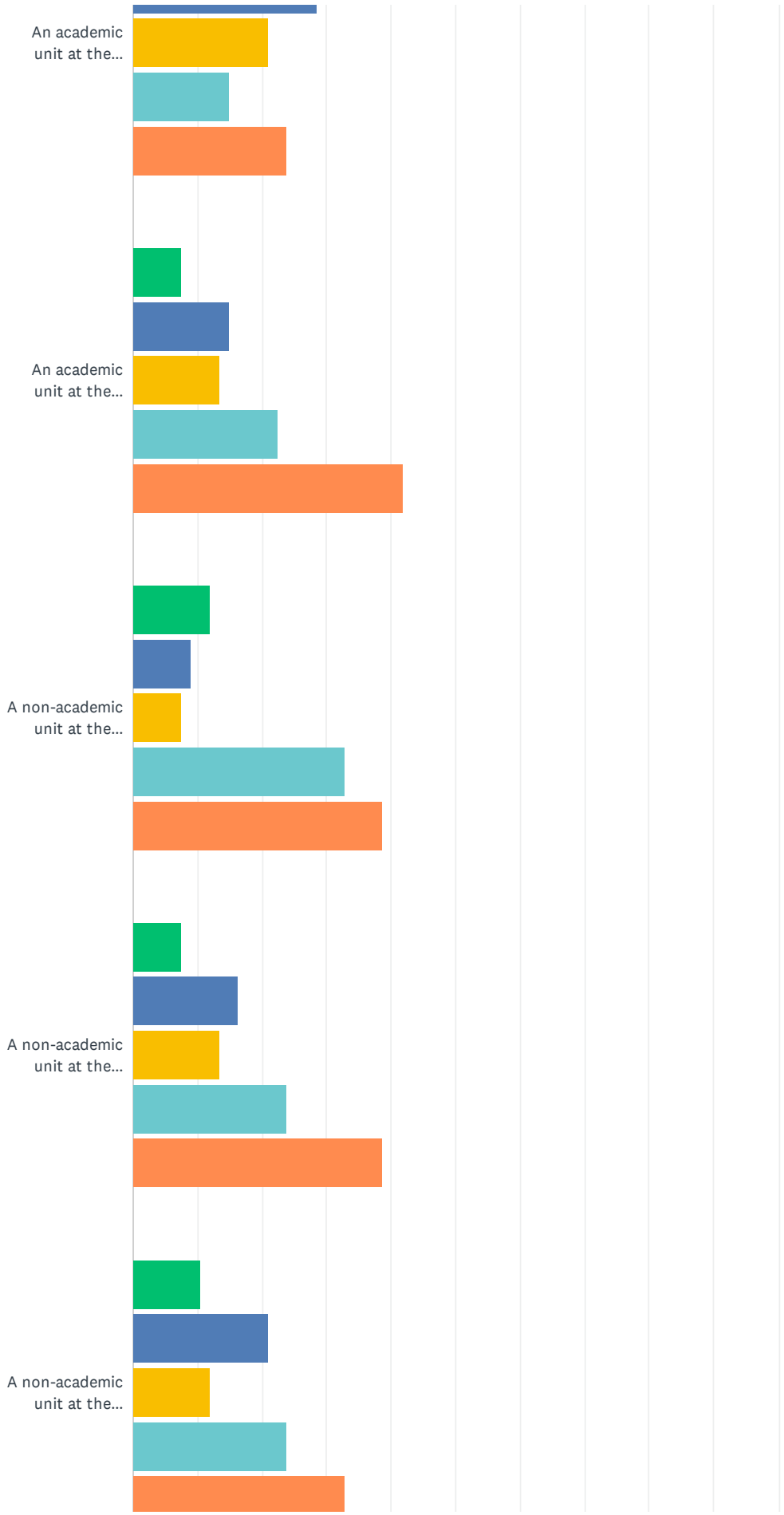
	STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE	TOTAL	WEIGHTED AVERAGE
The University should be allowed to use my personal data if it helps me with my degree	5.56% 4	22.22% 16	25.00% 18	37.50% 27	9.72% 7	72	3.24
The University should be allowed to use my personal data if it helps to improve my well-being	12.50% 9	16.67% 12	27.78% 20	34.72% 25	8.33% 6	72	3.10
The University should be allowed to use my personal data if it helps to improve the quality of my classes	6.94% 5	16.67% 12	29.17% 21	40.28% 29	6.94% 5	72	3.24
The University should be allowed to use my personal data if it helps to improve my experience at the U of C	11.11% 8	16.67% 12	25.00% 18	38.89% 28	8.33% 6	72	3.17

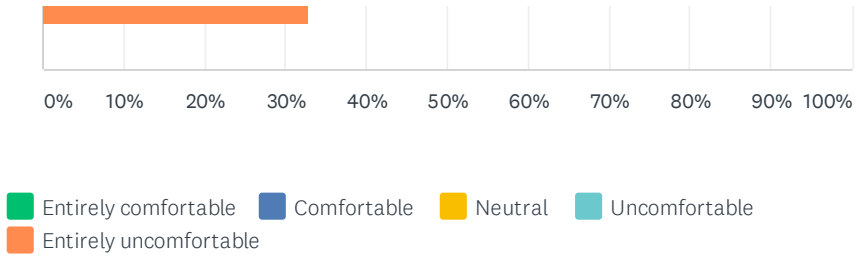
Q11 To what degree are you or would you be comfortable with the following:

Answered: 69 Skipped: 11




A department or unit at t...
A department or unit at t...
A third party initiates...
An academic unit at the...


An academic unit at the...
An academic unit at the...
A non-academic unit at the...
A non-academic unit at the...
A non-academic unit at the...

