Better the Devil You Know: Exposing the Data Sharing Practices of Smartphone Apps

Max Van Kleek* {max.van.kleek}
Ilaria Liccardi† {ilaria}
Reuben Binns* {reuben.binns}
Jun Zhao* {jun.zhao}
Daniel J. Weitzner† {djweitzner}
Nigel Shadbolt* {nigel.shadbolt}

*{}@cs.ox.ac.uk
Department of Computer Science
University of Oxford, UK

†{}csail.mit.edu
MIT CSAIL
Cambridge, MA, USA

ABSTRACT
Most users of smartphone apps remain unaware of what data about them is being collected, by whom, and how these data are being used. In this mixed methods investigation, we examine the question of whether revealing key data collection practices of smartphone apps may help people make more informed privacy-related decisions. To investigate this question, we designed and prototyped a new class of privacy indicators, called Data Controller Indicators (DCIs), that expose previously hidden information flows out of the apps. Our lab study of DCIs suggests that such indicators do support people in making more confident and consistent choices, informed by a more diverse range of factors, including the number and nature of third-party companies that access users' data. Furthermore, personalised DCIs, which are contextualised against the other apps an individual already uses, enable them to reason effectively about the differential impacts on their overall information exposure.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; K.4.1 Computers and Society: Privacy

Author Keywords
Decision-making; Personal data; Mobile apps; Privacy Indicators

INTRODUCTION
Apps provide a huge variety of functionality that an estimated 1.91 billion smartphone users worldwide take advantage of each day [49]. At the same time, apps are conduits through which substantial volumes of information about people are being captured, and transferred into the hands of advertisers, market researchers, e-commerce companies, among others [42, 54]. Increasingly, these troves of personal data are being used to determine prices, opportunities, and decisions both online and offline [19, 38, 43, 55].

Faced with this opaque world of data harvesting, smartphone users have been left feeling that they lack adequate understanding to make informed decisions regarding their privacy [17, 21, 34, 48]. As a result, some users have chosen to withdraw from using apps to their full extent, in an attempt to limit data exposure [11, 40, 41]. Previous research has shown that users have skewed expectations about how their data are being collected and by whom [26, 29], and overestimate the procedures in place to vet which apps are available through app stores [27].

Greater transparency may go some way to alleviating these fears. If the collection activities of the many first and third party trackers were open to scrutiny, users might be able to develop better mental models, better evaluate risks, and ultimately make more informed privacy choices. In turn, these choices could exert competitive pressure on developers to bring their privacy practices in line with actual users preferences expressed in the market, such as by reducing the range and type of third-party trackers included in their applications.

This paper presents the results of a mixed methods exploration of making the invisible visible, namely, allowing end-user individuals to inspect and assess the hidden data transmission behaviours of smartphone apps. Through an iterative design process, we devised a new class of privacy indicators [15] called *Data Controller Indicators* (DCIs), that disclose the kinds of data sent by apps to various organisations, background information about each, and the likely purposes for which the data will be used. We also developed a user-personalised version of a DCI, that shows when data being transmitted to a particular organisation is also already provided by other apps they already use. We tested Data Controller Indicators in an app-choice task, and found that they caused participants to consider more potential privacy implications, while simultaneously allowing them to make more consistent choices, with more confidence, than with traditional interfaces.

The objectives of this work are to answer the following questions:

1. *Designing Indicators* - How should data collection activities be conveyed to support privacy-related decisions? Should such activities be contextualised against other apps the individual uses?

2. *Effects on Decision Making* - If indicators revealing data dissemination behaviours are presented at time of an app-choice decision, will it affect the decision? What aspects of such decisions are affected, e.g. would it change the decision process, degree of confidence, or change the outcome entirely?

CHI 2017, May 06-11, 2017, Denver, CO, USA

 ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025556

We pursued a mixed methods approach to investigate these questions. We first designed and built Data Controller Indicators (DCIs), including infrastructure to intercept information sent by smartphone apps, and a technique for transforming raw app data transmission information into descriptive high-level characterisations of *what* was being shared with *whom* (#1). We then conducted a lab study (#2), consisting of a series of app choice role-play tasks in which participants were asked to think-aloud as they were given different interfaces with which to choose an app to install and use. This was followed by a series of semi-structured interviews, in which they were asked to reflect on the interface conditions. We found that when using our novel DCI interfaces, participants chose different applications and had more confidence in their choices, compared to other interfaces. They were also able to consider a larger range of factors, and able to reason about how the introduction of a new app might affect the overall exposure of their information to third parties.

BACKGROUND

Users' willingness to share information is dependent on a variety of factors, including the type of information collected, how sensitive it is, how long it will be stored, how it will be used, or levels of control [30, 31]. In the context of smartphones these preferences can be highly idiosyncratic and context-dependent, sensitive to individual nuances and the impact of exogenous events [37].

Under the existing 'permissions' model used by smartphone platforms to inform users of the kinds of data accessed by apps, users generally do not understand the implications or pay attention to such requests [21, 27], have sufficient information to make rational decisions [17], or feel able to represent the kinds of complex rules about information sharing that they would like to specify [10]. Given such limitations, many researchers have proposed and evaluated alternative approaches, which often involve contextualising [47], summarising [34] or augmenting existing permission information [33]. Since users' expectations are frequently at odds with the actual behaviours of apps [45], another promising approach involves focusing user attention on app behaviours that they are not likely to expect [34].

In addition to exploring new ways to present privacy-related information to users, researchers have also explored *when* to present it; for instance, either before the installation of an app [28, 32] or during the usage of an app [2, 6, 8, 23]. A 'nudge' approach, in which interventions are made at key moments in order to sway user behaviour towards more privacy-respecting behaviour, has also been found to be effective [2, 7, 52].

In this paper, we seek to build upon recent work on *privacy indicators*. Privacy indicators are an approach to simplifying potential privacy risks by summarising them visually or some other method, such as a score or grade [33, 34]. Some privacy indicators have explored displaying a summary whether an app exposes data to third parties, and tested the effect of such indicators on user's privacy decisions at app install time [6, 28, 51]. Others have addressed the problem of individual differences in individual privacy preferences, by creating personalised privacy indicators and agents acting as adaptive indicators [35, 36, 53]. Beyond apps, related research pertaining to the *Internet-of-Things* and technologies for the *Quantified Self* has explored how visualisations for supporting better end-user to understanding of personal data flows [13, 50]. Our work extends such efforts with a new kind of privacy indicator called Data Controller Indicators (DCIs), that breaks down the data being collected by smartphone apps by the organisations doing the collecting, alongside with rich contextual information about these organisations and their purposes for collecting the data.

The other contribution of our work is exploring the effectiveness of a *personalised privacy indicator*, i.e. a personalised DCI. Drawing inspiration from recent proposals which provide users with an overview of what data they may have already given away [3, 25], our personalised DCIs will highlight the new privacy cost for installing an app, in contrast to existing privacy loss of an individual, based on the apps already installed on the smartphone. This means that if installing a new app will not introduce new types of personal data to be accessed by a third party that has never accessed an individual's data yet, then the privacy risk associated with this new app is low for that individual. We are the first study to explore how this personalised privacy risk presentation could impact on an individual's privacy decision making.

DATA CONTROLLER INDICATORS

This section describes the design and development of our *Data Controller Indicators*. To ensure ecological validity, we started by investigating a variety of approaches for analysing information collection activities of real apps. This initial investigation informed and provided essential constraints for our iterative design phase, in which we explored different visual representations of this information to derive the final design we used in the subsequent lab study.

