Lehtiniemi, Tuukka; Kortesniemi, Yki

Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach

Published in:
BIG DATA & SOCIETY

DOI:
10.1177/2053951717721935

Published: 01/01/2017

Document Version
Publisher's PDF, also known as Version of record

Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2017: 1–11

DOI: 10.1177/2053951717721935
journals.sagepub.com/home/bds
SAGE

Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach

Tuukka Lehtiniemi[1] and Yki Kortesniemi[2]

Abstract

In privacy self-management, people are expected to perform cost–benefit analysis on the use of their personal data, and only consent when their subjective benefits outweigh the costs. However, the ubiquitous collection of personal data and Big Data analytics present increasing challenges to successful privacy management. A number of services and research initiatives have proposed similar solutions to provide people with more control over their data by consolidating consent decisions under a single interface. We have named this the '*consent intermediary*' approach.

In this paper, we first identify the eight obstacles to privacy self-management which make cost–benefit analysis conceptually and practically challenging. We then analyse to which extent consent intermediaries can help overcome the obstacles. We argue that simply bringing consent decisions under one interface offers limited help, but that the potential of this approach lies in leveraging the intermediary position to provide aides for privacy management. We find that with suitable tools, some of the more practical obstacles indeed can become solvable, while others remain fundamentally insuperable within the individuated privacy self-management model. Attention should also be paid to how the consent intermediaries may take advantage of the power vested in the intermediary positions between users and other services.

Keywords

Personal data, informed consent, privacy self-management, privacy, cost–benefit analysis, consent intermediary

Introduction

The production of personal data – any information relating to an identified or identifiable natural person (European Union, 2016) – seems to be ever increasing as activities performed with information technology have become daily routines, and companies use Big Data analytics to produce potentially detailed pictures of us. This extensive use of personal data can benefit individuals themselves, as personalisation can make services more valuable to use, and business models based on profiling often make services available free of charge. But the associated cost is the impact on privacy as people reveal more information about themselves to service providers.

EU legislation has long held privacy as a fundamental right of the individual (Wachter, 2017) and places strict limits on the processing of personal data. The new General Data Protection Regulation (GDPR) (European Union, 2016) states that personal data may only be processed based on one of the following six grounds: it is required by a legal obligation, it is carried out to protect a vital interest of the individual, it is carried out for the public interest, it falls within a legitimate interest of the data controller, it is necessary for the performance of a contract, or it is based on the consent of the individual. The informed consent approach, which is also used in many other jurisdictions, allows people the freedom to agree to many types of data processing. However, with a contract, processing is limited to data which is strictly necessary for its fulfilment, and for the

[1]Department of Computer Science, Aalto University, Finland; Faculty of Social Sciences, University of Turku, Finland
[2]Department of Computer Science, Aalto University, Finland

Corresponding author:
Tuukka Lehtiniemi, Aalto University, PO Box 15600, Espoo 00250, Finland.
Email: tuukka.lehtiniemi@iki.fi

other bases, the individual can either do nothing or can at most object to some of it.

In this article, we focus exclusively on consent-based processing as it places the greatest demands on the individual's ability to make informed decisions. People are expected to manage their privacy by weighing the subjective costs and benefits of data collection in each case (Solove, 2013). In practice, however, many are neither well informed on the uses of their personal data nor feel in control of it (European Commission, 2015; Turow et al., 2015). A fundamental dilemma underlies the concept of informed consent: meaningful cost–benefit analysis on personal data is anything but straightforward in the context of Big Data analytics, data aggregation, and opaque data flows. But for the moment, we will live with the model of informed consent, as in many jurisdictions it is codified in legislation. This has sparked an ongoing discussion about how to make the model work better. For example, Custers (2016) discusses expiry dates for consents and calls for further discussion on the issue of consents in the Big Data era, a call to which we respond with the present article.

