RESEARCH ARTICLE

Paths to social licence for tracking-data analytics in university research and services

Joshua P. White[1]*, **Simon Dennis**[1], **Martin Tomko**[2], **Jessica Bell**[3], **Stephan Winter**[2]

[1] Complex Human Data Hub, Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Victoria, Australia, [2] Melbourne School of Engineering, The University of Melbourne, Melbourne, Victoria, Australia, [3] Melbourne Law School, The University of Melbourne, Melbourne, Victoria, Australia

* josh.white@unimelb.edu.au

OPEN ACCESS

Citation: White JP, Dennis S, Tomko M, Bell J, Winter S (2021) Paths to social licence for tracking-data analytics in university research and services. PLoS ONE 16(5): e0251964. https://doi.org/10.1371/journal.pone.0251964

Editor: Barbara Guidi, University of Pisa, ITALY

Received: October 14, 2020

Accepted: May 6, 2021

Published: May 21, 2021

Data Availability Statement: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supporting Information. The complete data and code for reproducing the results in this article are available on the Center for Open Science website https://osf.io/mx948/, doi:10.17605/OSF.IO/MX948.

Funding: This research is part of the LUMAS project, funded by the Networked Society Institute at the University of Melbourne (https://research.unimelb.edu.au/research-at-melbourne/multidisciplinary-research/melbourne-interdisciplinary-research-institutes/former-

Abstract

While tracking-data analytics can be a goldmine for institutions and companies, the inherent privacy concerns also form a legal, ethical and social minefield. We present a study that seeks to understand the extent and circumstances under which tracking-data analytics is undertaken with social licence—that is, with broad community acceptance beyond formal compliance with legal requirements. Taking a University campus environment as a case, we enquire about the social licence for Wi-Fi-based tracking-data analytics. Staff and student participants answered a questionnaire presenting hypothetical scenarios involving Wi-Fi tracking for university research and services. Our results present a Bayesian logistic mixed-effects regression of acceptability judgements as a function of participant ratings on 11 privacy dimensions. Results show widespread acceptance of tracking-data analytics on campus and suggest that trust, individual benefit, data sensitivity, risk of harm and institutional respect for privacy are the most predictive factors determining this acceptance judgement.

Introduction

The increased simplicity of acquiring tracking data (spatio-temporal data attached to an identifiable object or person), and their immense utility for logistics, traffic and space management, human resource management, and advertising has turned tracking-data analytics into a revenue goldmine. Yet, the inherent privacy concerns of tracking data turn them equally into a minefield.

One source for tracking data are Wi-Fi networks [1, 2]. Campus-based universities, shopping malls, and hospitals are examples of complex, built environments with varied stakeholders and complex logistical issues that, on one hand, operate such Wi-Fi networks, and on the other need data-driven support for efficient operations. The realisation that tracking data contain information about staff attendance, can support health and safety reporting and compliance monitoring, and can inform space-use optimisation leads to an increased interest on behalf of institutions in active tracking technology deployments [2, 3].

While the operators of the infrastructure will use this data for the optimization of their services, other parties have an interest in this data as well [4]. Other interested parties, such as those whose data is being collected, may consider the use of tracking-data analytics to be

melbourne-interdisciplinary-research-institutes/networked-society-institute). Funding was awarded to S.W and S.D. The funders did not play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

invasive, controlling, and only to the benefit of the institution collecting the data, while other possible uses might be accepted by the users as useful, legitimate, or desirable. For example, using tracking-data to monitor staff attendance may be of benefit to an institution, but could largely be to the detriment of workers for whom it could provoke mistrust and threaten their sense of agency and privacy. Clearly, as with any technological advancement, the use of tracking data analytics involves the balancing of various social, ethical, and legal issues. Yet while infrastructure operators often focus only on satisfying the legal requirements that underlie operations, recent scholarship has emphasised the importance of going beyond this minimal threshold to technology deployment, and secure a broader basis for community acceptance.

This notion—termed '*social licence to operate*'—while developed primarily in the corporate social responsibility literature to describe the support and acceptance within the community of private companies' operations, has recently gained traction in the context of large-scale personal data collection and analysis. There are a number of examples of recent controversies in the healthcare sector that exemplify the degree to which a failure to secure social licence for research initiatives can undermine project success; e.g., My Health Record in Australia and Care.data in the UK [5]. Such failures to secure social licence, however, are not limited to the healthcare sector. Indeed, the release of information about the use of Wi-Fi and other big data in learning analytics in Australian universities has been similarly controversial and received backlash from students, regulators and the media [6–9]. Such controversies have also occurred in US universities [10, 11], and may cause negative media attention and attitudes towards institutions, as well as negatively affect the quality of the data collected through reduced compliance or cooperation. These examples illustrate the importance of building social licence and trust amongst the community in a university and education setting if the potential benefits of use of Wi-Fi tracking data are to be realised in the future.

Extensive literature covers the ethically normative requirements for using big data in a variety of contexts [4, 12], including in universities [13]. However, most work in relation to social licence focuses on the legal (i.e., what can entities legally do) and normative frameworks (i.e., what entities should do). However, little work has considered the social licence for big data usage from an empirical or descriptive perspective. While there has recently been an increase in this empirical work, it has generally been limited to the social licence for big data usage in governmental services or the mining industry [12, 14–16]. Here, we complement this gap by focusing on the quantitative analysis of the social licence for data use in research, in the context of data collected and analysed at Universities.

On university campuses, tracking data are often applied to the optimization of campus management, operations, and university services [3]. While, these uses are often legally covered by the Wi-Fi user agreements, a failure to ensure that participants broadly accept their operation can cause negative media attention and attitudes towards the university [6, 9, 10]. Further, these data are also of interest for academic research from a more fundamental perspective of advancing knowledge about the applications and limits of tracking data in a variety of domains. This includes research into privacy protection [17, 18] and obfuscation [19, 20], data cleaning and interpolation [1, 21], pattern detection and prediction, contact tracing [22], and the integration of tracking-data with social media sentiments [23] or with student or staff performance [24]. The use of the Wi-Fi tracking data for research purposes may not be covered in Wi-Fi user agreements, and may require additional informed consent. Wi-Fi users—the whole campus community—may feel uncomfortable with the topics, methodologies, or potential applications of some research and might thus resist having their data contribute towards them. On the other hand, many on campus might benefit from more permissive use of such data through the provision of better university services and management, greater research opportunities for graduate students, and more diffuse public benefits at large.

We thus argue that Universities are a particularly interesting environment to study from the perspective of social licence. The use of Wi-Fi location data in university research and services presents both great opportunities, but also risks such as loss of privacy, community backlash, data breaches etc. This data can be used to greatly benefit the community, by providing the ability to quickly conduct cutting-edge research with real and varied populations, the benefits of which to the community may be immense. In addition to legal requirements, institutional research ethics frameworks also govern Universities, providing extra oversight for data use. There is also a high level of collegiality amongst staff and students. For a productive and flourishing University community, it is crucial that this confidence is not undermined. Trust and confidence are particularly important because in many University institutions, individuals are a 'captive' population; individuals need to use the Wi-Fi network to effectively conduct their work and studies. As such, an individual's 'freedom' to opt-out may be fundamentally compromised. For this reason, it is vital to ensure protections are in place for stakeholder privacy, security and risk of harm and it is equally crucial to ensure social licence for trustworthy Wi-Fi location data use before implementation on campus.

We seek to understand to what extent and under what circumstances tracking-data analytics at universities is undertaken under a *social licence*—i.e., with broad community acceptance beyond formal compliance with legal requirements [25]. Understanding the acceptance by Wi-Fi users may help (a) academics to understand sensitivities and limits in advance, (b) ethics boards with formulating policies and guidelines, (c) university governance with a realistic evaluation of user sentiments, and (d) with a broader public debate on acceptable and unacceptable use of tracking data.