Building Models of Information Flows

Growing concern over data collection practices has motivated privacy researchers to develop methods to detect apps' data collection and dissemination activities. We first conducted a wide survey of work in this area, discovering three general approaches: static analysis, in which app binaries are decompiled and analysed to identify indicators of data collection activities (e.g. [4, 9, 16]), OS instrumentation, in which sensitive data is tagged and monitored as it flows through the system (e.g. TaintDroid [18], and AppTrace [44]), and network traffic monitoring (e.g. Recon [46]) in which data transmitted by apps is intercepted and analysed. For our project, we chose network monitoring because it was reliable [54], simple, and device and OS agnostic–meaning that practically any network-connected device could be inspected in the same way. There are, however, limitations to this approach, which we discuss at the end of the paper.

Apparatus, Modelling Workflow, and Data Capture Method

Figure 1 illustrates our data capture and modelling workflow. A smartphone was configured to connect to the Internet through a proxy server running the mitmproxy software [14]. The app was then launched on the smartphone, and a 10 minute tour was taken of the app, with the goal of invoking each piece of functionality at least once. This method was devised to ensure that a representative sample of traffic data would be generated by the app and captured. The raw log files were then transformed to higher-level descriptions through a dual-data processing pipeline, illustrated in the bottom of Figure 1. In the first, data detectors were applied to the raw log files to identify kinds of data being sent. For the purposes of this study, four broad categories of data were used: (*Phone ID*, *Phone characteristics*, *Location*, *Personal attributes*).

The second pipeline mapped individual hosts to organisations that owned them. For each organisation, a profile was then gathered, which included information on parent and subsidiary organisations (if any), information about the type of business, server geographic locations, organisational country of origins, date founded, number of employees/members, names of its directors, and a short textual history. Data sources that were used for this include *Crunchbase*, *Wikipedia* and *OpenCorporates*. For the purposes of building an initial collection of app data sharing models for the study, we selected the top 50 most popular apps available from the iTunes and Google Play stores (combined, ranked by number of downloads), and analysed them using the method described above.

Designing the Indicators

To design the indicators, we started with a set of prototypes inspired by Privacy Leaks [5], app information flows [54], and trace views [3]. An iterative, user-centred design process was then followed to refine the prototypes (visible in Figure 2) to yield the final set of *Data Controller Indicators*. During this process, we pursued several representations simultaneously: a directed graph (inspired by trace views), a tabular design (inspired by Privacy Leaks), and a nested hierarchical box representation (not shown). In its final design, two representations remained; a table representation (left), in which each row represents a distinct organisation collecting data in an app, of the types indicated by the columns, for the purposes indicated by particular colours. The Sankey version (right) has flowing paths that connect to boxes representing organisations, data types, and data uses/purposes. In both representations, detailed information about each company, data type, or purpose is displayed in a pop-up box that appears for each item hovered upon.

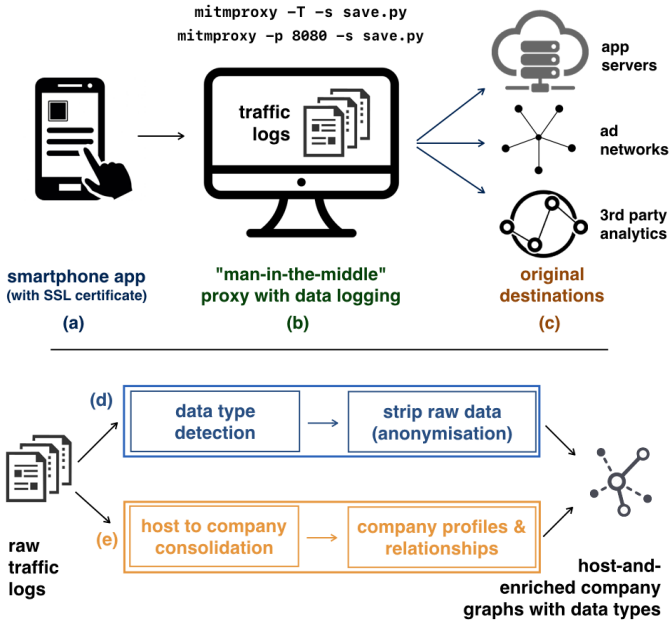

Figure 1. ***Data capture and modeling workflow*** **- A smartphone (a) is configured with the special mitm SSL certificate, and configured to use proxy server (b) as an HTTP proxy. Experimenter uses apps as a normal user would. Traffic generated is captured and written to log files, and forwarded on to original destinations (c). Resulting data logs are processed in two pipelines, data detection pipeline (d) and company detection pipeline (e).**

Support for Personalisation (exploring differential exposure) A key hypothesis we wished to explore was whether contextualising a particular app's data dissemination activities against those of other apps a person uses might allow them to reason about *differential exposure*, i.e. the relative risk of introducing a new app to the set of apps they already use. To investigate this question, we extended both interfaces to show those activities that are unique to a particular app, versus those in common with others. We refer to these indicators as PDCIs (Personalised Data Controller Indicators) to emphasise that they are personalised to a person's particular app ecosystem, i.e., the other apps they use. In the Table view in Figure 2, for instance, the PDCI view shows an app contextualised against 2 other apps; rows in yellow above the grey line are those that are unique to the app being examined, while those beneath are in common with the others. In the Sankey version (right), all apps are displayed on the left, unique information flows are highlighted in pink.

METHODOLOGY

Our main study was designed to test Data Controller Indicators in a realistic privacy-related decision-making setting. Specifically, we sought to measure whether people would make different decisions with DCIs than without them. Second, we wished to understand whether their decisions would follow different lines of reasoning, or consider different aspects when making choices with DCIs. Finally, we aimed to identify whether people would prefer DCIs, and the particular forms that they found most useful.

Framing these question as hypotheses, we wished to test the following :

- *Compared to permissions-based interfaces, Data Controller Indicators (DCIs) enable a more diverse range of strategies in people's privacy decisions, (*H:Strategy*), resulting in greater consistency of choices (*H:AppChoice*), and higher confidence (*H:Confidence*). However, choices with DCIs take slightly longer (*H:Time*), due to additional time needed to consider all the information presented.*
- *Compared to visualising traffic by hostname or domain, DCI's approach of representing destinations as entities (i.e. organisations and companies) (*H:Entities*), annotated with background information (*H:Background*), allows people to make better judgements about the companies that collect their data.*
- *Personalising DCIs, by contextualising the data collection activities of an app against other apps people already use, helps people to focus on new risks posed by that app (*H:Personalise*).*
- *People prefer DCI and PDCI interfaces to permissions-based interfaces for privacy-related decisions because of the information these indicators provide (*H:Preference*).*

For the privacy-related decision, we focused on *choosing a new app to install and use*. While this is far from the only kind of app-related privacy decision people make, it is both commonly occurring and important, because choosing a malicious or privacy-invasive app could drastically impact a person's privacy profile[1].

[1]While there is some evidence that privacy notices have greater salience *after* installation [8], by then it is usually too late and data has already been shared.