Within the last few years, a number of initiatives to give people better control over their personal data have started to appear. Proponents of personal information management systems (PIMS) (Abiteboul et al., 2015; European Commission, 2016) recognise the current inability of people to meaningfully control the uses of their data, and seek to redress the situation with personal data stores and features for managing data use permissions. We consider these emerging services *consent intermediaries* (CIs). With CIs, people themselves still manage their own privacy, but the intermediary consolidates all management to a single place. In this article, we investigate the concept of CIs and ask two questions: (1) to what extent can CIs help people in making informed privacy decisions and (2) is it even possible to overcome all of the obstacles?

The rest of the paper is organised as follows: we first review the privacy self-management model and identify eight obstacles which currently stand in the way of informed privacy decisions. We then proceed to describe the CI approach and analyse its potential to tackle these obstacles. We find that CIs form a platform for building tools to help with the more practical obstacles, but obstacles arising from the privacy self-management model's individuated nature are not as easy to solve without relaxing the model's individuated assumptions. Finally, we conclude by discussing the implications of the CI approach.

Privacy self-management

The right to informational privacy is essentially a decision right. Zuboff (2015), for example, conceptualises it as the ability to choose one's position along the spectrum between secrecy and transparency. Altman (1975) refers to the same phenomenon as boundary regulation of privacy and publicness. Solove (2013) refers to the current approach of privacy regulation as *privacy self-management*; people have the right to *notice* of the upcoming collection and use of personal data and have the *choice* whether or not to consent to such processing. Armed with these rights, people are expected to make privacy decisions based on cost–benefit evaluations and to disclose data only when the benefits outweigh the costs.

Privacy self-management has to take into account the highly divergent preferences people have on the desirable position along the secrecy-transparency spectrum. Westin's well-known classification identifies privacy *fundamentalists*, who have high privacy concerns, *pragmatists*, who have some concerns but favour individual choice, and the *unconcerned*, who have low concerns and tend to trust data collectors (Hoofnagle and Urban, 2014). Further, individuals' preferences on privacy can change over time, and are also highly context-dependent (Acquisti et al., 2015; Coll, 2014; Hoofnagle and Urban, 2014).

Privacy self-management relies on individuals being informed and making decisions based on subjective analysis of this information, and it therefore places a lot of faith in their rational capabilities. But in practice, decision-making is only partially the result of rational cost–benefit analysis. Decisions are also affected by social norms, emotions and heuristics (Acquisti et al., 2015), and people are only boundedly rational (Gigerenzer and Selten, 2001), due to limitations in information, cognitive capabilities, and available time. Therefore, individuals are often not that well-informed when consenting; they do not always read privacy policies (Custers, 2016) and can operate under misinformed assumptions about these policies' purpose and contents (Turow et al., 2015). In a recent Eurobarometer, only 18% of respondents reported reading privacy policies fully and 49% partially, length and complexity being typical reasons for not reading them (European Commission, 2015). In fact, many habitually accept consent dialogues without even glancing the provided information (Böhme and Köpsell, 2010).

Unsurprisingly, people do not feel in control of personal data but nevertheless see no alternatives to disclosing data in order to gain access to services (European Commission, 2015). The feelings of powerlessness to contest the data collection practices speak of the same issue (Andrejevic, 2014). To overcome the experienced lack of control, people employ *implicit* control mechanisms to regulate the quantity and quality of data, including maintaining multiple or pseudonymous profiles, providing incorrect information,

and refraining from providing data whenever feasible (Snell et al., 2012). Overall, however, people's actions seem to demonstrate the privacy paradox: despite indicating concerns about privacy, they part with intricate details about their private life – in other words, behavioural intentions towards privacy are not reflected in actual behaviour (Norberg et al., 2007). One explanation is that broad attitudes to privacy may measure different things than contextual decisions (Acquisti et al., 2015). Another could be that when people do consider the costs and benefits of the options provided, if the cost of achieving better privacy is high – for example the inability to use a social networking site, or a significant effort to configure the privacy settings – people do not always see the benefits of better privacy as worth the cost.

We can conclude that even if the privacy self-management model expects people to behave rationally, this is not always the case. It is therefore worth exploring whether the proposed new ways of consenting can help people make better privacy decisions.