To work towards a more nuanced understanding of the parameters of a social licence for the research use of Wi-Fi data on university campuses, in this paper we specifically address two research questions: first, which scenarios for Wi-Fi location tracking do university community stakeholders find acceptable?; and second, what are the factors within those scenarios that can predict this acceptability?

To answer these questions, participants in our study were exposed to a questionnaire presenting hypothetical scenarios depicting the use of Wi-Fi tracking data in research and support services at the University of Melbourne, and rated these scenarios on 11 privacy dimensions before indicating whether each scenario was an "acceptable" use of university tracking data. Using this survey data, we generalise predictive factors of community acceptability for the collection and use of university Wi-Fi location data, and develop a forward-looking predictive model to allow administrators, ethics boards, and other decision-makers to foreshadow the level of university community acceptance for new research and service projects.

Materials and methods

Participants

Participants were recruited via online and physical advertisements at the University of Melbourne and were compensated for their participation by being entered into a draw to win $500. To be included in the study, participants were required to be from the University of Melbourne community defined to include students, staff, and any others who regularly attend the University of Melbourne campus. There were no exclusion criteria.

Materials

Materials were 11 hypothetical scenarios, each of which described a project involving the use of Wi-Fi location data on the University of Melbourne campus. The scenarios were co-developed by a group of key University of Melbourne stakeholders including students, academics, and professional staff to cover a range of plausible university research and service projects that may require the use of tracking-data analytics on campus, and to cover a range of projects with different purposes, benefits, and funding arrangements. All 11 scenarios are reproduced in full at S1 Text, but a summary of each is shown in Table 1. The scenarios all depict uses of tracking-data that are likely to provide benefits to the university community (e.g., students, staff, researchers, people attending campus) or the wider community. This was done to place tracking-data use in a real-world context that requires participants to undergo a 'privacy calculus' [26] about whether the benefits of the project outweigh its privacy costs, and give a more ecologically valid understanding of participants attitudes to privacy. By using these real world scenarios with appreciable benefits, we hope to avoid the 'privacy paradox' in which people

Table 1. Summary of scenarios presented to participants.

Scenario Name	Scenario Summary
Work Records	University services develop a web-based platform to enable workers on campus to log their working hours and track their working locations through Wi-Fi data, allowing the University to improve service quality and worker safety. Workers usage of the web-based platform is compulsory.
Memory for Where	Researchers collect participants' location via University Wi-Fi data, and later undertake an experiment in which participants recall where they were at certain times. This research seeks to understand people's memory for location and determine what factors influence the errors they make. Participants must opt-in to the study and are paid $50 for their data and $15 for completing the memory experiment.
Serving you Better	The University uses Wi-Fi location data and analysis techniques to monitor the use of university food outlets and common areas to better understand the factors that drive the use of these facilities, and how to provide students with a better university experience. No analysis of individual patterns will be undertaken, only aggregated statistical trends. Students can opt-out of the project if they wish.
Safe Campus	Academic researchers collaborate with a start-up firm to use university Wi-Fi data to develop a smartphone app that allows people to be matched with 'walking buddies' to keep them safe at night on campus.
Student Wellbeing Project	The University uses student data, including their W-Fi location data, in a learning analytics system to help improve mental health and identify at-risk students. Students must opt in to participate, and are paid $20 for doing so.
Project Move	The University collaborates with Yarra Trams to use university Wi-Fi location and timetable data to improve public transport availability and provide information about likely wait-times and occupancy.
Project TRIIBE	University campus shopping and food retailers, collaborating with researchers, use location and internet history data to develop methods to capture and analyse indoor shopping behaviour across shoppers' physical, online, and social environments, to improve customer experience. Participants are paid $100 for their data and must opt in to participate.
Project QueueSense	Researchers collect Wi-Fi location data at select locations on campus (such as café outlets) to develop algorithms to reduce queueing times at these locations. Coffee will be discounted during the 1-month period of data collection
Project Fluloc	University researchers collect data of social interactions based on indoor Wi-Fi tracking and online health questionnaire data, to assess the role of professional and educational environments in the spread of influenza.
Project Precinct Change Management	University researchers, collaborating with University Services and the Facilities Management Metro, develop algorithms and computational methods to understand the impact of construction disturbances on campus operations and develop methods for improved space use.
Impact of Attendance on Academic Performance	Researchers use Wi-Fi location data to track a cohort of students for one semester to understand how university attendance and presence affects academic and other outcomes.

https://doi.org/10.1371/journal.pone.0251964.t001

actually give up a lot of their personal data and privacy while still professing strong attitudes favoring personal privacy [27, 28].

Measures

Participants rated these scenarios in relation to 11 privacy dimensions on a 0–5 Likert scale. Table 2 is a list of these dimensions, the question posed to participants in relation to each dimension, and the labels used for its Likert scale. Only the most extreme points of the Likert scale were defined, and a slider bar was used between these points.

These 11 privacy dimensions were developed from an extensive literature review and a workshop event in which University of Melbourne stakeholders including students, academics (from Law, Engineering, Computer Science and Psychology) and professional staff from University Services 'brainstormed' the relevant privacy and acceptability dimensions that may affect social licence. The literature review considered research that examined the dimensions underlying ethical tracking-data collection and use from both a normative [4, 29, 30] and descriptive/empirical perspective, both qualitative [12] and quantitative [14–16]. The final 11 dimensions were drawn qualitatively from an analysis of these sources. For further details of this process, please see the S1 Text and S5 Table.

Procedure

Participants completed the following experiment (scripted through Qualtrics) on their own electronic devices through their web browser. Participants were instructed that they were

Table 2. Summary of privacy dimensions on which each scenario was rated.

Dimension	Question	Likert response
Decline Difficulty	How easy is it for people to decline participation in the proposed research?	0 = 'Extremely easy'
		5 = 'Extremely difficult'
Private Benefit	How much would private entities benefit from the proposed research?	0 = 'Not at all'
		5 = 'Extremely'
Participant Benefit	How much would participants (i.e., the people whose data is being collected) benefit from the proposed research?	0 = 'Not at all'
		5 = 'Extremely'
Public Benefit	How much would the public benefit from the proposed research?	0 = 'Not at all'
		5 = 'Extremely'
Disproportionality	To what extent are the researchers only collecting the data necessary to achieve the purposes of the proposed research?	0 = 'Researchers collecting only necessary data'
		5 = 'Researchers collecting vast amounts of unnecessary data'
Sensitivity	How sensitive is the data to be collected by the proposed research?	0 = 'Not at all sensitive'
		5 = 'Very sensitive'
Risk of Harm	How serious is the risk of harm that could arise from the proposed research?	0 = 'Extremely low risk of harm'
		5 = 'Extremely high risk of harm'
Trust	How much do you trust the sponsor of the proposed research?	0 = 'Not at all'
		5 = 'Extremely'
Data Security	How secure is the data that would be collected from the proposed research?	0 = 'Not at all secure'
		5 = 'Extremely secure'
Ongoing Control	To what extent do participants have ongoing control of their data? This includes controlling how and when data is collected, and having access to view and delete data after it is collected.	0 = 'No control at all'
		5 = 'Complete Control'
Respect for Privacy	To what extent do you believe the proposed research respects participants' privacy?	0 = 'Not at all'
		5 = 'Extremely'

https://doi.org/10.1371/journal.pone.0251964.t002 going to be shown 3 scenarios that they should read carefully. These 3 scenarios were randomly drawn from the list of 11 scenarios and shown to them sequentially. For each scenario, participants were given an attention check question which asked, "What was the previous scenario about?". There were 4 randomly ordered response options to this question: the correct answer for the scenario they had just read, and 3 decoy answers drawn randomly from the correct answers for the other scenarios. Participants that failed at least one of these attention check questions were excluded from analysis.