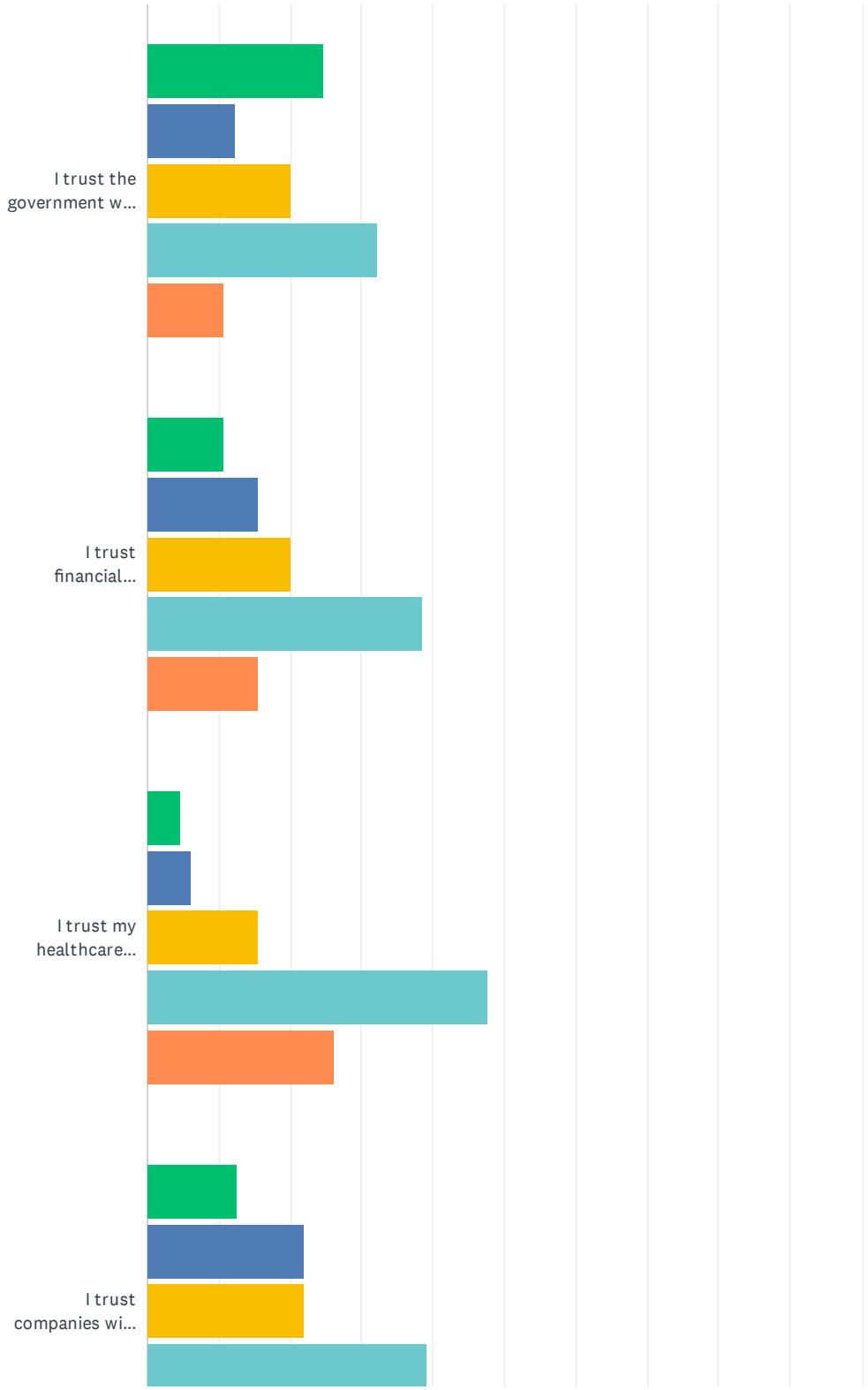

0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Entirely comfortable
Comfortable
Neutral
Uncomfortable
Entirely uncomfortable

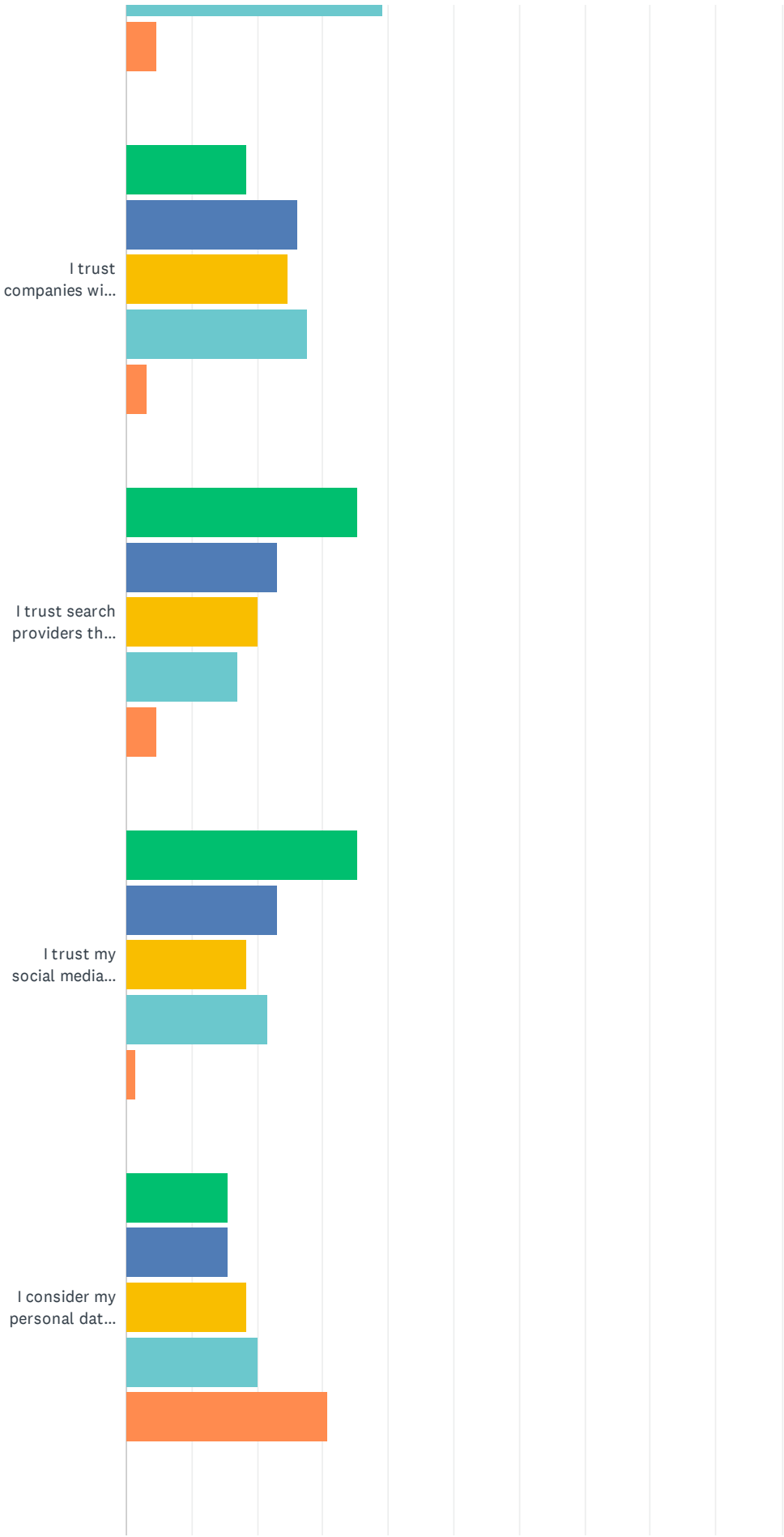
	ENTIRELY COMFORTABLE	COMFORTABLE	NEUTRAL	UNCOMFORTABLE	ENTIRELY UNCOMFORTABLE	TOTAL	WEIGHT AVERAG
The University sends correspondence to your university-provided email address	49.28% 34	30.43% 21	8.70% 6	7.25% 5	4.35% 3	69	1
The University sends correspondence to a private, non-university provided email address	10.45% 7	22.39% 15	23.88% 16	35.82% 24	7.46% 5	67	3
The University contacts you via social media to continue an interaction you initiated	13.43% 9	25.37% 17	20.90% 14	19.40% 13	20.90% 14	67	3
The University contacts you via social media to initiate an interaction	7.46% 5	11.94% 8	16.42% 11	34.33% 23	29.85% 20	67	3
A department or unit at the University you have interacted with corresponds with you via email	46.27% 31	34.33% 23	7.46% 5	8.96% 6	2.99% 2	67	1
A department or unit at the University you have not interacted with initiates correspondence with you via email	26.87% 18	32.84% 22	16.42% 11	14.93% 10	8.96% 6	67	2
A third party initiates correspondence with you with you via email, having received information about you from the university	10.45% 7	11.94% 8	16.42% 11	29.85% 20	31.34% 21	67	3
An academic unit at the university contacts you, with knowledge about your academic performance	17.91% 12	32.84% 22	22.39% 15	22.39% 15	4.48% 3	67	2
An academic unit at the	11.94% 8	28.36% 19	20.90% 14	14.93% 10	23.88% 16	67	3