Obstacles to privacy self-management

In this section, we distil findings from literature review into eight obstacles which summarise the challenges of privacy self-management. Our aim is to categorise and reformulate these findings so that they can be used to evaluate the effectiveness of attempts to improve privacy self-management. The obstacles are summarised in Table 1.

Table 1. Obstacles to privacy self-management.

Timing and duration	Estimating harms is difficult due to timing of decisions and the typically unlimited duration of the consent (Custers, 2016; Solove, 2013).
Non-negotiability	The terms are not negotiable enough (Custers, 2016).
Scale	Privacy self-management does not scale well enough (McDonald and Cranor, 2008; Solove, 2013).
Aggregation	Data is aggregated and analysed to produce new data, leading to implicit disclosure of latent data (Mai, 2016; Solove, 2013).
Downstream uses	Data flows to parties and purposes not foreseen at the time of consenting (Anthes, 2015; Crain, 2016; Turow et al., 2015).
Cognitive demands	The cognitive limitations of all human decision making hamper cost–benefit analysis (Solove, 2013).
Social norms	Pressure to conform can strongly affect the decisions people make (Acquisti et al., 2015; Andrejevic, 2014; Zuboff, 2015).
Social data	Privacy decisions are framed as individual choices, but the data and the decisions also affect others (Lampinen et al., 2011; Schneier, 2010; Taylor et al., 2017).

Timing and duration. A challenge with privacy is that it is an outcome of long-term information management, but the practical implementations of privacy self-management do not currently support this (Solove, 2013). The point of decision occurs when the collection of personal data is started, and individuals are then expected to assess all future harms and benefits. Decisions on disclosing personal data are also made in isolation from other similar decisions, and often they are made with the aim of gaining immediate benefits. And while immediate harms may be insignificant, long-term harms can develop gradually over time. Having to make the decision before the outcomes arise is arguably a feature of most human decision-making, but with personal data, the timing poses particular difficulties due to the inherent dynamics arising from the advancement of data analysis technologies (Custers, 2016). As harms and benefits may arise by mechanisms which are not discernible, or do not yet even exist, the consequences of a disclosure decision are a moving target. Yet a consent, once given, is typically in effect indefinitely.

Non-negotiability

The current implementation of notice and choice is usually based on terms dictated by the service provider (Custers, 2016), and users have to accept these terms in full to use the service. The other option, obviously, is not to use the service. This Hobson's choice does not match the preferences of those who are willing to agree to some subset of the terms in exchange for some subset of the service. Also, once the choice is made, the terms of personal data use are largely fixed. For example, the privacy settings within a service often only affect the visibility of personal data to third parties rather than, for example, what data gets collected. However, post-consent negotiations of sorts can arise when an organisation attempts to impose new terms which a large portion of the users find unacceptable. This is evidenced by the stir which Spotify's new privacy policy caused and the consequent changes made by the company (Kastrenakes, 2015). In addition, data protection regulation in the EU, for example, provides the possibility of withdrawing consent at will. But in practice, reconsidering a decision is impractical and potentially ineffective. When providing consent is an all-or-nothing decision, withdrawing consent involves ceasing the use of the service altogether. It also involves removing data

in the service provider's databases, but having data deleted has turned out to be a complex issue (Custers, 2016). Interestingly, the upcoming GDPR (European Union, 2016) addresses the current situation by stating that the availability of a service cannot be contingent on the individual consenting to data processing which is not essential to the service.

Scale

A practical obstacle to decision-making is that privacy self-management as currently implemented does not scale too well. Making people better informed in their decisions cannot be achieved simply by convincing people to read privacy policies better (Solove, 2013), because there is just too much information to study, and there are too many decisions to make. An estimated 80–300 hours are needed to familiarise oneself with just the privacy policies of the websites an individual visits in a year (McDonald and Cranor, 2008), and including other data-collecting entities only increases the time required. Also, as we will discuss below, some of the obstacles are due to people not being fully aware of the complex consequences of the decisions they make – and the more people are made aware of the consequences, the more problematic the scaling problem can become.