Participants were then asked to rate each scenario on a 0–5 Likert scale in relation to the 11 privacy dimensions. Participants rated each of their 3 randomly drawn scenarios contemporaneously on each dimension to allow for comparative scaling. While doing so, participants were provided with the scenario text below for them to refer to if necessary. The dimensions were presented to participants in the top–bottom order indicated in Table 2. After rating each scenario on each dimension, participants were asked 'whether the proposed use of university data in each scenario was acceptable' and gave a binary yes/no response. Finally, participants were asked some demographic questions: their age, gender, educational attainment, and their relationship to the University of Melbourne.

Ethics statement

This study received ethics approval from the University of Melbourne's psychology health and applied sciences human ethics sub-committee, approval number 1955555.1. All participants gave written informed consent prior to participating and were debriefed at the end of the experiment.

Statistical analysis

Descriptive model. Bayesian generalised linear mixed modeling (GLMM) was used to predict participant's acceptability judgements as an additive function of their privacy dimension judgements. Random intercept effects were included in the model to account for the dependency of the data introduced by different participants rating different scenarios. Thus, all random effects were blocked by participant and scenario. These random effects can be thought of as modelling the acceptability that is inherent in the participant or scenario that is not captured in the model's fixed parameters (the privacy dimension ratings and interactions therebetween). The privacy dimension Likert ratings were treated as numeric data for the purposes of modeling. The best-fitting model was also estimated with monotonic effects that preserve the ordinal nature of the Likert ratings [31]. This model performed worse for out-of-sample predictive accuracy (see S1 Text), providing support for treating these Likert ratings as numeric in all the analyses reported in this paper.

Posterior distributions of model parameters were estimated using Hamiltonian Markov Chain Monte Carlo No-U-turn Sampling implemented in the R package *brms* [32], a high level interface to Stan [33]. Four chains with 2000 iterations (including 1000 'burn-ins') were used. Uninformative priors were used for the intercept and random effect standard deviation parameters; both were Cauchy distributed with a location parameter of 0 and a scale parameter of 2.5 [34, 35]. Weakly informative priors were used for all fixed effect coefficients: a Laplacian (double exponential) distribution with a location parameter of 0 and a scale parameter of 0.2 (1 / range).

We developed two plausible candidate models: one which included one parameter for each privacy dimension, and one which also included some *a priori* interaction parameters between privacy dimensions (see S1 Table). Model selection between these models was conducted via Pareto-smoothed importance sampling leave-one-out cross validation using the *loo* package [36] in R: the model with the highest expected log predictive density (ELPD) was preferred.

To make inferences about the existence of effects, we used the Region of Practical Equivalence (to a null effect; ROPE) + 89% High Density Interval (HDI) decision rule: infer an effect if the entire 89% HDI falls outside of the ROPE [37, 38]. The region of practical equivalence was deemed to be any effect that causes less than a 5% increase or decrease from the mean acceptability proportion over the entire range of the variable. In this case, the ROPE was thus ± 0.098.

Predictive model. Data was randomly split into a training (80%) and testing set (20%) grouped by participant so that no participant's data was used in both training and testing. 3 models were fit on the training dataset via maximum likelihood estimation and then evaluated on the test dataset, using classification accuracy as the model selection metric. The logistic regression and the mixed effects logistic regression were estimated with *lme4* [39]. ElasticNet logistic regression was conducted using *caret* [40] and *glmnet* [41]. 5-repeat 10-fold cross validation was used to tune regularization hyperparameters, α and λ, via a grid search (all combinations of α and λ for $A = \{0, 0.05, 0.1, \ldots 1\}$ and $\Gamma = \{0, 0.05, 0.1, \ldots 2\}$), again using classification accuracy as the metric. For prediction in the test dataset, the random intercepts from the mixed effects model were dropped, and only fixed effects were used.

Results

Demographics

Participants were 314 members of the University of Melbourne community (198 female, 111 male, 5 prefer not to say, $M_{\text{age}} = 25.63$, $SD = 8.14$, range: 17–63 years). All participants gave their informed consent prior to the experiment. Twenty-seven participants failed at least one of three comprehension checks, leaving a final sample of 287 participants (184 female, 98 male, 5 prefer not to say, $M_{\text{age}} = 25.59$, $SD = 8.21$, range: 17–63 years).

We endeavoured to match our sample to the distribution of stakeholder types (e.g. undergraduate, postgraduate, academic employee etc.) of the university. Fig 1 shows the proportion of each type juxtaposed with the corresponding population figures taken from the University of Melbourne's 2019 Annual Report [42]. Our sample is representative of the students and staff of the University of Melbourne, but no population data is available for the number of other employees or others who regularly attend the university campus and our sample likely underrepresented these demographics.

Acceptance and perception of tracking-data analytics

Fig 2 shows the proportion of participants that labelled each scenario as acceptable. The mean acceptance proportion among the scenarios was 0.715 ($SD = 0.111$). Figs 3 and 4 show participants' perceptions of each of the scenarios in relation to each of the 11 privacy dimensions (ratings in Fig 3 are ordered by dimension, Fig 4 by scenario). For precise means and standard deviations of these privacy dimension ratings for each scenario see S2 Table.

Panels A–D of Fig 5 show mean participant acceptance judgements broken down by demographic variables. Omnibus ANOVA showed that mean acceptance did not differ by gender, $F(2, 284) = 0.449$, $p = 0.638$, relationship to University of Melbourne, $F(5, 281) = 0.605$, $p = 0.696$, educational attainment, $F(8, 276) = 0.993$, $p = 0.441$, or age group, $F(4, 276) = 0.377$, $p = 0.825$. Simple OLS regression analysis also showed that mean acceptance did not differ by age when treated as a continuous variable, $t(279) = -0.20$, $p = 0.844$.

Predicting acceptance of tracking-data analytics

Descriptive model. Bayesian mixed effects logistic regression was used to predict participants' acceptability judgements as an additive function of their privacy dimension judgements.

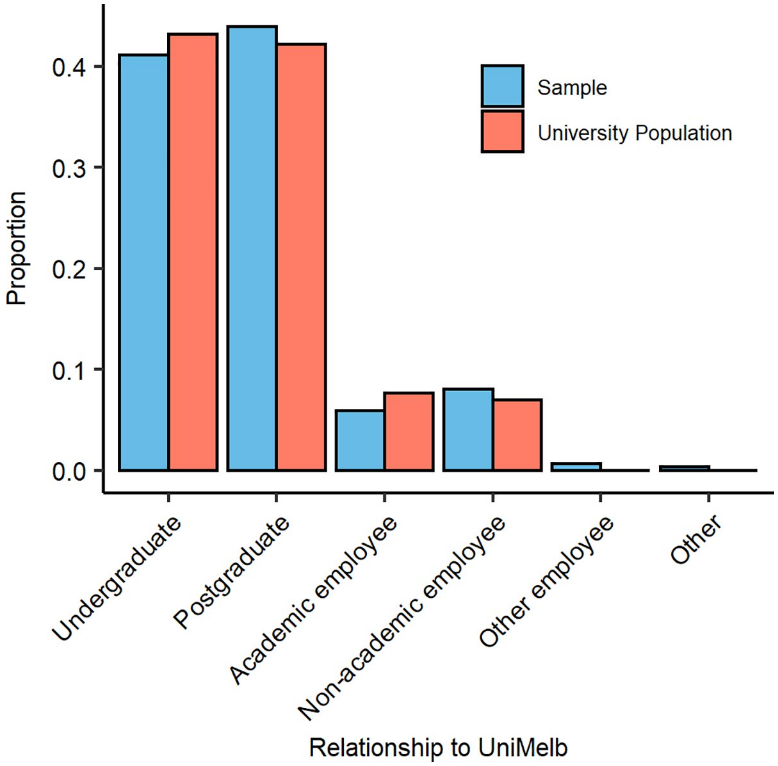

Fig 1. Sample characteristics compared to population. The proportion of participants in our sample (blue) from each 'relationship to UniMelb' category compared to that of the university population (red). Note, no population data was obtained for 'other employee' or 'other' categories.

https://doi.org/10.1371/journal.pone.0251964.g001

Two candidate models were estimated: one which included one parameter for each privacy dimension, and one which also included some *a priori* interaction parameters between privacy dimensions (listed in S1 Table).