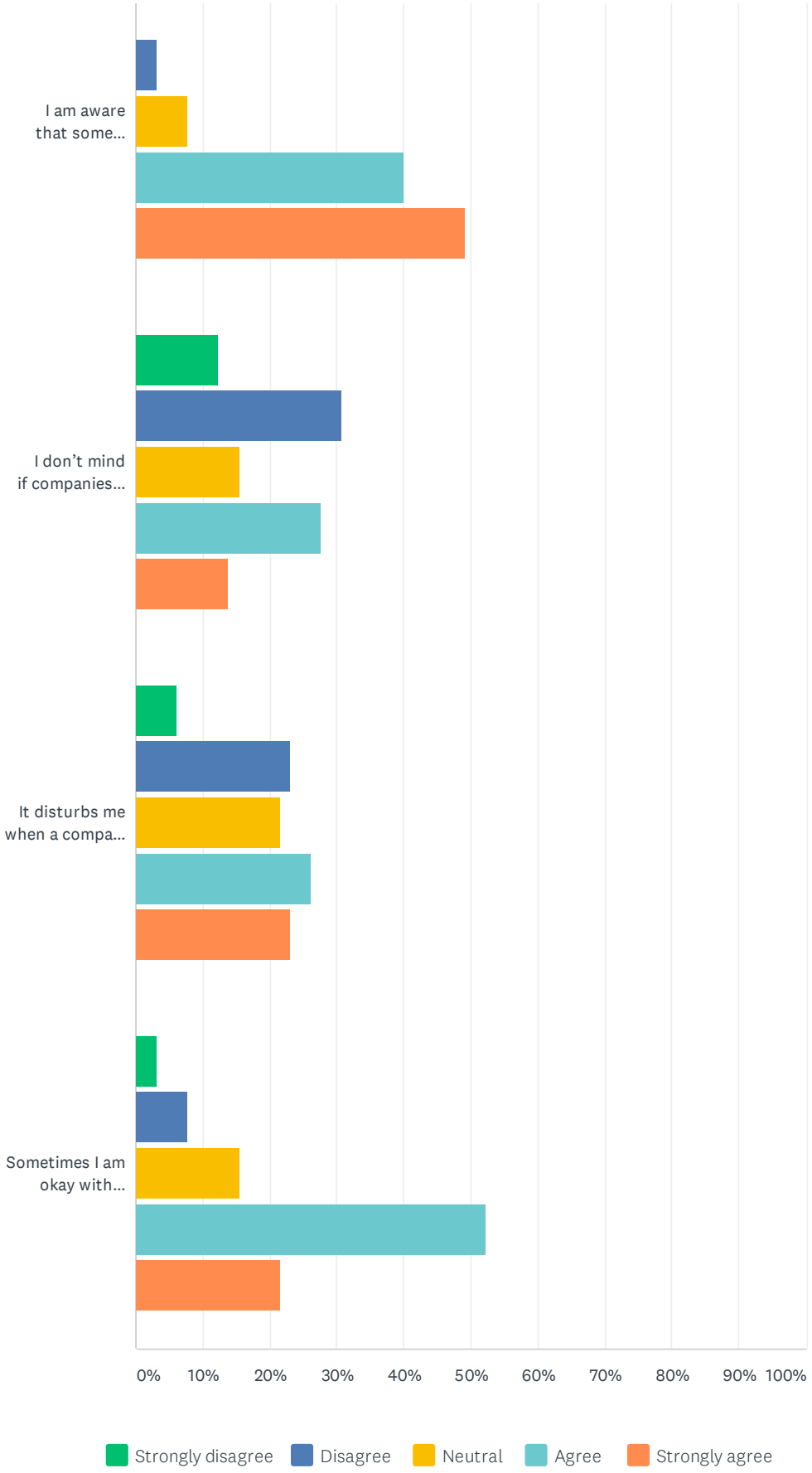
university contacts you, with knowledge about your non-academic activities on-campus							
An academic unit at the university contacts you, with knowledge about your activities off-campus	7.46% 5	14.93% 10	13.43% 9	22.39% 15	41.79% 28	67	3
A non-academic unit at the university contacts you, with knowledge about your academic performance	11.94% 8	8.96% 6	7.46% 5	32.84% 22	38.81% 26	67	3
A non-academic unit at the university contacts you, with knowledge about your activities off-campus	7.46% 5	16.42% 11	13.43% 9	23.88% 16	38.81% 26	67	3
A non-academic unit at the university contacts you, with knowledge about your non-academic activities on campus	10.45% 7	20.90% 14	11.94% 8	23.88% 16	32.84% 22	67	3

Q12 To what extent do you agree or disagree with the following statements? Please use the provided scale, where 1 means strongly disagree, and 7 means strongly agree. You may indicate where you are unsure or don't wish to answer.