Aggregation

Data-collecting entities often aggregate personal data across individuals and contexts, which can lead to revelation of new data through data analysis. We contrast this *latent data* to openly expressed and exhaust data (Kitchin, 2014). Openly expressed data is consciously provided by individuals about themselves, for example filled in a form, and exhaust data is produced by observing activities, for example, clickstreams on a website. Latent data, however, is fundamentally different; it is produced from other data by using inference techniques and is therefore implicitly shared alongside other data. Yet an explicit consent is never provided for latent data (Mai, 2016). Inference can produce seemingly unconnected results by treating the input data as proxy data for the unavailable information. Then, for example, demographic data can be deduced from location history alone (Bellovin et al., 2013). Aggregation also happens across not just numerous sources of data but also across individuals. Therefore, the costs and benefits of my disclosure decisions are affected by the decisions of others. So even if each disclosure decision were well-considered in isolation, the aggregation of data can lead to the overall effect being undesired. The production of latent data also hinders the effectiveness of refraining from disclosing data, as it may still be deduced from other data (Custers, 2016).

Downstream uses

Unexpected movements of personal data to new parties are to a large extent opaque to the individuals, complicating meaningful decision-making. We refer to these movements as downstream uses of data. There are multiple reasons for these movements. In downstream data markets, data brokers sell personal data compiled from public records and nonpublic sources, often without the knowledge of the individuals involved, even though they may have consented to such uses of data by the primary data collectors (Anthes, 2015; Crain, 2016; Turow et al., 2015). Changes in business models of data-collecting companies may result in new uses of the data contradicting the individual's' expectations from the time of decision, an example being direct-to-consumer personal genome testing and the subsequent medical research use of the collected data (Alba, 2015; Seife, 2013). Another reason for downstream uses is malicious actions of third parties. Well-known examples include publication of data hacked (Zetter, 2013) or otherwise collected (Zimmer, 2016) from dating services, resulting in personal data about customers being put to unforeseen uses. Personal data collected by private companies can also end up being aggregated in governmental databases, and superficially innocuous pieces of personal data may end up being highly consequential in practice. The upcoming GDPR, again, places some limitations on downstream uses of data, stipulating that all processing must have a legal basis.

Cognitive demands

People's ability to make informed and rational choices about personal data is not on par with requirements of privacy self-management, and people can end up making bad decisions with respect to disclosing personal data, regardless of the information and tools they have in use. The cognitive limitations hampering privacy self-management have been summarised by Solove (2013) as follows. To begin with, people are not very well informed about the decisions they make because they do not read privacy policies. If they do read them, they have difficulties understanding them. If they do understand them, they lack the necessary knowledge to make a truly informed choice. And even if they are well-informed, their decision-making capability is limited by difficulties which generally riddle human decision-making.

Social norms

As observed above, the decision not to disclose personal data often means non-participation in activities which include collecting data. As many online services are regarded an integral part of modern life (Andrejevic, 2014; Zuboff, 2015), non-participation

may simply be infeasible regardless of privacy preferences or subjective concerns over data disclosure. Thus, decision-making on personal data is subject to social norms (Acquisti et al., 2015) which regulate individual decisions. Another way of saying this is that private cost–benefit decisions to disclose data are embedded in a network of social relations, and looking at them from an individuated, under-socialised point of view is misleading (Granovetter, 1985). These norms are further reinforced by each individual decision; the more people conform, the harder it becomes to deviate from the norm regardless of the individual judgement of costs and benefits. Norms may also be at odds with attempts to implicitly control the quality and quantity of data once consent has been provided.

Social data

Privacy self-management frames the decision-making on personal data as an individual choice based on private cost–benefit analysis, despite personal data often also conveying information about others. Schneier (2010) uses the term *incidental data* to denote data which other people's activities leak about you. Any data about my interactions or relationship with you is also data about you. Health or genome data may implicate relatives in the case of hereditary diseases, shared photos can convey information about others, and consumption data may by nature concern a household. Decision to share location data may help predict the future locations of others (Bellovin et al., 2013), and the combined effect of two people sharing location data may reveal details about their relationship. In particular, latent data produced by Big Data analytics may by nature concern a group rather than individuals (Taylor et al., 2017). Privacy can, by various mechanisms, be affected by the choices others make (Lampinen et al., 2011) and the outcomes of data-sharing decisions are, therefore, not only private.