Model selection was conducted via Pareto-smoothed importance sampling leave-one-out cross validation to maximise the out-of-sample predictive accuracy of the model [36, 43] and

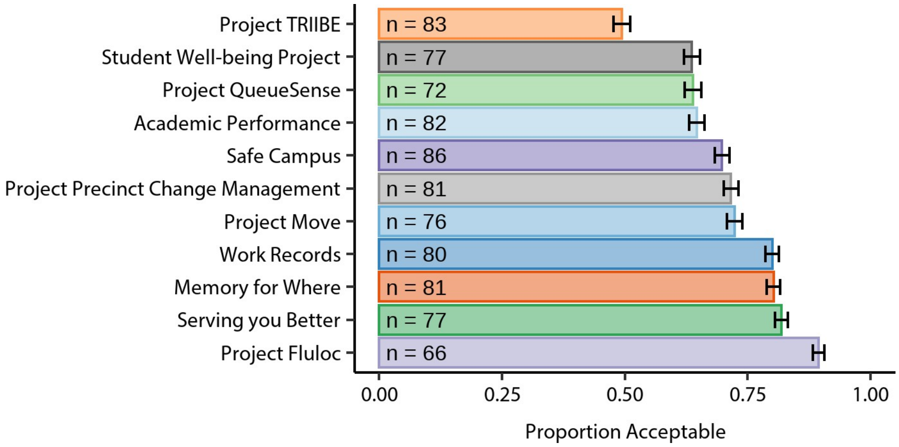

Fig 2. Participant perception of scenario acceptability. The proportion of 'acceptable' judgements for each scenario. The sample size for each scenario is listed at the base of the bar chart. Error bars represent standard errors of the mean.

https://doi.org/10.1371/journal.pone.0251964.g002

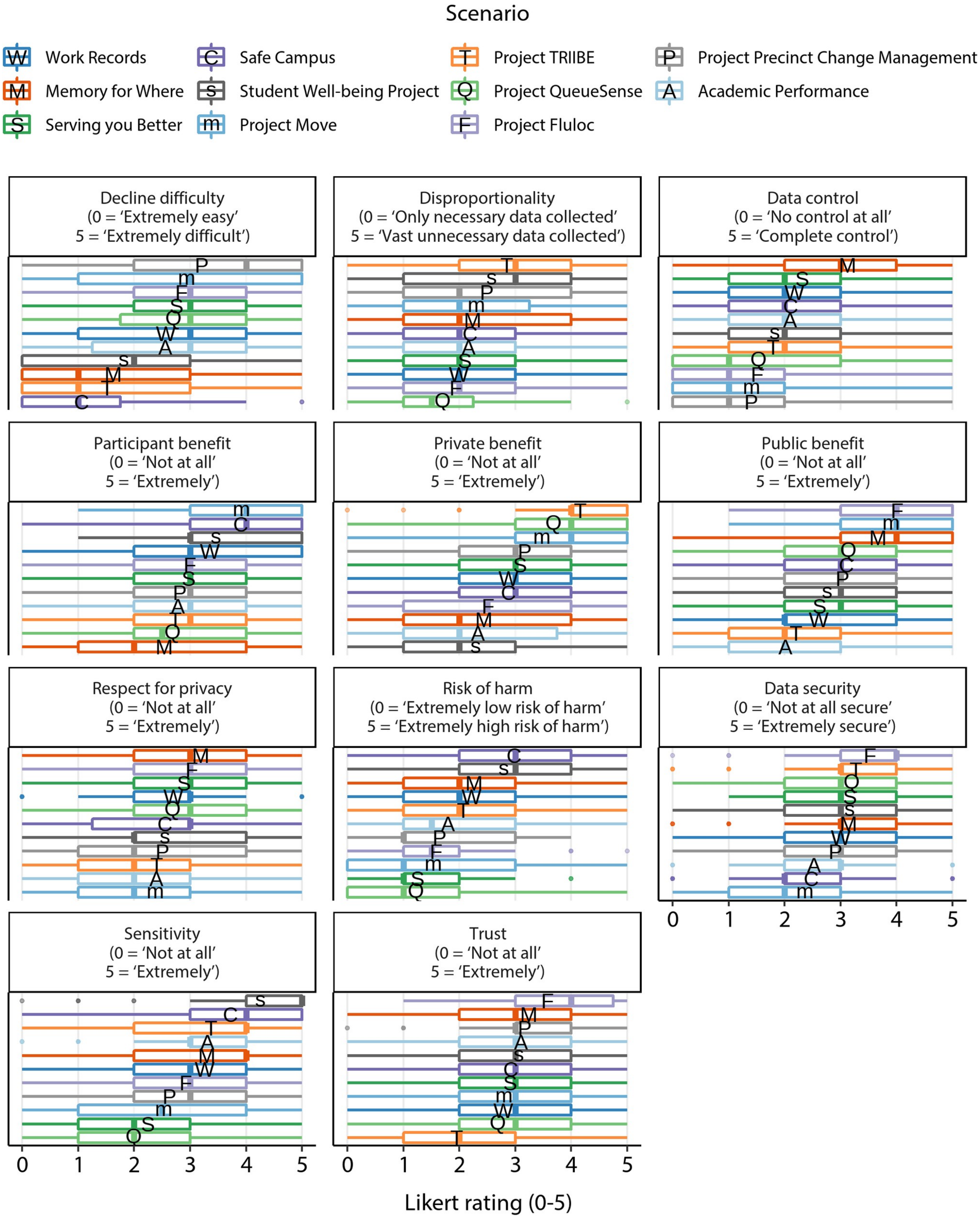

Fig 3. Participant ratings on privacy dimensions, by dimension. Participant ratings of each privacy dimension for each scenario, organised by privacy dimension. Letters on boxplots indicate the mean ratings of each privacy dimension for each scenario.