Answered: 65 Skipped: 15



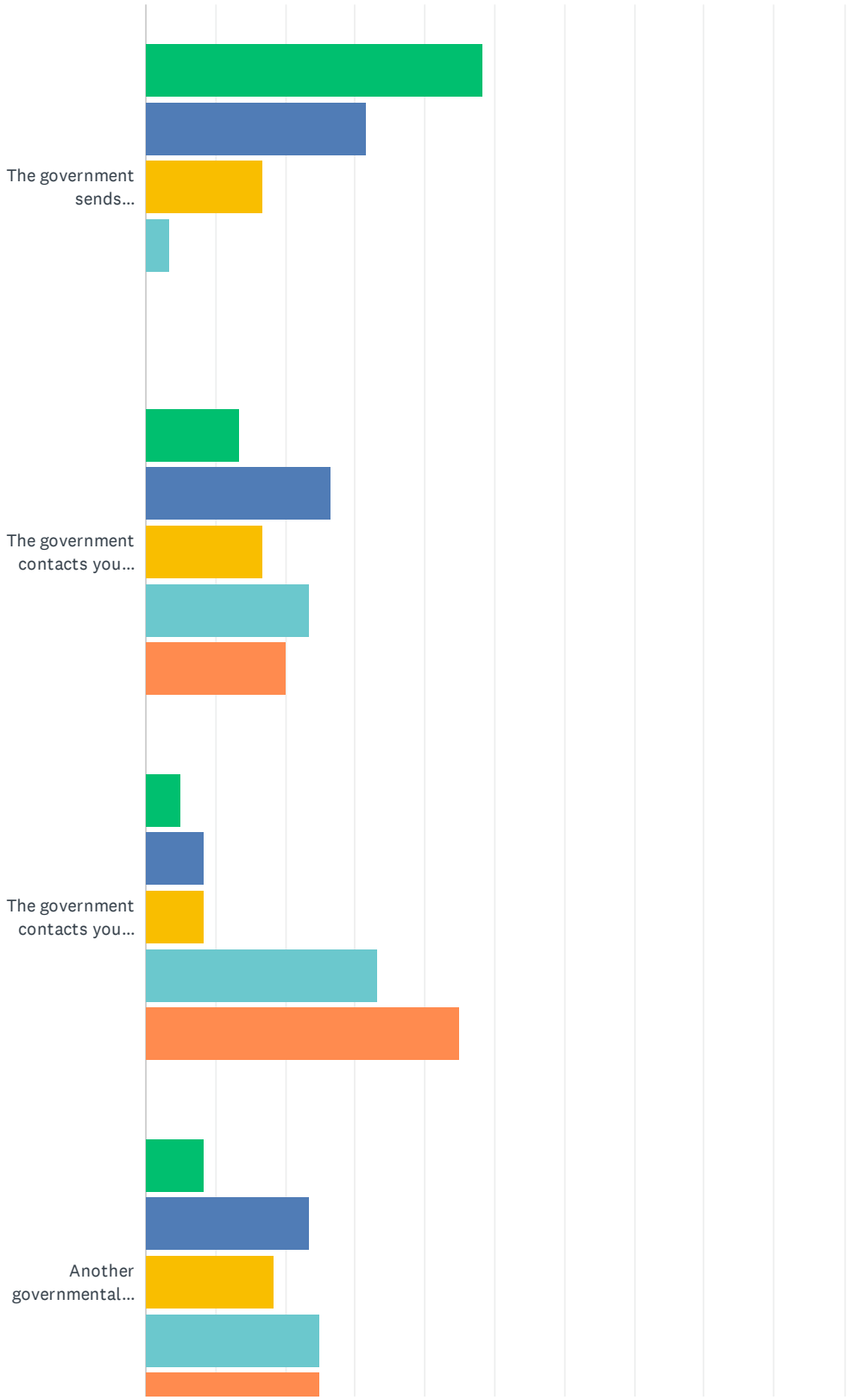

I trust companies wi...
I trust search providers th...
I trust my social media...
I consider my personal dat...

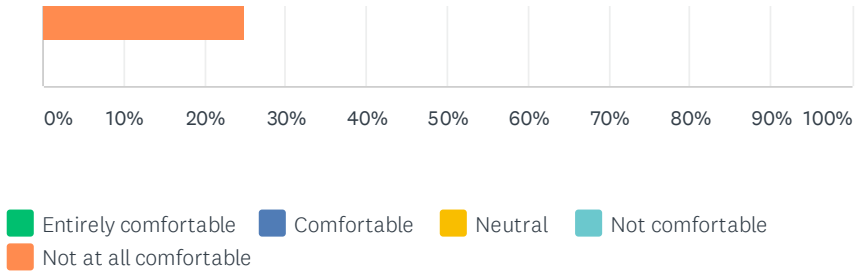

I am aware that some...
I don't mind if companies...
It disturbs me when a compa...
Sometimes I am okay with...
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Strongly disagree
Disagree
Neutral
Agree
Strongly agree

	STRONGLY DISAGREE	DISAGREE	NEUTRAL	AGREE	STRONGLY AGREE	TOTAL	WEIGHTED AVERAGE
I trust the government with my personal data	24.62% 16	12.31% 8	20.00% 13	32.31% 21	10.77% 7	65	2.92
I trust financial institutions with my personal data	10.77% 7	15.38% 10	20.00% 13	38.46% 25	15.38% 10	65	3.32
I trust my healthcare providers with my personal data	4.62% 3	6.15% 4	15.38% 10	47.69% 31	26.15% 17	65	3.85
I trust companies with whom I do business offline with my personal data	12.50% 8	21.88% 14	21.88% 14	39.06% 25	4.69% 3	64	3.02
I trust companies with whom I do business online with my personal data	18.46% 12	26.15% 17	24.62% 16	27.69% 18	3.08% 2	65	2.71
I trust search providers that I use to search the web with my personal data	35.38% 23	23.08% 15	20.00% 13	16.92% 11	4.62% 3	65	2.32
I trust my social media services with my personal data	35.38% 23	23.08% 15	18.46% 12	21.54% 14	1.54% 1	65	2.31
I consider my personal data private no matter which organization uses it	15.38% 10	15.38% 10	18.46% 12	20.00% 13	30.77% 20	65	3.35
I am aware that some companies use my private data for more than one thing	0.00% 0	3.08% 2	7.69% 5	40.00% 26	49.23% 32	65	4.35
I don’t mind if companies use my private data for more than one thing, as long as it’s clearly stated in their privacy policy.	12.31% 8	30.77% 20	15.38% 10	27.69% 18	13.85% 9	65	3.00
It disturbs me when a company uses my private data for more than one thing	6.15% 4	23.08% 15	21.54% 14	26.15% 17	23.08% 15	65	3.37
Sometimes I am okay with trusting one company with my private data, but not another.	3.08% 2	7.69% 5	15.38% 10	52.31% 34	21.54% 14	65	3.82

Q13 The following questions describe “the government” as a governmental body, such as the office of an elected representative, or services and agencies such as Alberta Registries or Parks Canada.