To summarise the obstacles, privacy decisions are made in a situation described by considerable information asymmetry; non-experts know little about collected personal data, what is done with the data, or the business operations of the data industry (Zuboff, 2015). Altogether, the obstacles affect privacy self-management by first making it hard to appraise the situation and then by diminishing the possibilities of actually making preferred decisions.

Next, we proceed to describe the CI approach and analyse its potential to tackle these obstacles.

Consent intermediaries

The last few years have seen an emergence of initiatives and services whose aim is to provide people with better control over the collection and sharing of their personal data. A report by the European Commission (2016) on PIMS included commercial service developers such as the personal cloud server Cozy Cloud (2017) and the personal information control services digi.me (2017) and Meeco (2017), as well as research-originated initiatives such as the networked personal data indexing device Databox (Chaudhry et al., 2015), personal data stores Hub of All Things (Hub of All Things, 2017) and OpenPDS (de Montjoye et al., 2014), and the personal data management model MyData (Poikola et al., 2015).

All of these services aim to provide more control to personal data and allow people to share their data with third parties, but two different means to achieve this can be identified: *storage spaces* accumulate personal data from various sources, whereas *permission-management services* only keep track of where data is stored. While their practical implementations and stages of maturity vary, conceptually these services propose to act as Consent Intermediaries (CIs) between individuals and data-using entities. From the perspective of individuals, CIs aim to consolidate the provisioning of consents under one control point, providing an access point through which individuals grant, view and withdraw consent to collect and use data. From the perspective of the services, CIs enable the outsourcing of privacy management. The CI, therefore, consolidates the consenting practices of many services and the consenting decisions of multiple individuals, in a conceptual change to the current dispersed practice as shown in Figure 1.

CIs strive to provide individuals with better control over their personal data, which is expected to lead to better privacy and larger benefits from their data. A recent opinion published by the European Data Protection Supervisor (2016), for example, sees the PIMS services, backed up by GDPR, as potentially leading to the empowerment of users. Yet empowerment and control might be illusory if making sense of the consequences of data disclosure decisions does not become easier than it currently is. In addition, CIs also introduce a new party in the consenting process, which may also have its own aims and incentives.

Analysis

We start by looking at the changes the intermediary may bring to the consenting process, and then go through how these changes could help overcome the obstacles. We conclude the analysis by addressing the nature of privacy self-management obstacles.

How could a consent intermediary change consenting?

Simply bringing consents under one control interface has the potential to make privacy self-management