https://doi.org/10.1371/journal.pone.0251964.g003

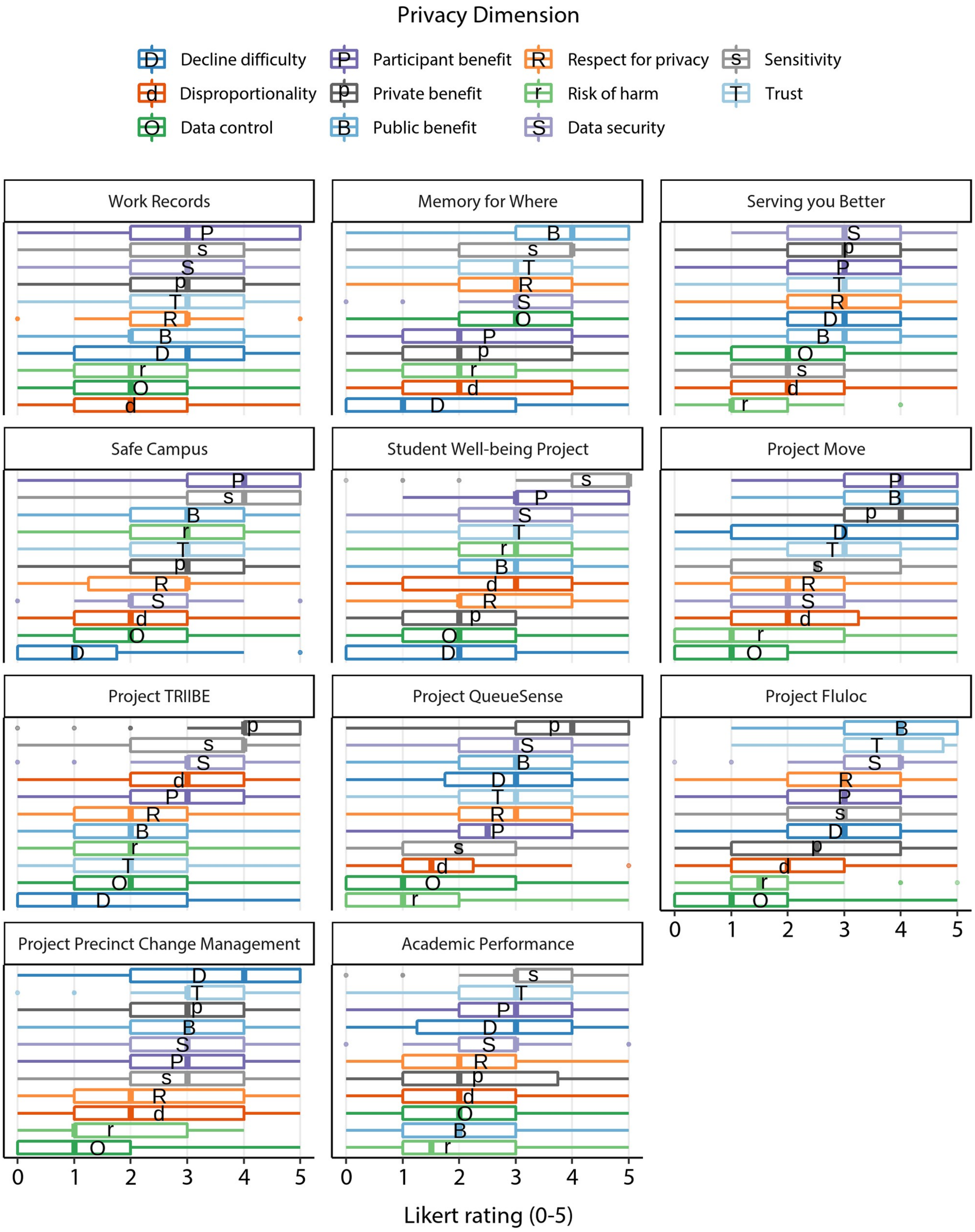

Fig 4. Participant ratings on privacy dimensions, by scenario. Participant ratings of each privacy dimension for each scenario, organised by scenario. Letters on boxplots indicate the mean ratings of each privacy dimension for each scenario.

https://doi.org/10.1371/journal.pone.0251964.g004

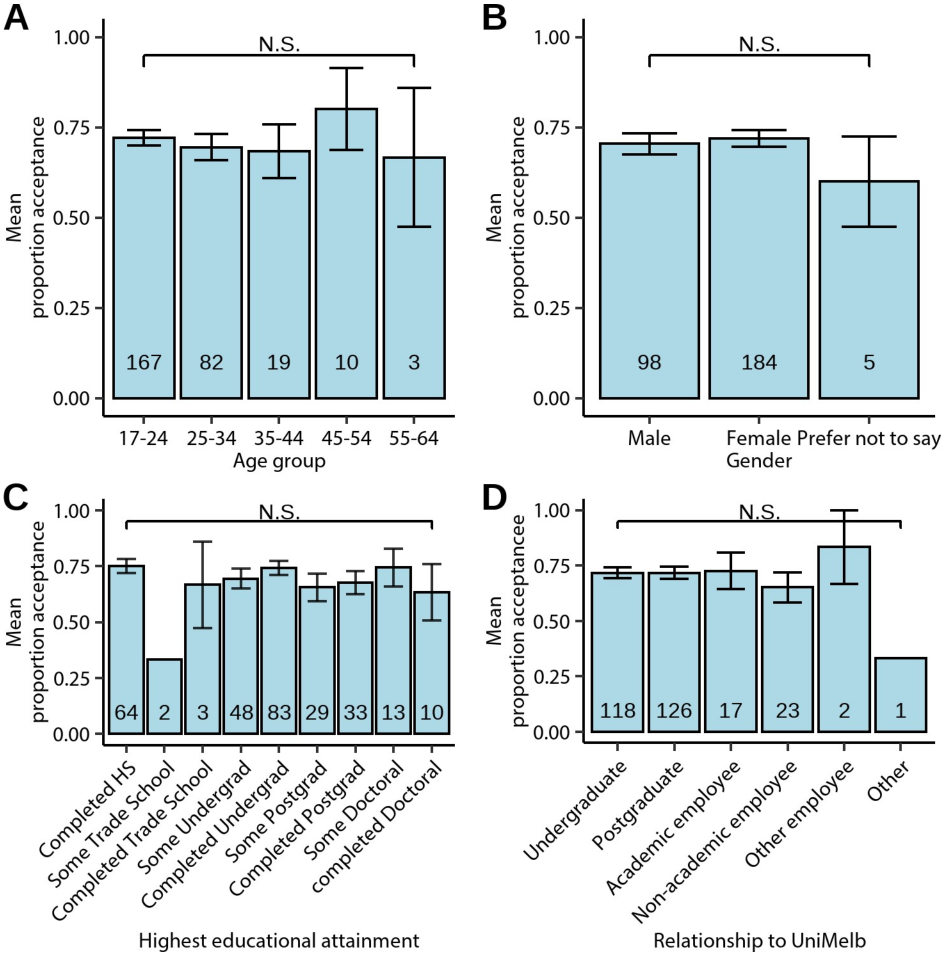

Fig 5. Acceptability of tracking-data analytics by demographics. (**A**-**D**) Mean participant acceptability proportion broken down by age group (**A**), gender (**B**), Educational attainment (**C**), and relationship to the University of Melbourne (**D**). Acceptability ratings did not differ by any of these demographic variables. Numbers at the bottom of the bars indicate the sample size in that group.

https://doi.org/10.1371/journal.pone.0251964.g005

the model with the highest expected log predictive density (ELPD) was preferred. On this basis, the model without interactions (ELPD = -345.22, *SE* = 17.54) was preferred to the model with interactions (ELPD = -349.66, *SE* = 18.80). The preferred model had a Nakagawa conditional R^2 of 0.454 (*SE* = 0.034), indicating that 45.6% of the variance in the data was explained by the random and fixed predictor variables, and a Nakagawa marginal R^2 of 0.398 (*SE* = 0.024) indicating that 39.8% of the variance in the data is explained by the fixed predictor variables, i.e., the privacy dimension ratings [44].

Posterior estimates of all fixed model parameters are shown in Fig 6 and listed in S3 Table. Following the Region of Practical Equivalence (to a null effect; ROPE) + 89% High Density Interval (HDI) decision rule, we inferred a practically relevant effect when the entire 89% HDI falls outside of the ROPE [37, 38]. On this basis, we inferred a predictive effect on acceptability judgements for ratings of respect for privacy, trust in the research sponsor, risk of harm, participant benefit from the tracking-data analytics, and data sensitivity. The belief that the research respected people's privacy had the biggest predictive effect on acceptability judgements, with a 1-unit increase in this rating resulting in a 1.72-factor increase in the odds of judging the project as acceptable. On the other hand, the smallest practically relevant effect