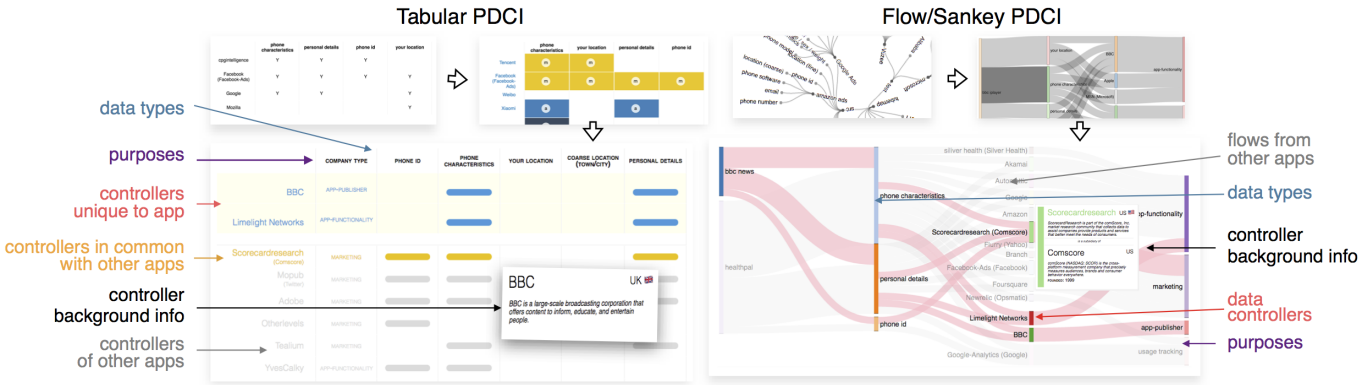

Figure 2. ***Designing the Data Controller Indicators*** **- Illustration of iterations of two different DCI visualisations: Tabular and Flow/Sankey. Each of three design iterations are displayed for each type, clockwise from the top left. Final versions of each type, with personalisation enabled, (PDCI) are at the bottom.**

Study Design

The main study consisted of a sequence of app-choice tasks, in which participants were asked to choose between two functionally equivalent apps for their personal use, in a simulated app store setting. During each task, participants were presented information about each app using one of five interfaces, shown in Figure 3. The **Permissions** interface reproduced the standard Android permissions interface presented at time of app installation, and was used as a control condition. The **Permissions + Purposes** interface condition was one in which annotations were added to permissions, specifying the purposes for each. This interface condition was designed in response to earlier work which found that smartphone users were most concerned with finding out such uses [47, 51]. The **Data Leaks** interface was inspired by the Privacy Leaks interface [6], and shows a list of third party destinations (identified by their host name) and the kinds of data they collect. The **DCI** and **PDCI** interfaces were as described earlier; participants could choose between matrix and Sankey representations for each.

Using only the information provided in each interface, participants were asked to choose one of the two apps, following a think-aloud process [39] while making their decision. After a decision was made, participants were asked to explain their decision, and indicate the level of confidence in their decision along a Likert scale. There was no time limit for each task.

Task Design

To create the tasks used in the study, 10 representative "seed" apps models were manually chosen from the 50 app models created using the network capture and analysis method previously described. These seed apps were used to generate fictional apps in 5 different popular app categories (productivity, health, personal finance, travel and utilities), to eliminate the possibility of accidental prior familiarity.

Using real app models of traffic data with fictional names ensured that the choices participants were faced with were as realistic as possible, and reflective of the true range of data tracking behaviours we observed across our entire data set of apps, whilst avoiding the problem of prior familiarity. Re-using the same 10 app models repeatedly across conditions (rather than collecting unique underlying data for each of the 25 app pairs) also enabled us to compare decisions by app-type, and to control for other potentially confounding factors that might arise from differences in the individual models. Finally, in order to focus participants' decisions on only the data collection aspects of the app choices, (rather than e.g. functionality or ratings), the apps in each pair were presented as having equivalent functionality and ratings.

Creating task sequences for participants required delicate scheduling of tasks. To avoid potential learning effects, we wished to ensure that each participant only saw each app pair at most once during the study. At the same time, we needed to ensure that each app pair appeared in the five different interface conditions described earlier. To achieve this, we adapted a between-groups design by dividing participants into five groups. For each of the 25 app pairs, one group made a judgement on that pair using the Permissions interface, while the other groups made a judgement on that pair using one of the other interfaces (Permissions + Purposes, Data Leaks, DCI, an PDCI). Through explicit scheduling, we ensured that each app pair was seen only once by one group, and each group saw five app pairs in each of the five interface conditions and five domain conditions. We wanted to also ensure that different interfaces were presented at each turn, so in order to overcome these issues and possible ordering biases, we permuted each app pair to ensure that no interface or domain condition appeared in adjacent order. Participants were randomly assigned to one of the five groups.

At the conclusion of the app choice tasks, participants were asked to reflect on all of the interface conditions, and to choose the interface condition that they found most helpful in making their decisions. Participants were asked to provide motivations and reasoning for their answers.

Participant Recruitment

Participants were recruited via social media, e-mail mailing lists at a large UK university, and paper flyers posted in public spaces across a small city in the UK. At time of sign-up, participants were asked to confirm eligibility by stating that they were at least 18 years old, owned and actively used a smartphone, and had installed at least 1 app on it themselves. This was done to ensure that participants had previously installed apps or would likely consider doing so in future. Prior to arriving at the lab, each participant was also asked to name up to 10 apps they used regularly. The traffic data from these apps were collected and used during the study to

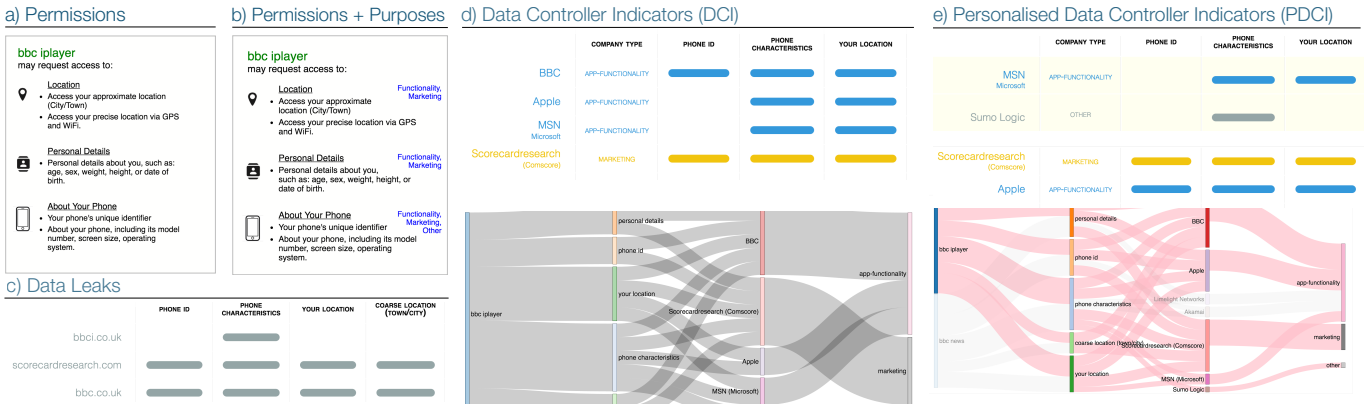

Figure 3. ***Experimental interface conditions*** **- (a) Permissions (Control), (b) Permissions + Purposes, (c) Data Leaks, and (d) DCI and (e) PDCI conditions each with tabular (top) and Sankey options (bottom).**

render a contextualised display in the PDCI interface. Participants were given a £10 gift voucher for completing the experiment.

Data Analysis

We examined the effect of different interfaces on participants choices using several measures. First, we computed the level of *inter-participant* agreement within each condition to see the extent to which participants in the same interface conditions made the same choice. Second, we tested *intra-participant* agreement between conditions, i.e., whether an individual participant was more likely to choose apps with different data collection characteristics between interface conditions. In order to test this, for each pair, we tested whether the app with the larger number of each feature (e.g., data controllers, purposes, and the data types) was selected significantly more often than the other, between interface conditions. Third, we calculated the average time taken to reach a decision under each interface condition. Finally, we performed a one-way ANOVA to test if there was a statistically significant difference in participants' reported levels of confidence in their app choices between interface conditions.

The think-aloud processes for all app choice rounds for all participants were recorded, transcribed and anonymised, along with app choice rationale. Then, an iterative thematic coding process was used to identify common themes for all aspects that were mentioned during the think-aloud process associated with each decision. After coding disjoint quarters of all rounds, 4 researchers convened and consolidated themes and then derived theme codes. Then, two researchers re-coded all rounds. Cohen's kappa k was run to determine if there was inter-coder agreement between the thematic tag assigned by each coder to each piece of text. A one-way ANOVA test was then performed on the thematic tags applied under each interface condition, to measure whether different interfaces resulted in participants engaging in more diverse decision making processes. For the exit survey, we recorded the total number of times each interface was regarded as the most helpful in making decisions.