Answered: 60 Skipped: 20

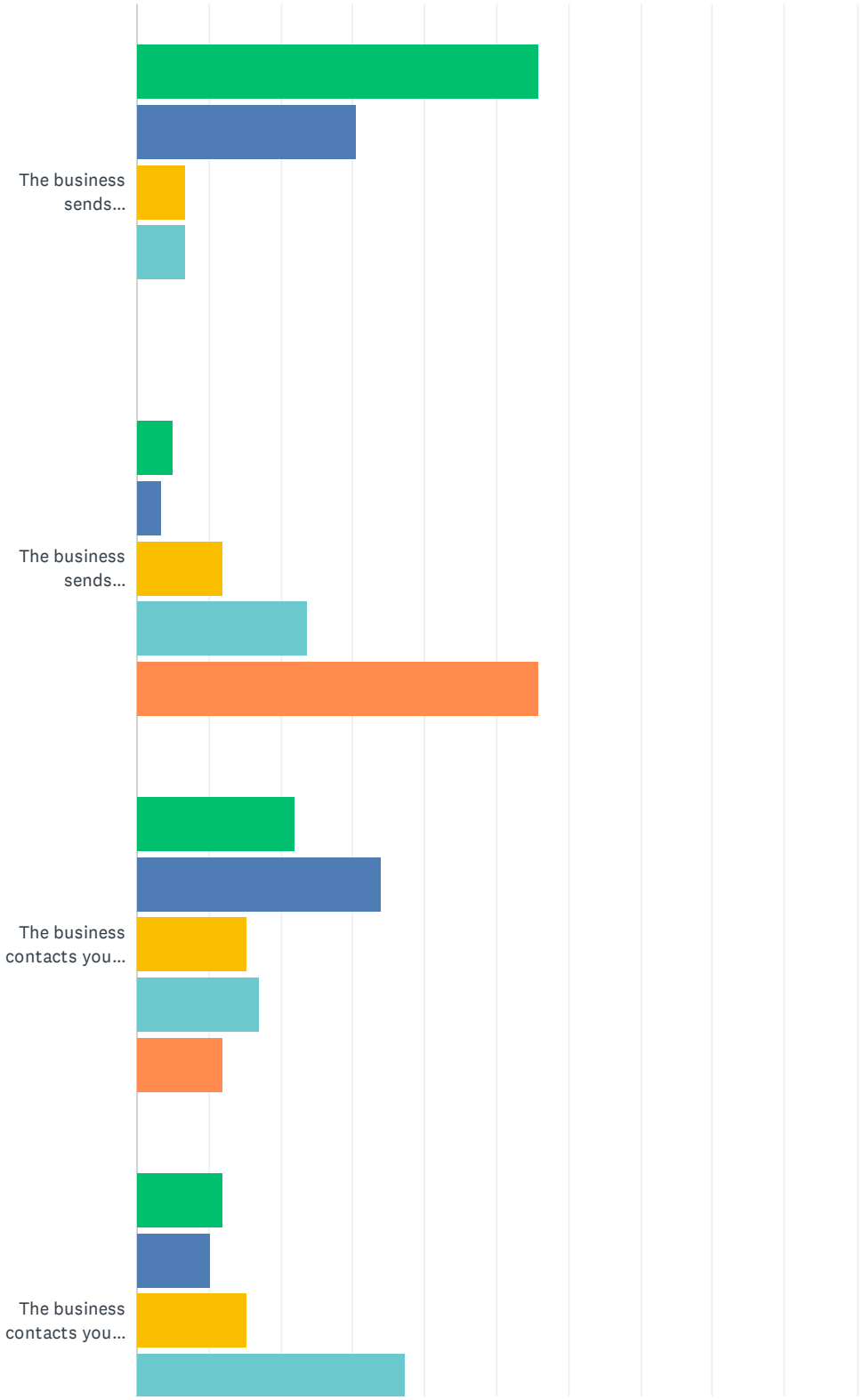


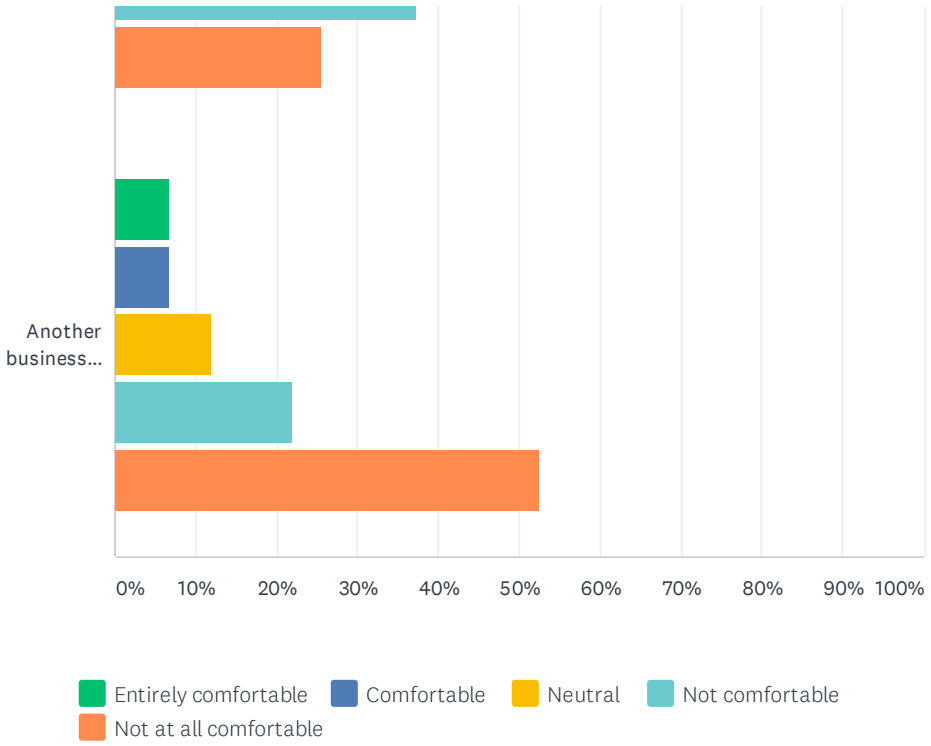


	ENTIRELY COMFORTABLE	COMFORTABLE	NEUTRAL	NOT COMFORTABLE	NOT AT ALL COMFORTABLE	TOTAL	WEIGHTED AVERAGE
The government sends correspondence to your private email address, which you have already provided them.	48.33% 29	31.67% 19	16.67% 10	3.33% 2	0.00% 0	60	2.27
The government contacts you via social media to continue an interaction you initiated	13.33% 8	26.67% 16	16.67% 10	23.33% 14	20.00% 12	60	4.17
The government contacts you via social media to initiate an interaction.	5.00% 3	8.33% 5	8.33% 5	33.33% 20	45.00% 27	60	5.45
Another governmental body initiates contact with you via email, having received your contact information and other data about you from a governmental body with which you have had contact.	8.33% 5	23.33% 14	18.33% 11	25.00% 15	25.00% 15	60	4.52

Q14 The following scenarios describe “a business” as a company with whom you have conducted some sort of business in the physical world. For example, a restaurant or a store where you have bought something.To what extent would you be comfortable with the following:

Answered: 59 Skipped: 21

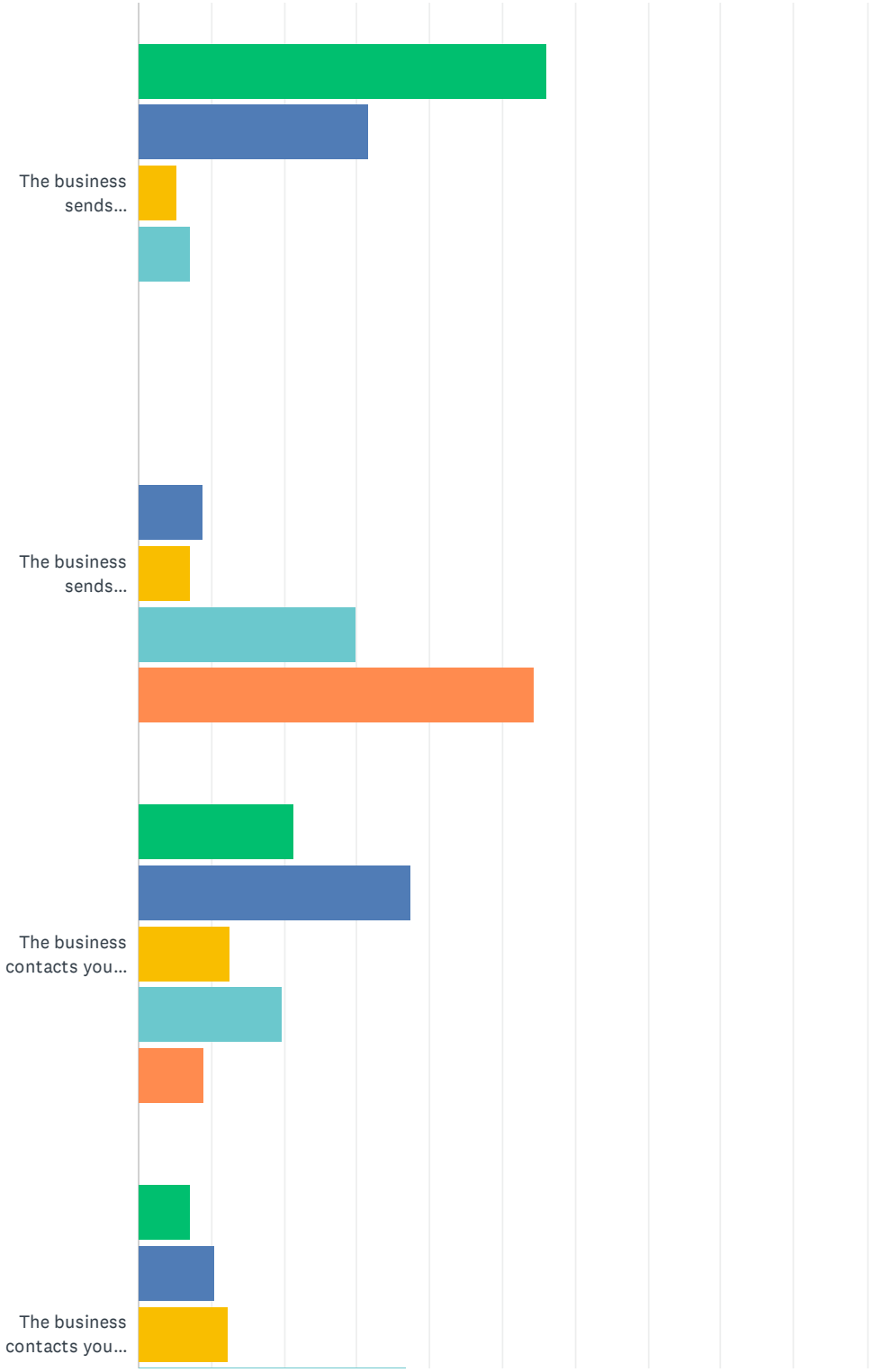


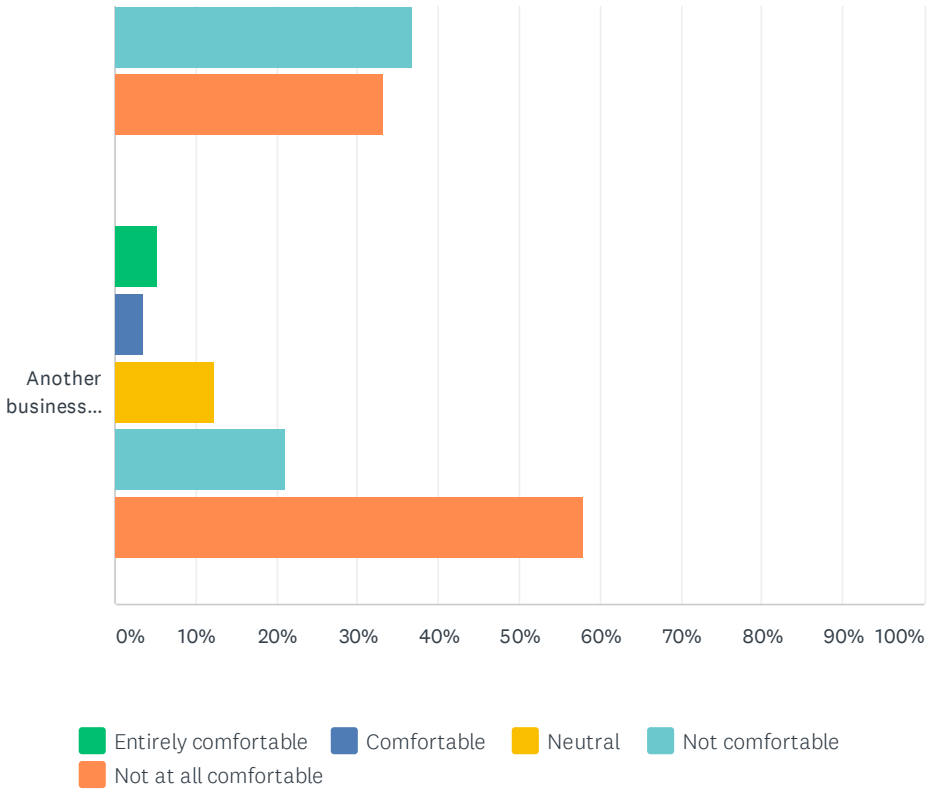

Another business...
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Entirely comfortable
Comfortable
Neutral
Not comfortable
Not at all comfortable

	ENTIRELY COMFORTABLE	COMFORTABLE	NEUTRAL	NOT COMFORTABLE	NOT AT ALL COMFORTABLE	TOTAL	WEIGHTED AVERAGE
The business sends correspondence to your private email address, which you have already provided them.	55.93% 33	30.51% 18	6.78% 4	6.78% 4	0.00% 0	59	1.64
The business sends correspondence to your private email address, and it is unclear how they have acquired your email.	5.08% 3	3.39% 2	11.86% 7	23.73% 14	55.93% 33	59	4.22
The business contacts you via social media to continue an interaction you initiated	22.03% 13	33.90% 20	15.25% 9	16.95% 10	11.86% 7	59	2.63
The business contacts you via social media to initiate an interaction.	11.86% 7	10.17% 6	15.25% 9	37.29% 22	25.42% 15	59	3.54
Another business initiates contact with you via email, having received your contact information and other data about you from this business.	6.78% 4	6.78% 4	11.86% 7	22.03% 13	52.54% 31	59	4.07

Q15 The following scenarios describe an “online business” as a company with whom you have conducted some sort of business, with whom you have only had contact with online. For example, an online retailer such as Amazon or a service (such as Spotify)To what extent would you be comfortable with the following:

Answered: 57 Skipped: 23

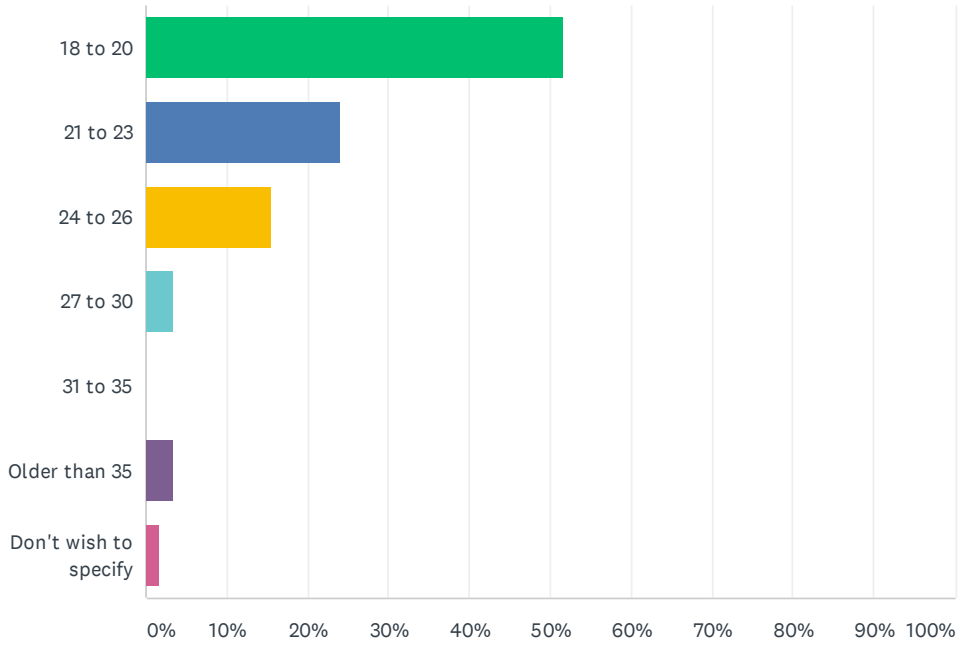



Another business...
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Entirely comfortable
Comfortable
Neutral
Not comfortable
Not at all comfortable