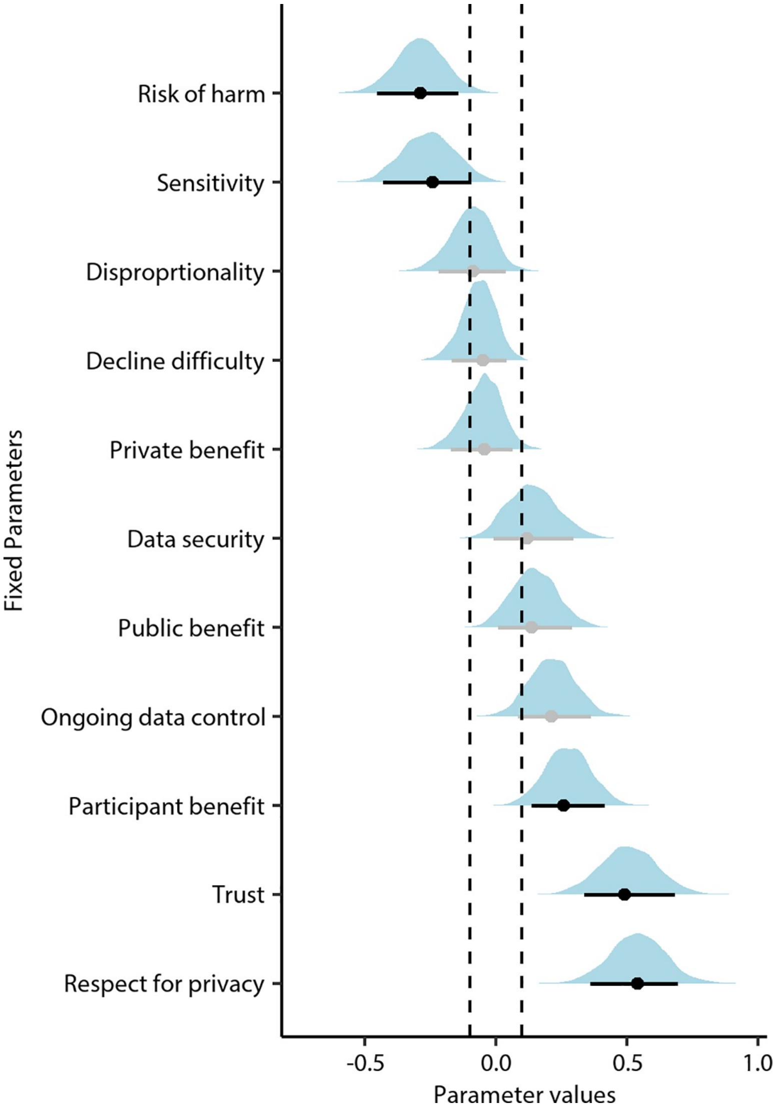

Fig 6. Fixed effect posterior distributions from the preferred model. Point estimates (dots) are posterior modes and intervals are 89% highest density intervals (lines). Parameters for which a practically significant effect was inferred have interval lines coloured black. The region of practical equivalence is shown between the dotted vertical lines. Practically significant effects were inferred when the entire 89% highest density interval fell outside the region of practical equivalence.

https://doi.org/10.1371/journal.pone.0251964.g006

was for the sensitivity of the data, with a 1-unit increase in this rating resulting in a 1.29-factor decrease (i.e., a 0.78-factor increase) in the odds of judging the project as acceptable. Finally, the posterior distributions of the random intercept effects are shown in S1 Fig.

Predictive model. We also developed a predictive model to assist decision-makers to estimate the social licence of future university research and service projects. We trained 3 models on a training dataset: a logistic regression, a mixed effects logistic regression with random intercepts for scenario and participant, and an ElasticNet logistic regression with regularisation hyperparameters, λ and α, chosen via 10-fold cross validation. The mixed effects logistic regression model had the best prediction accuracy (84.31%) on the test data set when compared to the logistic regression model (83.66%) and the ElasticNet logistic regression model

(83.66%). The regularisation hyperparameters of the ElasicNet regression were λ = 0 and α = 0.15.

The preferred model—the mixed effects logistic regression—had a Cohen's Kappa of 0.61, indicating that it had 'good' or 'substantial' classification performance [45, 46]. S3 Table shows the parameter coefficients for this model. Given privacy dimension ratings for a new project, these coefficients can be used to predict the likelihood of future acceptability judgements (via the logistic function). To facilitate the ease of this process, we have developed a simple online app (available at https://whitejp.shinyapps.io/lumas-predictive-model/) [47] in which, after entering ratings on each dimension, a user is given a probability in return. This probability can model the percentage of the university community that would accept a project with those ratings.

Discussion

In this study we aimed to explore and understand the social licence, or lack thereof, for the use of Wi-Fi location data in university research and service provision. More specifically, we sought to understand whether, to what extent, and under what circumstances, there is social license for location data collection and use on university campuses. To answer these questions, we presented participants—members of the University of Melbourne community (students, staff, and others)—with 11 hypothetical scenarios depicting the use of Wi-Fi location data on campus. We asked them to rate these scenarios in relation to 11 privacy dimensions, and finally to indicate whether they believed that the use of data in the scenario was acceptable. We then used participants' ratings on these 11 privacy dimensions to predict participant attitudes of acceptability, hoping to understand which dimensions are most predictive of social licence.

Understanding the social licence for tracking-data analytics

Our data generally support the notion that, for the hypothetical projects used in this study, there is wide support for the use of university Wi-Fi location data in university research and services. Indeed, the mean acceptance proportion of the scenarios was 71.5%. and all but one (Project TRIIBE) gained more than 50% acceptance. However, social licence is not conceptualised as merely a majoritarian enterprise but rather as requiring the broad support of the community [48]. Yet, even incorporating a higher acceptance criterion, our data is promising: four of the scenarios (Work Records, Memory for Where, Serving you Better, and Project Fluloc) exceeded 80% acceptability. Of course, exactly where the acceptability criterion should be set by decision-makers as the minimum to establish social licence is likely to depend on contextual factors, and is beyond the scope of this research.

Another consideration relevant to understanding social licence is not just the degree of acceptance of a project, but also how this acceptance is distributed; that is, support for a project must be manifest across the various constituent groups of a community [49]. Here, our results are promising: acceptability judgements did not differ as a function of gender, age, or educational attainment, suggesting that the acceptance of tracking-data analytics is not just concentrated to a few groups. Most relevantly, there were also no differences in acceptance judgements between staff (both academic and non-academic) and students (both under- and post-graduate). Given the different interests and incentives between these groups, it was perhaps surprising that acceptance ratings did not differ [48].

Factors predicting the acceptability of tracking-data analytics

Our results also give insight into the factors that are most predictive of the acceptable uses of tracking-data analytics on the University of Melbourne campus. Here, we find evidence that the most predictive factors are the perception that the research team respects individual privacy, the trust in the research team, the degree to which people personally benefit from the project, the sensitivity of the data being collected and used, and the risk of harm inherent in the project.

Surprisingly, however, the perception of the public benefit of a project did not show strong evidence of increasing the likelihood of project acceptability, nor did the perception of private or commercial benefit; only a personal benefit did. Thus, clearly identifiable and more imminent benefits that flow directly to those whose data are being collected and used, are more likely to increase social licence than more diffuse, amorphous public benefits.

Our findings also suggest limited predictive ability of a one-time opt-in/opt-out distinction on acceptability. Indeed, we found that perceptions of the initial difficulty of declining to have one's data collected and used in a project was less strongly predictive of acceptance than perceptions of how much ongoing control over their data users had (e.g., control over how and when one's data was collected, and personal ability to view and delete their data). This may have implications for privacy law which, in Australia and often elsewhere, considers consent as a requirement, not ongoing control of data [50].

Finally, the strong effect of trust in increasing the likelihood of acceptance in our sample is in line with previous work which suggested that trust was the primary factor underlying social licence in a very different context—for a proposed coal seam gas mine [51]. Similarly, the strong effect for 'respect for privacy' shows how acceptance of privacy-encroaching technologies relates to peoples' perceptions of how such technologies affect their privacy. As such, our findings are broadly in line with the notion that people undertake a privacy calculus in which they weigh the benefits of a technology with the negatives and risks of their privacy being compromised [52].

Estimating the acceptability of future tracking-data analytics projects

We further provide a simple approach to estimate the social licence of future university research and service projects that may be of use to university ethics boards or decision-makers. We have built a forward-looking predictive regression model which takes ratings of a university project that involves location data collection and analytics on 11 dimensions and predicts what proportion of the university community, given those ratings, will consider the project acceptable. Decision-makers could use this to estimate the acceptability for their proposed projects. To best do so, decision-makers could get a small sample of independent ratings for their project on the 11 dimensions, take the mean rating for each dimension, and input these into the model to obtain an estimate of the proportion of university community members who would likely view the scenario as acceptable, given the scenario ratings. To facilitate this, we have created an online app (available at https://whitejp.shinyapps.io/lumas-predictive-model/) [47]. This ability to prospectively estimate acceptability will allow decision-makers to either drop projects with little prospect of community support, or to make changes to project details (that are reflected by altered independent ratings on the 11 dimensions) to increase community acceptability.