For the concluding semi-structured interview questions, interview data were recorded, transcribed, and anonymised, and thematic analysis was applied to identify common themes in responses.

RESULTS

In total, 32 people participated. Data for the first two were removed, as the protocol was refined (we subsequently rebalanced the groups and their conditions). Thirty people completed the study (6 in each of the 5 in-between groups). Fifteen were male (avg. age 33), 13 were female (avg. age 36), and 2 decided not to disclose their gender (avg. age 28). Level of education varied from having completed high school (3), three-year college degree (11), four-year college degree (4), master degree (6), being a graduate student (5) to advanced graduate work or completed Ph.D (1). The study took an average of 68 minutes to complete (min=45; max=100).

Effect on Choice

When participants were asked to make a choice using the different interfaces they exhibited different levels of agree-

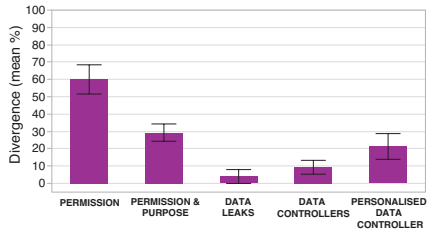

Figure 4. ***Choice Disagreement*** **- Inter-participant choice disagreement (divergence) per interface condition. Each error bar is constructed using 1 standard error from the mean.**

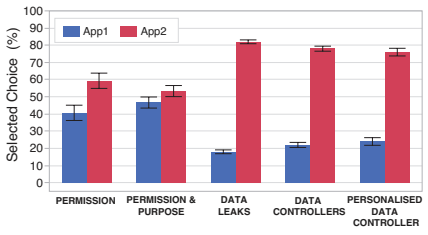

Figure 5. ***Choice of App w/ Fewer Controllers*** **- Counts of times an app with more data controllers was chosen (App 1) versus one with fewer. Error bar is 1 std. err from the mean.**

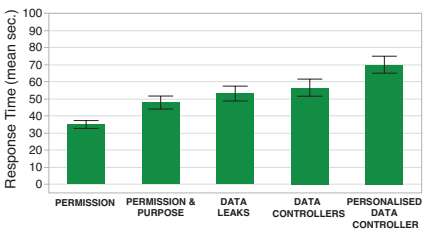

Figure 6. ***Response time*** **- Mean of time (seconds) taken to choose an app, grouped by interface conditions. Each error bar is constructed using 1 standard error from the mean.**

Code & Description		Example
DATA TYPES *(WHAT)*		
DT (182):	Data Types	*"this one is not taking my location"*
NT (125):	Num. of Types	*"it requests fewer pieces of information"*
PURPOSE OF USAGE *(WHY)*		
PU (231):	Purposes	*"if it's to market me more stuff, i don't want that."*
NP (55):	Num. of Purposes	*"it uses my data for fewer purposes."*
NE (39):	Need	*"it was for app functionality which i feel was helpful"*
DO (27):	Domain	*"Why would a notepad app access my data"*
DATA CONTROLLERS INFORMATION *(WHO)*		
NC (361):	Num. of Companies	*"they seem to be giving away my information to less people"*
TR (81):	Trust	*"i trust that developer (company) more"*
FM (102):	Familiarity	*"I've heard of them", "i don't know them"*
CA (17):	Company Attributes	*"it sounds like a good company"*
AC (45):	Already Collecting	*"facebook knows everything about my world anyway"*
UN (38):	Uniqueness	*"my data is already going to those places from other apps"*
OTHERS		
SA (138):	Same	*"they look exactly the same to me"*
NA (83):	Name	*"I don't like some of the names"*
RS (23):	Resignation	*"I'm just resigned to Google"*
RG (30):	Regulatory	*"they are operating in a grey area of the law."*

Table 1. Tags and tag definitions derived from thematic coding of app selection think-aloud process.

ment/disagreement about which app to choose (a one-way ANOVA test confirmed that there was a statistically significant difference in the levels of agreement/disagreement between interfaces ($F(4,24) = 13.26, p < 0.0001$)). Using the Permissions interface lead to the highest inter-personal disagreement (60%) (Figure 4), followed by the Permissions + Purposes interface (30%). When participants were given details about third-party information flows, their answers tended to converge on a common choice (the app with the fewest data controllers, or unique/new controllers in PDCI); disagreement was lowest in Data Leaks (4%), followed by DCI (9%) and PDCI (21%).

Did participants choose the app with fewer data controllers? Yes, for the Data Leaks, DCI and PDCI interface conditions, participants were significantly (one-way ANOVA ($F(4,145) = 20.59, p < 0.0001$)) more likely to choose the app that shared data with fewer organisations than in the permissions-based interface conditions. A Tukey post-hoc test revealed significant differences between the Permissions condition and the rest, including Permissions + Purposes ($p < 0.0142$), Data Leaks ($p < 0.0001$), DCI ($p < 0.0001$), and PDCI ($p < 0.0004$). The Permissions + Purposes interface also presented significant differences compared to Data Leaks ($p < 0.0001$), DCI ($p < 0.0001$) and PDCI ($p < 0.0001$). No pairwise significant differences were observed among Data Leaks, DCI and PDCI interface conditions. Figure 5 shows choices per condition, where "App 1" represents a choice of the app with *more* controllers while "App 2" designates the app with fewer was chosen.

We did not find any statistically significant effect of other company attributes (e.g. company type, number of purposes, or type of data requested) on app choice.

Effect on Time Taken

A one-way ANOVA test confirmed that there was a statistically significant difference in the time taken by participants to perform each app choice task (Figure 6) among the different interface conditions ($F(4,145) = 9.2141, p < 0.0001$). The average time taken to make a decision was the smallest in the Permissions condition ($\mu = 34.99s, \sigma = 12.5$), followed by Permissions + Purposes ($\mu = 47.81s, \sigma = 20.83$), Data Leaks ($\mu = 53.07s, \sigma = 23.78$) and DCI ($\mu = 56.5s, \sigma = 27.3$). The longest response time was reported using in the PDCI condition ($\mu = 69.9s, \sigma = 27.12$).

Effect on Confidence

A one-way ANOVA test confirmed that there was a statistically significant difference in participants' reported levels of confidence in their app choices between interface conditions($F(4,145) = 27.14$, $p < 0.0001$).

A Tukey post-hoc test revealed that participants' reported confidence was significantly lower in Permissions conditions ($\mu = 0.27$, $\sigma = 0.2$, $p < 0.0001$) compared to all the others ($p < 0.0001$). Participants' confidence increased when more detailed information was included as part of the interface (Figure 7). Merely including the purpose of use in Permissions interfaces (i.e. Permissions + Purposes) substantially raised confidence levels ($\mu = 0.62$, $\sigma = 0.2$, $p < 0.0001$), and levels subsequently increased when using Data Leaks ($\mu = 0.70$, $\sigma = 0.2$, $p < 0.0001$), DCI ($\mu = 0.78$, $\sigma = 0.18$, $p < 0.0001$) and PDCI interfaces ($\mu = 0.74$, $\sigma = 0.2$, $p < 0.0001$). There were no significant differences between the Data Leaks, DCI and PDCI interface conditions.