	ENTIRELY COMFORTABLE	COMFORTABLE	NEUTRAL	NOT COMFORTABLE	NOT AT ALL COMFORTABLE	TOTAL	WEIGHTED AVERAGE
The business sends correspondence to your private email address, which you have already provided them.	56.14% 32	31.58% 18	5.26% 3	7.02% 4	0.00% 0	57	1.63
The business sends correspondence to your private email address, and it is unclear how they have acquired your email.	0.00% 0	8.77% 5	7.02% 4	29.82% 17	54.39% 31	57	4.30
The business contacts you via social media to continue an interaction you initiated	21.43% 12	37.50% 21	12.50% 7	19.64% 11	8.93% 5	56	2.57
The business contacts you via social media to initiate an interaction.	7.02% 4	10.53% 6	12.28% 7	36.84% 21	33.33% 19	57	3.79
Another business initiates contact with you via email, having received your contact information and other data about you from this business.	5.26% 3	3.51% 2	12.28% 7	21.05% 12	57.89% 33	57	4.23

Q16 I am:

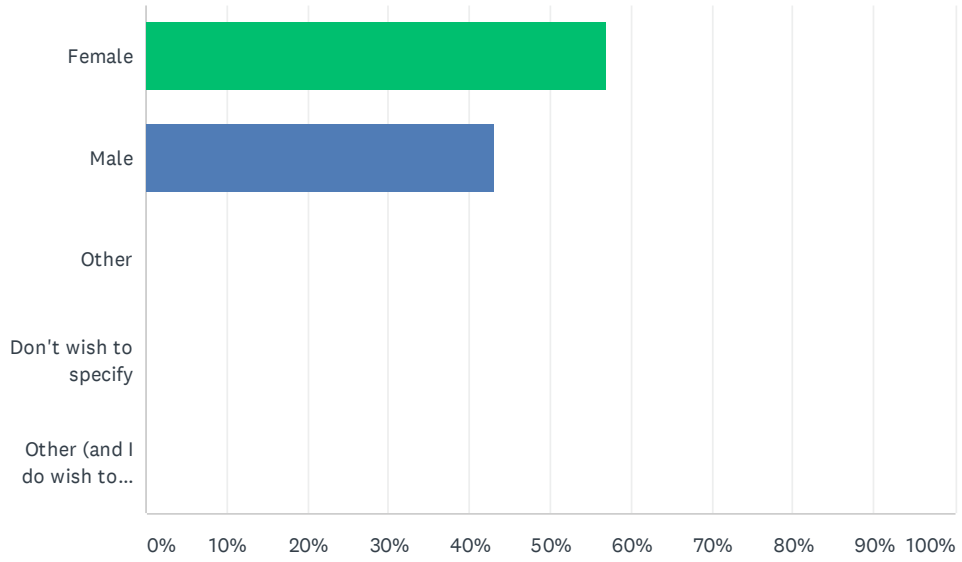
Answered: 58 Skipped: 22



ANSWER CHOICES	RESPONSES	
18 to 20	51.72%	30
21 to 23	24.14%	14
24 to 26	15.52%	9
27 to 30	3.45%	2
31 to 35	0.00%	0
Older than 35	3.45%	2
Don't wish to specify	1.72%	1
TOTAL		58

Q17 I am

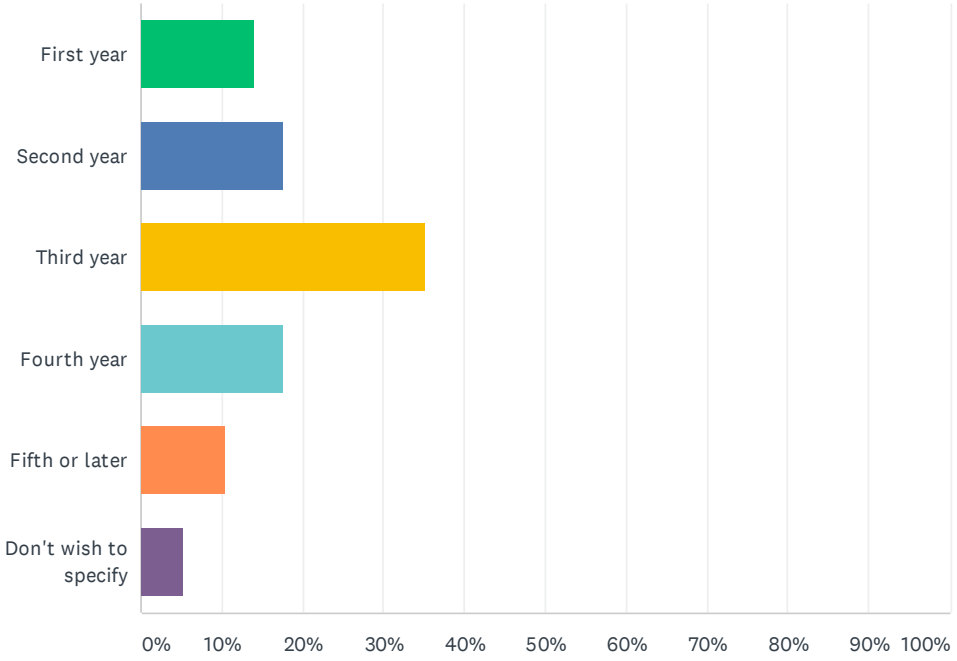
Answered: 58 Skipped: 22



ANSWER CHOICES	RESPONSES	
Female	56.90%	33
Male	43.10%	25
Other	0.00%	0
Don't wish to specify	0.00%	0
Other (and I do wish to specify)	0.00%	0
TOTAL		58

Q18 My current year of program is:

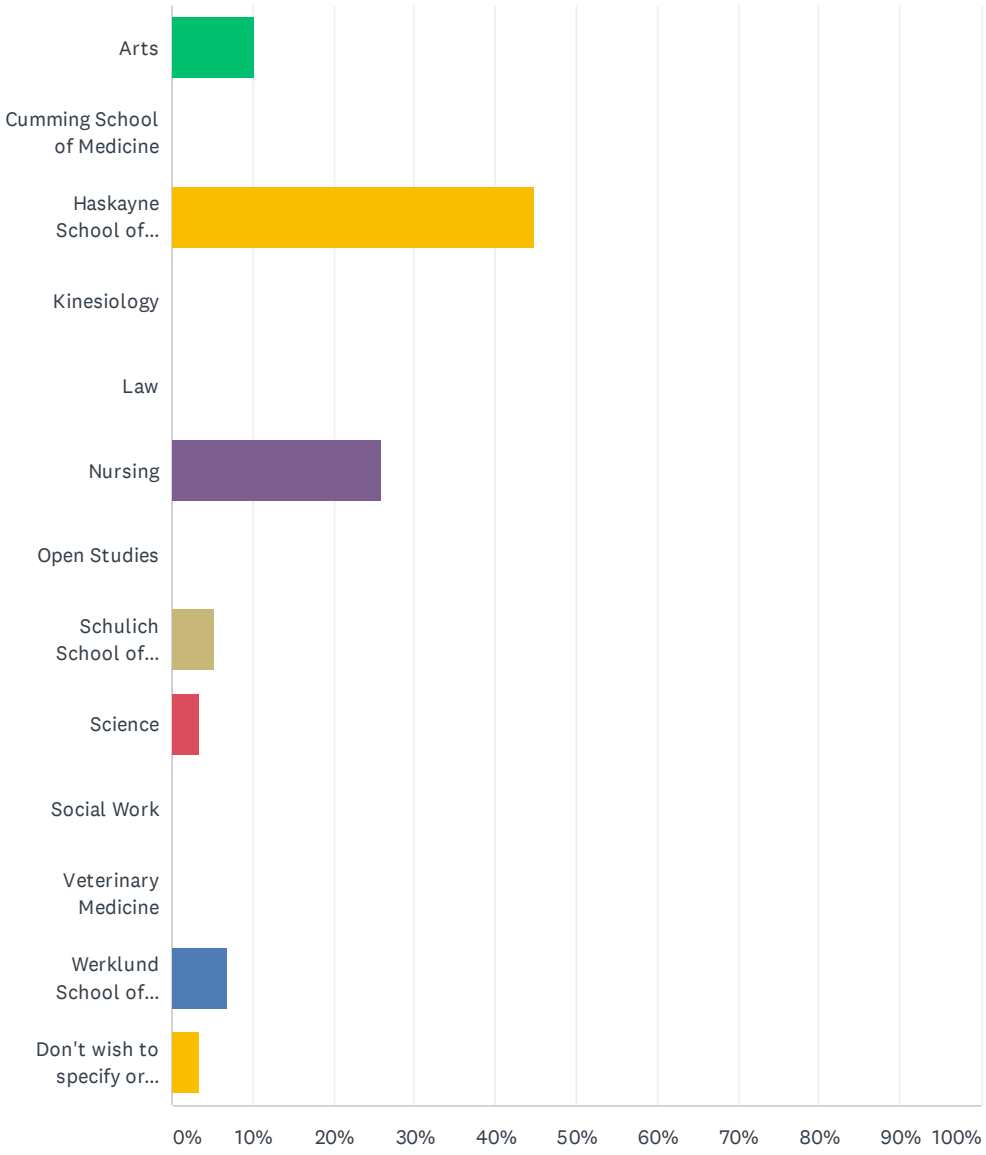
Answered: 57 Skipped: 23



ANSWER CHOICES	RESPONSES	
First year	14.04%	8
Second year	17.54%	10
Third year	35.09%	20
Fourth year	17.54%	10
Fifth or later	10.53%	6
Don't wish to specify	5.26%	3
TOTAL		57

Q19 My program belongs to the following faculty:

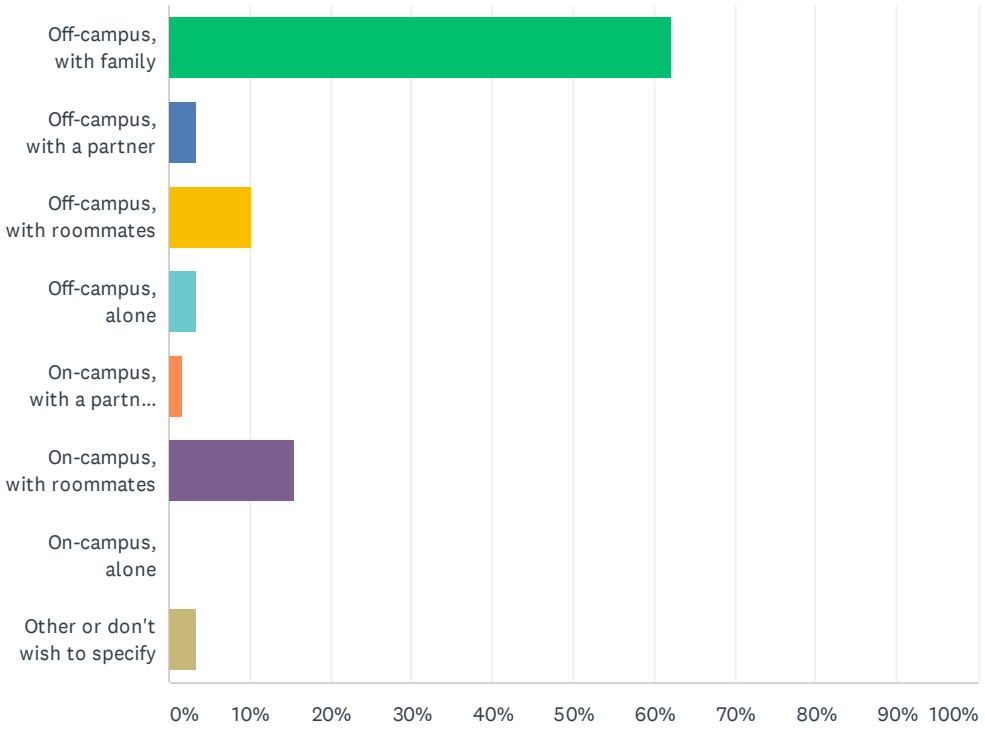
Answered: 58 Skipped: 22



ANSWER CHOICES	RESPONSES	
Arts	10.34%	6
Cumming School of Medicine	0.00%	0
Haskayne School of Business	44.83%	26
Kinesiology	0.00%	0
Law	0.00%	0
Nursing	25.86%	15
Open Studies	0.00%	0
Schulich School of Engineering	5.17%	3
Science	3.45%	2
Social Work	0.00%	0
Veterinary Medicine	0.00%	0
Werklund School of Education	6.90%	4
Don't wish to specify or don't know	3.45%	2
TOTAL		58

Q20 The following best describes my current living situation:

Answered: 58 Skipped: 22



ANSWER CHOICES	RESPONSES	
Off-campus, with family	62.07%	36
Off-campus, with a partner	3.45%	2
Off-campus, with roommates	10.34%	6
Off-campus, alone	3.45%	2
On-campus, with a partner or family	1.72%	1
On-campus, with roommates	15.52%	9
On-campus, alone	0.00%	0
Other or don't wish to specify	3.45%	2
TOTAL		58

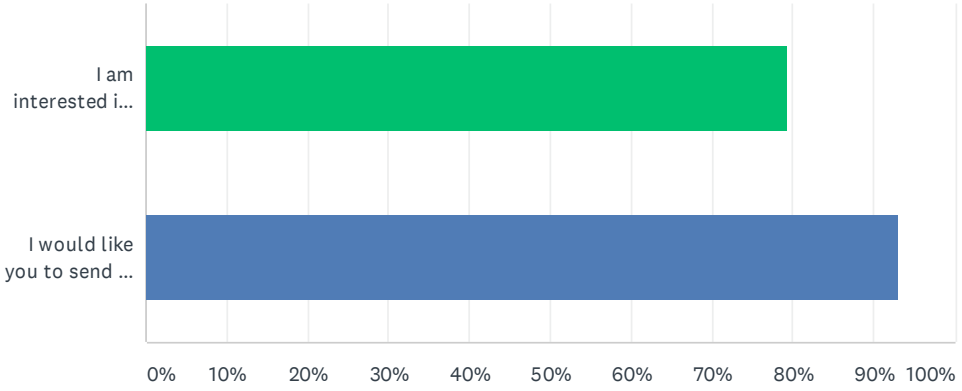
Q21 Please enter your email address here:

Answered: 28 Skipped: 52

ANSWER CHOICES	RESPONSES	
Name	0.00%	0
Company	0.00%	0
Address	0.00%	0
Address 2	0.00%	0
City/Town	0.00%	0
State/Province	0.00%	0
ZIP/Postal Code	0.00%	0
Country	0.00%	0
Email Address (optional)	100.00%	28
Phone Number	0.00%	0

Q22 Please select the options which apply.

Answered: 29 Skipped: 51



ANSWER CHOICES	RESPONSES	
I am interested in participating in a future, follow-up study about this topic	79.31%	23
I would like you to send me the results of your analysis, once complete.	93.10%	27
Total Respondents: 29		