Limitations

We should be careful not to extend these findings too far beyond the university domain. In particular, we should be careful not to generalise to significantly different contexts such as government surveillance for which there are specific considerations and a large literature [53–56]. Future work should consider the social licence for tracking-data use and collection in a wider context and sample [57–59]. Crucially, social licence requires openness and transparency to inspire trust and confidence in uses of data. The authors' note with emphasis that social licence is highly context dependent, and hope that the innovative methodological process described in this paper can be used to help uncover trustworthy uses of data in a given context, as part of an open dialogue with those who provide the data in the first instance. While the methodology offers a process for predicting acceptable uses, it is not a panacea but an additional tool in the suite of governance mechanisms for trustworthy data use.

Conclusion

These findings show that a large majority of the University of Melbourne community find the hypothetical scenarios presented to them acceptable. Further, our research suggests that trust of the research sponsor, the belief that they respect people's privacy, the benefit to participants of the research or service, and the sensitivity of the data collected, and the risk of harm imposed by the research, are all important factors which help determine the social licence of tracking-data analytics on university campuses. Researchers, university management, and those collecting location and other sensitive data, could use these results to cater their data collection and analytics methodology to community commands and expectations. Indeed, doing so will generally increase the efficacy of the data collection and avoid pushback, protest and negative publicity [60] that may arise from inadvertently stepping out of their social licence.

Considerations of the parameters of social licence of tracking-data analytics are only likely to grow in coming years as tracking-data collection and analysis methodologies become easier and more pervasive, and as a result of the widespread uptake of privacy-encroaching tracking technologies by governments around the world in response to the COVID-19 pandemic [57–59, 61].

Supporting information

S1 Text. Supplementary materials and methods, and results.
(DOCX)

S1 Fig. Random effect posterior distributions. **(A)** Posterior distributions of scenario random intercepts from the preferred model. Point estimates are posterior modes and intervals are 89% highest density intervals. **(B)** Summary of participant random intercept posterior distributions. Black point estimates are posterior means, red point estimates are posterior modes, and blue intervals are 89% highest density intervals.
(TIF)

S1 Table. Interaction effects included in candidate model.
(DOCX)

S2 Table. Mean and standard deviations (SD) of participant ratings on privacy dimensions for each scenario.
(DOCX)

S3 Table. Parameter posterior distribution summary statistics for model parameters.
(DOCX)

S4 Table. Parameter coefficients for predictive model.
(DOCX)

S5 Table. Developing final privacy dimensions from workshop dimensions.
(DOCX)

Acknowledgments

The authors would like to thank all members of the University of Melbourne community that participated in the LUMAS Social Licence Workshop which helped refine and conceive the idea and methods for this research. We would also like to thank the participants who completed the experiment forming the basis of this paper.

Author Contributions

Conceptualization: Simon Dennis, Martin Tomko, Jessica Bell, Stephan Winter.

Data curation: Joshua P. White.

Formal analysis: Joshua P. White.

Funding acquisition: Simon Dennis, Stephan Winter.

Investigation: Joshua P. White.

Methodology: Joshua P. White, Simon Dennis, Martin Tomko, Jessica Bell, Stephan Winter.

Project administration: Joshua P. White, Simon Dennis, Stephan Winter.

Resources: Joshua P. White, Simon Dennis, Martin Tomko, Stephan Winter.

Supervision: Simon Dennis.

Writing – original draft: Joshua P. White.

Writing – review & editing: Joshua P. White, Simon Dennis, Martin Tomko, Jessica Bell, Stephan Winter.

References

1. Sapiezynski P, Stopczynski A, Gatej R, Lehmann S. Tracking Human Mobility Using WiFi Signals. PLOS ONE. 2015; 10: e0130824. https://doi.org/10.1371/journal.pone.0130824 PMID: 26132115
2. Kalogianni E, Sileryte R, Lam M, Zhou K, Van der Ham M, Van der Spek S, et al. Passive WiFi monitoring of the rhythm of the campus. 2015. pp. 9–14.
3. Menon ME, Terkla DG, Gibbs P. Using data to improve higher education: research, policy and practice. 2014.
4. Ballantyne A. Where is the human in the data? A guide to ethical data use. GigaScience. 2018; 7. https://doi.org/10.1093/gigascience/giy076 PMID: 29961892
5. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics. 2015; 41: 404–409. https://doi.org/10.1136/medethics-2014-102374 PMID: 25617016
6. Cook H. University students, you are being watched. The Age. 11 Aug 2016. Available: https://www.theage.com.au/education/university-students-you-are-being-watched-20160811-gqqet7.html. Accessed 29 Aug 2019.
7. The Australian Broadcasting Corporation. University of Melbourne defends wi-fi tracking as planning move amid privacy concerns. ABC News. 12 Aug 2016. Available: https://www.abc.net.au/news/2016-08-12/university-of-melbourne-tracking-students-through-wifi/7723468. Accessed 29 Aug 2019.
8. The University of Melbourne. University statement on Wifi analysis on campus. In: The Univerisity of Melbourne Newsroom [Internet]. 12 Aug 2016 [cited 29 Aug 2019]. Available: https://about.unimelb.edu.au/newsroom/news/2016/august/university-statement-on-wifi-analysis-on-campus
9. Xiao A. University rejects outrage over mobile phone tracking. Australian Broadcasting Corporation. 3 Feb 2020. Available: https://www.abc.net.au/news/2020-02-04/newcastle-university-tracking-student-attendance-through-mobile/11915502. Accessed 16 Sep 2020.
10. Harwell D. Colleges are turning students' phones into surveillance machines, tracking the locations of hundreds of thousands. Washington Post. 24 Dec 2019. Available: https://www.bostonglobe.com/news/nation/2019/12/24/colleges-are-turning-students-phones-into-surveillance-machines-tracking-locations-hundreds-thousands/mxer5QFzMaFaa7eXCoEZcJ/story.html. Accessed 16 Sep 2020.

11. Mogg T. Alabama University Tracks Students to See If They Leave Football Games. Digital Trends. 13 Sep 2019. Available: https://www.digitaltrends.com/mobile/university-of-alabama-tracks-students-to-see-if-they-leave-football-games/. Accessed 16 Sep 2020.

12. Gulliver P, Jonas M, McIntosh T, Fanslow J, Waayer D. Qualitative research: Surveys, social licence and the integrated data infrastructure. Aotearoa N Z Soc Work. 2018; 30: 57.

13. Prinsloo P, Slade S. Student data privacy and institutional accountability in an age of surveillance. Using Data to Improve Higher Education. 2014. pp. 195–214. Available: https://brill.com/view/book/edcoll/9789462097940/BP000013.xml

14. Data Futures Partnership. A path to social license: Guidelines for trusted data use. New Zealand: Data Futures Partnership; 2017.

15. Data Futures Partnership. Our data, our way: What New Zealand people expect from guidelines for data use and sharing. New Zealand: Data Futures Partnership; 2017.