Effects on Decision-Making Thought Process

Our thematic analysis of participants' think-aloud process resulted in the set of codes described in Table 1. There was substantial agreement between the coders' judgements (Cohen's kappa $k = .776$; $p < .0001$). As shown in Figure 8, participant's

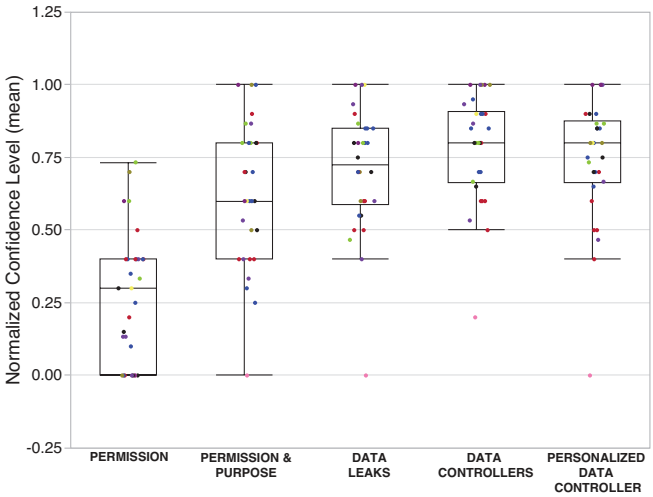

Figure 7. *Confidence per condition* - Box plot showing participants' normalised confidence level across different conditions. All participants' individual normalised means for each interface conditions are also shown.

think-aloud processes suggested that each interface resulted in participants using different thought processes and motivations for their choices (even if the same choice was made) (H:Strategy).

There was a significant difference in the number of factors participants reported considering between each interface condition, as determined by one-way ANOVA ($F(4,145) = 5.848$, $p < 0.0001$).

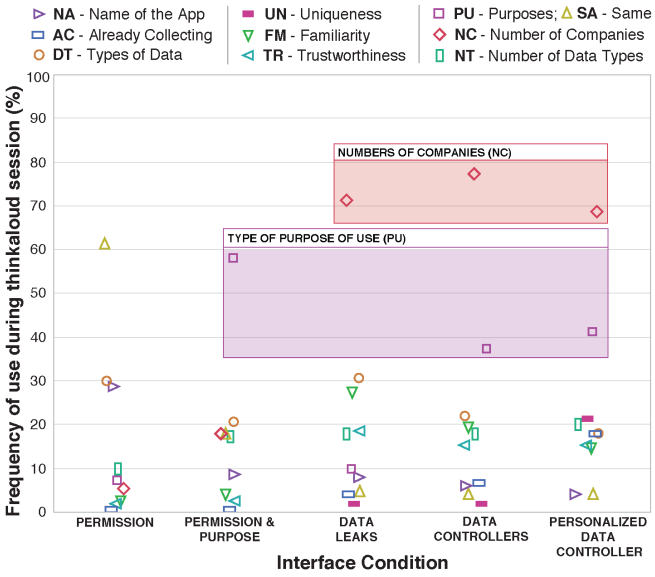

Figure 8. ***Decision-making factors*** **- Frequency (per task) that each factor was considered per interface condition. Each factor (corresponding to the tags in Table 3) is represented as a distinct symbol. In the Permission condition, most factors were seldom considered, except 3: Types of Data (DT), Name of App (NA), and the similarity between the choices (SA) were mentioned, whilst in DCI and PDCI conditions, these factors were rarely mentioned, instead replaced by a spread of others.**

In the Permissions condition, participants primarily considered the types of data being collected (**DT**). Most often, however, they mentioned not having sufficient information to make a decision (e.g. *"not a lot of difference between them"*) (P7), and often mentioned that they chose randomly because they looked the same (**SA**) or because they liked the app's name (**NA**), such as P12 who said: *"I'll go with [appname] because they're basically the same, but the name sounds more professional"*. In other conditions, participants very rarely mentioned using an app name as part of their decision making process.

The purpose of use (**PU**) was mentioned most often when participants were presented with the Permissions + Purposes interface. Participants paid close attention to the purposes and made judgements based on which type of purpose they believed more reasonable and necessary. For instance P2 preferred usage tracking to marketing, reporting that *"usage tracking could be a good thing"*, whereas *"marketing is too much in-your-face"*. Purpose was not always deemed useful, as P9 reported *"those words don't mean much to me"*.

When distinct controllers were presented (as in Data Leaks, DCI and PDCI conditions), the number of data controllers (**NC**) was the most prominent factor discussed by participants. For example P17 reported, *"it instinctively makes me feel less threatened, knowing that my information will be shared with fewer people [companies]"*. Similarly, P13 rejected an app because it *"seems to be sending your data to random places that i don't think are going to be beneficial to anyone"*.

While the number of controllers was the predominant factor that participants mentioned in these conditions, they also reported basing their decision on the familiarity (or lack thereof) (**FM**) and the trust (**TR**) they had towards particular companies. For example P9 discussed the importance of recognising a specific brand *"there's some comfort in Google being a brand that you recognise, as opposed to The [company name] Company. Like, who are they?"*. P6 reported that *"my choice would be done based on which [app] is [owned by] the large company. [...] I would go for a google one because it's more likely that they already have my information"*.

When faced with an unfamiliar entity, several participants described the importance of companies' *jurisdiction* and *location* (**CA**). P26 explained his reasoning for not wanting to choose the app which shared his data with a foreign organisation, because his data would be *"sent by one legislative zone to another"*. Four participants (P10, P25, P26, P29) discussed their concerns about having their data shared with companies based in specific countries. P25 reported *"[entity] oh those are both Chinese. [...] Every problem I've had recently is due to Chinese software vulnerabilities"*.

In the PDCI condition, participants frequently focused on whether installing an app would result in their data being exposed to new data controllers who hadn't previously accessed it (**UN** & **AC**). P4 reported *"contextualisation shows that there's only one unique data flow there"*. P1 discussed making a trade-off between sharing his data with more, but familiar, companies versus fewer but less familiar ones: *"Although [app name] doesn't have as many new companies that I need to consent to, I don't know who these are!"*.

Participants' decision making processes were also influenced by the specific domain (finance, health, note-taking, utilities and travel) of the app they were viewing (**DO**). There was a statistically significant difference in the number of factors participants considered between each type of domain, with regard to data type (*what*), purpose (*why*) and companies (*who*), as determined by one-way ANOVA ($F(4,145) = 4.0123, p < 0.0001$), ($F(4,145) = 7.8325, p < 0.0001$), ($F(4,145) = 14.3364, p < 0.0001$) respectively. The type of domain of each app pair presented different expectations and resulted in different types of considerations. For example, P12 disliked finance apps that either tracked her or sent data to third parties, arguing that *"it's a money management app, and I don't want my location tracked!"*. Others considered whether an app in a given domain could conceivably have legitimate need to collect certain data (**NE**), e.g. P15 reported *"for [...] a note taker, there is no reason to collect my location information"*.

Helpfulness, Preferences and Reflections on Interfaces

At the end of the study, we asked participants to state which interface they found most helpful when deciding which app to choose. The majority of participants (16) preferred the PDCI interface (Figure 9). The DCI and Permissions + Purposes interfaces were preferred by 6 participants each, making them the second most liked interface conditions.

Permissions The standard (and currently used) Permissions interface was regarded by many participants (26) as the least useful one, a *"standard form with no detail"* (P30), which

did not allow one to *"see anything about where [ones] data goes"*(P23). However, there were 2 participants, P5 and P11 that reported to prefer this interface, since they believed it *"provide[s] enough"*(P5), or that it is the *"clearest because it says exactly what it is being access[ed]"* (P11).

Permissions+Purposes 6 participants preferred the balance between details and clarity that the Permissions + Purposes interface provided. P14 reported *"I just want to know roughly what kind of data and why."*. This interface was also reported to be the *"easiest to read"* (P21) and more *"straightforward"* (P24) than the rest. However, the remainder of participants (the majority) stated that this interface *"felt uncontextualised"* (P9) and lacking in *"critical information"* (P1), since like the Permissions interface, it *"doesn't actually tell you [...] who it's sharing the information with, which is quite important"*(P28).

Data Leaks The Data Leaks interface was the least favoured of all the interfaces. Participants stated that the lack of company information was disconcerting; since as P13 described *"it makes me more inherently suspicious about the destination if you don't know why it's going there, and if you only have a website rather than company information that raises alarm bells"*. The only participant who reported finding it helpful (Figure 9) appeared to be recognising many of the host names directly (P24).