16. Boutilier RG. A Measure of the Social License to Operate for Infrastructure and Extractive Projects. Rochester, NY: Social Science Research Network; 2017 Nov. Report No.: ID 3204005. Available: https://papers.ssrn.com/abstract=3204005

17. Veale M, Binns R, Ausloos J. When data protection by design and data subject rights clash. Int Data Priv Law. 2018; 8: 105–123. https://doi.org/10.1093/idpl/ipy002

18. Cunche M. I know your MAC address: targeted tracking of individual using Wi-Fi. J Comput Virol Hacking Tech. 2014; 10: 219–227. https://doi.org/10.1007/s11416-013-0196-1

19. Krumm J. A survey of computational location privacy. Pers Ubiquitous Comput. 2009; 13: 391–399. https://doi.org/10.1007/s00779-008-0212-5

20. Ardagna CA, Cremonini M, Damiani E, De Capitani di Vimercati S, Samarati P. Location Privacy Protection Through Obfuscation-Based Techniques. In: Barker S, Ahn G-J, editors. Data and Applications Security XXI. Berlin, Heidelberg: Springer; 2007. pp. 47–60. https://doi.org/10.1007/978-3-540-73538-0_4

21. Zou H, Chen Z, Jiang H, Xie L, Spanos C. Accurate indoor localization and tracking using mobile phone inertial sensors, WiFi and iBeacon. 2017 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL). 2017. pp. 1–4. https://doi.org/10.1109/ISISS.2017.7935650

22. How To Use Wi-Fi Networks To Ensure a Safe Return to Campus—IEEE Spectrum. In: IEEE Spectrum: Technology, Engineering, and Science News [Internet]. [cited 25 Aug 2020]. Available: https://spectrum.ieee.org/view-from-the-valley/telecom/wireless/want-to-return-to-campus-safely-tap-wifi-network

23. Ciuccarelli P, Lupi G, Simeone L. Visualizing the Data City: Social Media as a Source of Knowledge for Urban Planning and Management. Springer International Publishing; 2014. Available: https://www.springer.com/gp/book/9783319021942 PMID: 25115552

24. Scanlon P, Smeaton AF. Using WiFi Technology to Identify Student Activities Within a Bounded Environment. In: Lavoué É, Drachsler H, Verbert K, Broisin J, Pérez-Sanagustín M, editors. Data Driven Approaches in Digital Education. Cham: Springer International Publishing; 2017. pp. 473–478. https://doi.org/10.1007/978-3-319-66610-5_45

25. Boutilier RG. Frequently asked questions about the social licence to operate. Impact Assess Proj Apprais. 2014; 32: 263–272. https://doi.org/10.1080/14615517.2014.941141

26. Krasnova H, Spiekermann S, Koroleva K, Hildebrand T. Online Social Networks: Why We Disclose. J Inf Technol. 2010; 25: 109–125. https://doi.org/10.1057/jit.2010.6

27. Kokolakis S. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Comput Secur. 2017; 64: 122–134. https://doi.org/10.1016/j.cose.2015.07.002

28. Barth S, de Jong MDT. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. Telemat Inform. 2017; 34: 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

29. Marx GT. Ethics for the New Surveillance. Inf Soc. 1998; 14: 171–185. https://doi.org/10.1080/019722498128809

30. Pounder CNM. Nine principles for assessing whether privacy is protected in a surveillance society. Identity Inf Soc. 2008; 1: 1–22. https://doi.org/10.1007/s12394-008-0002-2

31. Bürkner P-C, Charpentier E. Modelling monotonic effects of ordinal predictors in Bayesian regression models. Br J Math Stat Psychol. n/a. https://doi.org/10.1111/bmsp.12195 PMID: 31943157

32. Bürkner P-C. brms: an r package for bayesian multilevel models using stan. J Stat Softw. 2017; 80: 1–28. https://doi.org/10.18637/jss.v080.i01

33. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Softw. 2017; 76: 1–32. https://doi.org/10.18637/jss.v076.i01

34. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 2006; 1: 515–534. https://doi.org/10.1214/06-BA117A

35. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat. 2008; 2: 1360–1383. https://doi.org/10.1214/08-AOAS191

36. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017; 27: 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

37. Kruschke JK. Rejecting or Accepting Parameter Values in Bayesian Estimation: Adv Methods Pract Psychol Sci. 2018 [cited 17 Jun 2020]. https://doi.org/10.1177/2515245918771304

38. Kruschke JK, Liddell TM. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychon Bull Rev. 2018; 25: 178–206. https://doi.org/10.3758/s13423-016-1221-4 PMID: 28176294

39. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015; 67: 1–48. https://doi.org/10.18637/jss.v067.i01

40. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw. 2008; 28: 1–26. https://doi.org/10.18637/jss.v028.i07 PMID: 27774042

41. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33: 1–22. https://doi.org/10.18637/jss.v033.i01 PMID: 20808728

42. The University of Melbourne. University of Melbourne 2019 Annual Report. 2019. Available: https://about.unimelb.edu.au/__data/assets/pdf_file/0024/156291/AR-PRINT-2019-WEB.pdf

43. Vehtari A, Simpson D, Gelman A, Yao Y, Gabry J. Pareto smoothed importance sampling. ArXiv150702646 Stat. 2019 [cited 17 Aug 2020]. Available: http://arxiv.org/abs/1507.02646

44. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol. 2013; 4: 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

45. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. John Wiley & Sons; 2013.

46. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977; 33: 159–174. https://doi.org/10.2307/2529310 PMID: 843571

47. Joshua P White. Predicting the Social Licence of University Tracking-data Analytics Web App. [cited 18 Apr 2021]. Available: https://whitejp.shinyapps.io/lumas-predictive-model/

48. Wilburn KM, Wilburn R. Achieving social license to operate using stakeholder theory. J Int Bus Ethics. 2011; 4: 3–16.

49. Gunningham N, Kagan RA, Thornton D. Social License and Environmental Protection: Why Businesses Go Beyond Compliance. Law Soc Inq. 2004; 29: 307–341. https://doi.org/10.1111/j.1747-4469.2004.tb00338.x

50. Parliament of the Commonwealth of Australia. Privacy Act 1988 (Cth), Schedule 1 (Australian Privacy Principles). Sect. Schedule 1.

51. Moffat K, Zhang A. The paths to social licence to operate: An integrative model explaining community acceptance of mining. Resour Policy. 2014; 39: 61–70. https://doi.org/10.1016/j.resourpol.2013.11.003

52. Dienlin T, Metzger MJ. An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample. J Comput-Mediat Commun. 2016; 21: 368–383. https://doi.org/10.1111/jcc4.12163

53. Nam T. What determines the acceptance of government surveillance? Examining the influence of information privacy correlates. Soc Sci J. 2018 [cited 2 Aug 2019]. https://doi.org/10.1016/j.soscij.2018.10.001

54. Nam T. Untangling the relationship between surveillance concerns and acceptability. Int J Inf Manag. 2018; 38: 262–269. https://doi.org/10.1016/j.ijinfomgt.2017.10.007

55. Nam T. Does ideology matter for surveillance concerns? Telemat Inform. 2017; 34: 1572–1585. https://doi.org/10.1016/j.tele.2017.07.004

56. Friedewald M, Burgess JP, Čas J, Bellanova R, Peissl W. Surveillance, Privacy and Security: Citizens' Perspectives. Taylor & Francis; 2017.

57. Garrett PM, White JP, Lewandowsky S, Kashima Y, Perfors A, Little DR, et al. The acceptability and uptake of smartphone tracking for COVID-19 in Australia. PLOS ONE. 2021; 16: e0244827. https://doi.org/10.1371/journal.pone.0244827 PMID: 33481841

58. Garrett PM, Wang Y, White JP, Hsieh S, Strong C, Lee Y-C, et al. Young adults view smartphone tracking technologies for COVID-19 as acceptable: the case of Taiwan. Int J Environ Res Public Health. 2021; 18: 1332. https://doi.org/10.3390/ijerph18031332 PMID: 33540628

59. Lewandowsky S, Dennis S, Perfors A, Kashima Y, White JP, Garrett P, et al. Public acceptance of privacy-encroaching policies to address the COVID-19 pandemic in the United Kingdom. PLOS ONE. 2021; 16: e0245740. https://doi.org/10.1371/journal.pone.0245740 PMID: 33481877

60. Franks DM, Davis R, Bebbington AJ, Ali SH, Kemp D, Scurrah M. Conflict translates environmental and social risk into business costs. Proc Natl Acad Sci. 2014; 111: 7576–7581. https://doi.org/10.1073/pnas.1405135111 PMID: 24821758

61. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, Nadai MD, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv. 2020; 6: eabc0764. https://doi.org/10.1126/sciadv.abc0764 PMID: 32548274