DCI Participants who preferred DCI (6) reported that they found it helpful to understand and reason about unfamiliar companies using the background information the interface provides. P10 was interested in company size, as he reasoned that large companies might *'pay more attention'* to laws on data protection, although he would be willing to forgive certain small start-up companies *'because they are trying their best not to misuse it'*. In addition to jurisdiction and size, corporate relationships were also seen as relevant. In some cases, being a subsidiary of a known brand offered reassurance to participants and indicated that the company was reputable; *"it was useful knowing it was [owned by] Twitter"*(P10). However, this detailed information could also be considered a disadvantage, since some participants (4) felt that the DCI presented too much information that required too much time to be interpreted. For instance, (P14 stated that *"I am not able to take the time to make much sense of that"*), while for (P26), the company descriptions were *"too broad"* to be useful.

PDCI The majority of participants (16) preferred PDCI. They reported valuing the ability to understand and differentiate between companies who had *already* accessed their data (via other applications) to the ones who had not. For example P1 stated *"I feel that the more new companies [there are] involved in collecting my data, the more I feel like I'm increasing my exposure to the world and to risks [...] increasing the danger of it being leaked"*. Similarly P10 remarked: *"it's kind of in for a penny, in for a pound with Google"*, implying that since Google already had his information, there would be no danger in sharing it with them again.

There were some participants (4) who had reservations about PDCI due to the complexity and difficulty of understanding the variety and the number of unique data flows, as well as fears that it may encourage complacency in the face of over-zealous trackers: (*"just because another app is doing it doesn't make it ok. [PDCI] is basically saying two wrongs make a right — that's completely silly!"* (P24)). This potential for complacency was recognised even by the participants who stated a clear preference for PDCI. P4 noted that while the PDCI interface was useful, there was a danger of *"a ratcheting up effect, where you think well because my phone is already doing this, then I'm ok with it happening [...] so it is a double-edged sword"*. However, participants did recognise that this was highly dependent on the company, for example P23 described *"if it's Facebook Ads [...] I don't want them to know about me through [this] app as well as the other ways they already know about me"*.

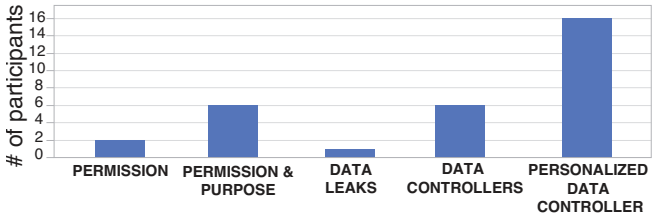

Figure 9. ***Participants' preferred interface*** **- Count of interface selected as preferred choice across all participants.**

DISCUSSION

Revisiting our hypotheses, we found significant support for *H:Strategy*. Somewhat unsurprisingly, interface conditions that provided more varied information resulted in more diverse decision-making strategies. However, there was also considerable contrast in the factors considered in each; the factors used most in the control conditions (such as app name and data types), were much less frequently considered in the information-rich conditions. Instead, other factors were considered first, such as the number of data destinations and purposes of use, along with details about the destination organisations, such as their reputation, trustworthiness, and country of origin. For PDCI, the additional factor of whether choosing an app sent data to organisations that already had their data, or widened their exposure served as an important focus. Such findings suggest that these factors were more valuable to app-choice decisions than those that currently take centre stage in platform permission interfaces.

In addition to variations in strategy, the results also revealed significant effects of interface condition on app choice, lending support to *H:AppChoice*. Such differences included both a higher rate of inter-participant agreement when DCI and PDCI interfaces were used, as well as a significant bias towards apps that shared data to fewer destinations, compared to when Permissions-based interfaces were used. Participants took significantly longer (*H:Time*) with both DCI and PDCI over the other conditions. On one hand, taking longer in these conditions could be seen as drawback due to the need for more time and effort; on the other hand, however, choices with these interfaces were made considering more factors, and were ultimately made with greater confidence (*H:Confidence*), particularly in the DCI condition. Finally, a majority of participants said they preferred the PDCI interface (*H:Preference*) overall, because it was perceived as most helpful in reasoning about how their degree of exposure would change as the result of their choice.

Individually, none of these measures is sufficient to draw a definitive evaluation of the interfaces examined in this study.

H:STRATEGY; H:ENTITY; H:BACKGROUND	
Thinkaloud themes	***Yes, Supported*** *- Participants considered more factors in the DCI/PDCI conditions than others. For control, participants focused on such features as : data shared (DT), app name (NA); in the DCI/PDCI factors varied more, included number of destinations (NC), familiarity with organisation (FM), trust (TR), and purpose (PU), as well as organisational background info such as its country of origin (CA). For PDCI, participants also considered the uniqueness of destinations (UN) and whether organisations already had their data (AC).*
H:APPCHOICE	
Inter-part agreement	***Yes, Supported*** *- Compared to Permissions (60%) and Permissions + Purposes (30%), inter-participant disagreement was lowest in Data Leaks (4%), DCI (9%), PDCI (21%).*
Data controllers	***Yes, Supported*** *- In DCI, PDCI and Data Leaks conditions, participants were significantly more likely to choose the app that sent data to fewer destinations.*
App type	***Yes, Supported*** *- Type of app (e.g. Productivity, Health, Finance) was shown to have a significant effect on number of factors considered.*
Other DC attributes	***No Support*** *- No significant effect was found of other data controller attributes (e.g. company type, size of company, age of company, or purpose) on choice.*
H:CONFIDENCE	
Confidence	***Yes, Supported*** *- DCI, PDCI and Data Leaks resulted in significantly higher confidence scores than either Permissions or Permissions+Purpose.*
H:TIME	
Time taken	***Yes, Supported*** *- Decisions took significantly* longer *in the DCI and PDCI conditions than the other conditions.*
H:PERSONALISE	
Post-task reflection	***Yes, Supported*** *- 15/30 of the participants preferred the PDCI, making it the most popular interface condition; however, some participants found other interfaces easier to use.*
H:PREFERENCE	
Post-task reflection	***Yes, Supported*** *- (22/30) chose either DCI or PDCI as their favourite interface, indicating that these interfaces were significantly more preferred than the other conditions.*

Figure 10. ***Summary of findings*** **- Summary of findings contextualised with hypotheses they were seen to support. Details of each finding are described under *Results*, and discussed in *Discussion*.**

For instance, the speed with which user makes a decision says nothing about the quality of the decision; likewise, while greater confidence may generally be a positive thing, confidence can be misplaced [12]. Evaluating the effectiveness of a privacy indicator therefore necessitates consideration of multiple dimensions — as well as some notion of what constitutes a 'good' privacy decision, which is itself highly contested. By providing these multiple measures we hope to enable such multi-faceted evaluation, and allow for a more nuanced deliberation on the various merits and trade-offs associated with each type of indicator.

Relation to Prior Work

As noted in the introduction, we sought to extend prior work on privacy indicators in several ways. First, we have extended previous work on exposing third-party libraries [6], with our (P)DCI approach, which represents third-party entities not simply by a hostname, but rather as companies, incorporated and embedded in a particular social, legal and economic context. Our study suggests that users did indeed find this more helpful (supporting *H:Entities* and *H:Background*). Furthermore, to our knowledge, our study is unique in directly comparing two other styles of indicator, namely Privacy Leaks [6] and purpose-based interfaces (such as [47, 51]), via a controlled study.

Second, our PDCI interface explores the notion of *differential risk* of installing an app: that is, the extent to which installing an app would expose an individual's data to entities who did not already have access to such data via other apps. Participants were able to use this to reason effectively about the differential impacts on their overall information exposure. This extends related work exploring similar approaches in the context of files shared to cloud backup services [25], and visual representations of a user's existing overall information exposure [3]. To our knowledge, our study is the first to explore differential risk with the typical range of data collected and shared with third parties through a smartphone app. Our results show that a large proportion of users may find this form of personalised privacy indicator useful in making decisions.

Implications: Better Decisions Through Transparency?

Our findings suggest providing transparency about tracking behaviours is important for several reasons. First, there was substantial variability in people's sensitivity to being tracked. A few disliked the idea of being tracked altogether, and, lacking the ability to control or know about what was going on, had resorted to the strategy of using apps minimally and avoiding installing new apps altogether. By providing greater transparency, these individuals could purposely find and choose apps that disclosed minimally to entities they trusted, and be more confident in their use.

A majority of people, however, were pragmatic. One of the most important findings of our work is the variation in the ways such pragmatists chose apps in the information-rich conditions. While certain heuristics (such as number of data controllers) were often important, there was no universal tactic for anticipating what people would consider a "better" choice; it was clear that different factors mattered for different people. For instance, while some expressed their love for and trust in Google; others felt resigned to Google having "everything". Still others were resolutely against Google gaining any more information about them. Even for lesser-known controllers, participants attitudes ranged from thinking that they were "probably harmless" to being deeply suspicious. These highly idiosyncratic forms of reasoning suggest that interfaces such as (P)DCI allow non-experts to effectively make choices that reflect their individual preferences, biases and world-views. They achieve this by re-contextualising the decision, representing it as a choice between ecosystems of data controllers as social entities, rather than as a choice between one app or another. Ultimately, these kinds of transparency mechanisms might encourage users to choose apps with more reasonable data collection and sharing practices, thus exerting pressure on app developers and third party services to reform standards in the app market.

Study Findings and Modelling Accuracy

With regard potential limitations, it might be assumed that the study results depended crucially on the accuracy of the techniques used to model app data sharing behaviours. However, we believe this to not be the case; the apps and their descriptions were entirely fictional, and although the models were derived from real apps, they were randomly assigned from among the archetypes selected from a much larger set. Therefore, we believe the results could be attributed more to the kinds of information and means

of conveyance, rather than the specifics of what was conveyed, namely data in the captured models.

This is not to say that accuracy would not matter in a real deployment; for privacy indicators to carry any weight, people would need to be able to trust their accuracy and completeness. Even when participants were told in the study that the apps were fictional, some participants were intrigued as to how the data represented in the DCI was detected, and keen to confirm their veracity. If such trust were compromised, people would likely ignore the indicators outright, making them useless.

This leads to the obvious question of how such app behaviours should be measured, by whom, and how they should be verified. While app developers themselves generally know what their apps are doing in their code, they may have a perverse incentive to not report them for fear of discouraging potential users (creating a problematic information asymmetry [1]). Companies that run app marketplaces, such as Google and Apple, are another obvious possibility, as the anchors of trust that arbitrate which apps can enter their ecosystems. However, leaving such analysis to the operators of these platforms raises its own problems, as they themselves are purveyors of apps which may collect large amounts of personal data. In some cases, these companies are beneficiaries (or even owners) of the very third party advertising networks that would be the subject of their audit activity, creating a conflict of interest [24].

Moreover, if data collection and dissemination behaviours become widely known, it is possible, even likely, that third party entities will start to resort to covert methods that are harder to detect. For example, instead of simply encrypting data via HTTPS, apps could further encrypt the payload, or use techniques such as certificate pinning [20]. Such technical challenges mean that alternative approaches, such as static analysis or runtime instrumentation, may ultimately need to be used. However, even these "deeper" approaches are at risk of being ineffective (or illegal) if developers start to employ DRM to hide their information collecting activities, since reverse engineering DRM can fall foul of copyright law in many countries [22]. Thus, the ultimately sustainability of approaches to achieving transparency may depend on either voluntary compliance by platform providers, or an external force such as a regulatory policy change requiring such disclosure.

LIMITATIONS

There are several potential limitations of our primary findings. The first pertains to the ecological validity of our results. Since our methodology used a lab study requiring role-play and decisions about fictional apps, there are many potential ways that a real field deployment of (P)DCI indicators might yield different results. In designing the lab study, however, we took steps to make the study more realistic, and make it easier for people to treat the role-playing tasks as genuine. First, we chose a common role-play task that would be as familiar as possible to our participant population, who we selected to ensure all had experience installing apps. Second, to ensure that the fictional apps were realistic, we used real models to drive all interface conditions, which we fictionalised only by changing the name of the app and company. Thus, the interfaces were as close to real as possible, while avoiding the use of real apps which would have presented the challenge of prior familiarity. Finally, we attempted to recreate platform interfaces for control conditions as close to the originals as possible, to the extent that this was feasible.

A second limitation of the study pertains to its scope: we focused on a single, common decision-making context. As discussed already, there are many other privacy-related decisions people make routinely regarding their apps, and whether indicators like PDCI will help in these other contexts will likely vary upon the particular situation. Moreover, even for app choice, our tasks were simplified to involve choosing only between two apps, whilst in real life people may be selecting among dozens or even hundreds.

While using real app models was beneficial for ecological validity, it simultaneously meant we were limited in the extent to which we could examine the decisions that were made. For example, we were not able to tweak the models in each round to measure specific thresholds for decisions being made, such as by incrementally varying the number of data collection hosts or company reputations. One reason that we felt this was out of the scope of what we wanted to look at initially is that it was likely that these specific thresholds would likely vary significantly, from one person to another, or even for a single person between situations. We intend to examine such detailed preferences and variation in future work.

CONCLUSION

This study revealed that making apps' data collection and sharing behaviours transparent to users was beneficial in several ways. Participants were able to consider to what extent using an app would increase their exposure to third parties, and weigh a much larger variety of factors to determine whether they were comfortable with the parties gaining access to their data. The result was that people made different choices than with a traditional permissions interface, and had higher confidence in their choices. However, the continued ability to be able to detect and discern such behaviours of apps is under threat, as apps and platforms are likely to hide the particular information pathways we used to collect data from scrutiny in the future. We believe keeping pathways open and scrutable, and making information collection behaviours accountable, will lead to greater trust in apps in the future.

ACKNOWLEDGEMENTS

Max Van Kleek, Reuben Binns, Jun Zhao, Nigel Shadbolt were supported under *SOCIAM: The Theory and Practice of Social Machines.* The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/2 and comprises the University of Oxford, the University of Southampton, and the University of Edinburgh.

Ilaria Liccardi and Daniel J. Weitzner were supported by the *William and Flora Hewlett Foundation* and the *NSF 1639994: Transparency Bridges* grant.

REFERENCES

1. George Akerlof. 1995. The market for "lemons": Quality uncertainty and the market mechanism. In *Essential Readings in Economics*. Springer, 175–188.

2. Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your location has been shared 5,398 times!: A field study on mobile app privacy

nudging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 787–796.

3. Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. Usable transparency with the data track: a tool for visualizing data disclosures. In *Proceedings of the Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1803–1808.
4. Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Octeau, and Patrick McDaniel. 2014. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. *ACM SIGPLAN Notices* 49, 6 (2014), 259–269.
5. Rebecca Balebako and Lorrie Cranor. 2014. Improving app privacy: Nudging app developers to protect user privacy. *IEEE Security & Privacy* 12, 4 (2014), 55–58.
6. Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. 2013. Little brothers watching you: Raising awareness of data leaks on smartphones. In *Proceedings of the Symposium on Usable Privacy and Security*. ACM, 12.
7. Rebecca Balebako, Pedro G Leon, Hazim Almuhimedi, Patrick Gage Kelley, Jonathan Mugan, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2011. Nudging users towards privacy on mobile devices. In *Proceedings of Workshop on Persuasion, Nudge, Influence and Coercion*.
8. Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. 2015. The Impact of Timing on the Salience of Smartphone App Privacy Notices. In *Proceedings of the ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 63–74.
9. Leonid Batyuk, Markus Herpich, Seyit Ahmet Camtepe, Karsten Raddatz, Aubrey-Derrick Schmidt, and Sahin Albayrak. 2011. Using static analysis for automatic assessment and mitigation of unwanted and malicious activities within Android applications. In *Malicious and Unwanted Software (MALWARE), 2011 6th International Conference on*. IEEE, 66–72.
10. Michael Benisch, Patrick Gage Kelley, Norman Sadeh, and Lorrie Faith Cranor. 2011. Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing* 15, 7 (2011), 679–694.
11. Jan Lauren Boyles, Aaron Smith, and Mary Madden. 2012. Privacy and Data Management on Mobile Devices. *Pew's Report: Mobile Identity* (September 5th 2012), 1–19.
12. Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced confidences privacy and the control paradox. *Social Psychological and Personality Science* 4, 3 (2013), 340–347.
13. Michael Brown, Tim Coughlan, Glyn Lawson, Murray Goulden, Robert J Houghton, and Richard Mortier. 2013. Exploring Interpretations of Data from the Internet of Things in the Home. *Interacting with Computers* 25, 3 (2013), 204–217.
14. A Cortesi and M Hils. 2013. "mitmproxy: a man-in-the-middle proxy. (2013).
15. Lorrie Faith Cranor. 2006. What do they indicate?: evaluating security and privacy indicators. *Interactions* 13, 3 (2006), 45–47.
16. Manuel Egele, Christopher Kruegel, Engin Kirda, and Giovanni Vigna. 2011. PiOS: Detecting Privacy Leaks in iOS Applications.. In *NDSS*. 177–183.
17. Serge Egelman, Adrienne Porter Felt, and David Wagner. 2013. Choice architecture and smartphone privacy: There'sa price for that. In *The economics of information security and privacy*. Springer, 211–236.
18. William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. 2014. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)* 32, 2 (2014), 5.
19. Steven Englehardt and Arvind Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of ACM Computer and Communications Security 2016*.
20. Chris Evans, Chris Palmer, and Ryan Sleevi. 2015. *Public key pinning extension for HTTP*. RFC 7469. RFC Editor. `https://tools.ietf.org/html/rfc7469`
21. Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of Symposium on Usable Privacy and Security*. ACM, 3.
22. Electronic Frontier Foundation. 2016. Coders' Rights Project Reverse Engineering FAQ. (2016). `https://www.eff.org/issues/coders/reverse-engineering-faq`
23. Huiqing Fu, Yulong Yang, Nileema Shingte, Janne Lindqvist, and Marco Gruteser. 2014. A field study of run-time location access disclosures on android smartphones. *Proceedings of USEC* 14 (2014).
24. Sanford J Grossman and Oliver D Hart. 1983. An analysis of the principal-agent problem. *Econometrica: Journal of the Econometric Society* (1983), 7–45.
25. Hamza Harkous, Rameez Rahman, and Karl Aberer. 2016. Data-Driven Privacy Indicators. In *Symposium on Usable Privacy and Security*. USENIX Association.
26. Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Proceedings of Symposium On Usable Privacy and Security*. 39–52.
27. Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. 2012. A conundrum of permissions: installing applications on an android smartphone. In *International Conference on Financial Cryptography and Data Security*. Springer, 68–79.

28. Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3393–3402.
29. Jennifer King. 2012. How Come I'm Allowing Strangers to Go Through My Phone? Smartphones and Privacy Expectations. *Smartphones and Privacy Expectations* (2012).
30. Pedro Giovanni Leon, Ashwini Rao, Florian Schaub, Abigail Marsh, Lorrie Faith Cranor, and Norman Sadeh. 2015. Privacy and Behavioral Advertising: Towards Meeting Users' Preferences. In *Proceedings of the Symposium on Usable Privacy and Security*.
31. Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. 2013. What matters to users?: factors that affect users' willingness to share information with online advertisers. In *Proceedings of Symposium on Usable Privacy and Security*. ACM, 1–7.
32. Ilaria Liccardi, Joseph Pato, and Daniel J. Weitzner. 2014a. Improving Mobile App Selection through Transparency and Better Permission Analysis. *Journal of Privacy and Confidentiality: Vol. 5: Iss. 2, Article 1.* (2014), 1–55.
33. Ilaria Liccardi, Joseph Pato, Daniel J. Weitzner, Hal Abelson, and David De Roure. 2014b. No Technical Understanding Required: Helping Users Make Informed Choices About Access to Their Personal Data. In *Proceedings of International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 140–150.
34. Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of Conference on Ubiquitous Computing*. ACM, 501–510.
35. Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I Hong. 2014. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Symposium On Usable Privacy and Security*. 199–212.
36. Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun Aerin Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. 2016. Follow My Recommendations: A Personalized Assistant for Mobile App Permissions. In *Proceedings of the Symposium on Usable Privacy and Security*.
37. Kirsten Martin and Katie Shilton. 2016. Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society* 32, 3 (2016), 200–216.
38. Akiva A Miller. 2014. What Do We Worry About When We Worry About Price Discrimination? The Law and Ethics of Using Personal Information for Pricing. *Journal of Technology Law & Policy* 19 (2014), 41.
39. Janni Nielsen, Torkil Clemmensen, and Carsten Yssing. 2002. Getting access to what goes on in people's heads?: reflections on the think-aloud technique. In *Proceedings of the Nordic conference on Human-computer interaction*. ACM, 101–110.
40. Kenneth Olmstead and Michelle Atkinson. 2015. App Permissions in the Google Play Store: Chapter 1: The Majority of Smartphone Owners Download Apps. *Pew's Report: Numbers, fact and Trends Shaping the world* (November 19th 2015), 1–19.
41. TNS Opinion. 2015. Special Eurobarometer 431, Data protection. (2015). `http://ec.europa.eu/public_opinion/archives/eb_special_439_420_en.htm#431`
42. Paul Pearce, Adrienne Porter Felt, Gabriel Nunez, and David Wagner. 2012. Addroid: Privilege separation for applications and advertisers in android. In *Proceedings of the Symposium on Information, Computer and Communications Security*. ACM, 71–72.
43. Jeremy Pounder. 2015. For what it's worth - the future of personalised pricing. (2015).
44. Lingzhi Qiu, Zixiong Zhang, Ziyi Shen, and Guozi Sun. 2015. AppTrace: Dynamic trace on Android devices. In *2015 IEEE International Conference on Communications*. IEEE, 7145–7150.
45. Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online.
46. Jingjing Ren, Ashwin Rao, Martina Lindorfer, Arnaud Legout, and David Choffnes. 2016. Demo: ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services Companion (MobiSys '16 Companion)*. 117–117.
47. Fuming Shih, Ilaria Liccardi, and Daniel Weitzner. 2015. Privacy tipping points in smartphones privacy preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 807–816.
48. Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. 2014. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2347–2356.
49. Statistica. 2016. Number of smartphone users in the United States from 2010 to 2019 (in millions). (2016). `http://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/`
50. Peter Tolmie, Andy Crabtree, Tom Rodden, James Colley, and Ewa Luger. 2016. "This has to be the cats": Personal Data Legibility in Networked Sensing Systems. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 491–502.
51. Na Wang, Bo Zhang, Bin Liu, and Hongxia Jin. 2015. Investigating Effects of Control and Ads Awareness on Android Users' Privacy Behaviors and Perceptions. In *Proceedings of Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 373–382.

52. Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2367–2376.

53. Jason Watson, Heather Richter Lipford, and Andrew Besmer. 2015. Mapping user preference to privacy default settings. *ACM Transactions on Computer-Human Interaction* 22, 6 (2015), 32.

54. Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney. 2015. Who knows what about me? A survey of behind the scenes personal data sharing to third parties by mobile apps. *Proceeding of Technology Science* (2015).

55. Frederik J Zuiderveen Borgesius. 2015. Improving privacy protection in the area of behavioural targeting. *SSRN 2654213* (2015).