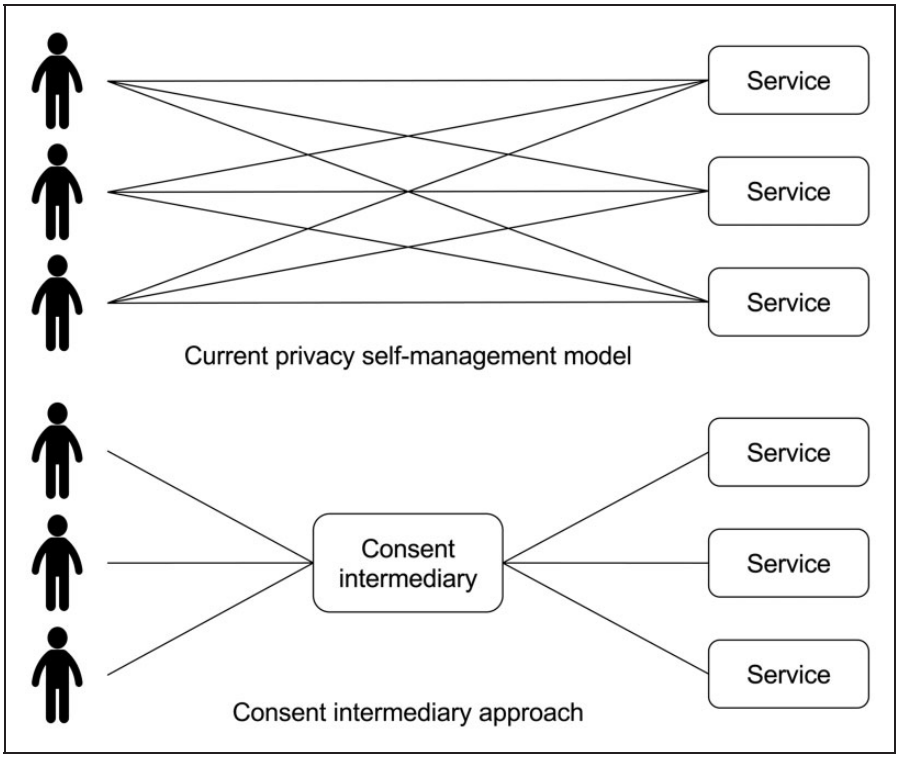

Figure 1. The conceptual change of the consent intermediary approach.

easier. It can help individuals make sense of the whole, be aware of past decisions, and take them into account in future decisions. This is particularly true if the CI presents consents in a comparable format. An overall view can help in situations in which privacy management fails due to individuals not being aware of the totality of their own decisions, which is, in light of the identified obstacles, a part of the problem but does not cover nearly all of its aspects.

Bigger changes can happen if the CI takes advantage of its intermediary position and builds new tools to aid decision-making, for example, by employing concepts which are already commonly used in other online services. Without an intermediary between individuals and data users, it would be much harder to build these tools.

First, online services and marketplaces routinely employ recommendations, predictions, ratings, and crowdsourcing to provide their users tailored information. In smartphone platforms, users give permissions to applications they install, and making privacy-conscientious decisions requires accounting for how the applications likely use those permissions. Liu et al. (2016) propose a 'privacy assistant' which provides personalised recommendations for application permissions based on user profiling. It is possible to apply a similar approach to the more general issue of providing consent. An intermediary service can leverage consent metadata, including information contained in consents themselves, and information on other users' actions regarding consents to provide more information at the point of decision. In experimental settings, timely presentation of privacy information has been found to lead to more privacy-protecting decisions (Kelley et al., 2013), and designs which highlight the implications of decisions have been found to have a similar effect (Harbach et al., 2014).

Second, there are several ways to automate actions based on, for example, rules and profiles. With a CI, it might be possible to automate some practical consent decisions. This might include straightforward recommendations based on preferences; users indicate their preferences, and the intermediary then recommends actions based on them. Privacy preferences can also be deduced automatically using data analysis, as has been done for privacy settings on Facebook (Fang and LeFevre, 2010) and for mobile applications (Liu et al., 2016). In the context of mobile applications, recommendations for access permissions by experts (Rashidi et al., 2015) and crowdsourcing (Agarwal and Hall, 2013) have also been proposed. With consents, individuals could similarly choose to automatically follow the recommendations formed collaboratively by engaged users or provided, for example, by a privacy advocacy group or a commercial provider.

Third, companies are fundamentally dependent on individuals as their data sources, and this position

could be leveraged for more favourable terms of data use. Currently, the position of individuals is characterised by low bargaining power over these terms. Attempts to balance similar asymmetries are in many other contexts based on the disadvantaged parties organising as a collective actor rather than as individuals, including well-known examples of consumer interest lobbies and unions in labour market negotiations. The CI could act as a platform for collective action to balance these power asymmetries by leveraging the presence of others in the same decision-making situation.

How could these changes help overcome the obstacles?

Timing and duration. As noted above, making information on consents viewable from a single point has the potential to increase individuals' awareness of the long-term aspects of privacy management. This could help individuals make decisions in a more systematic manner, particularly by mitigating the timing issue in the sense that the long-term effects of decisions can be better taken into account. Making sense of the whole can also be made possible by making use of consent metadata. For example, an individual might be made aware of all actors who have access to certain kind of data. It would be straightforward for the CI to employ nudges to revisit previous decisions (Liu et al., 2016) to see whether they still accurately represent current preferences. Prompts to re-evaluate consent might be issued periodically or be based on changed conditions such as the provision of new consent for similar purposes. Nudges and prompts could bring benefits similar to those of proposals for periodically expiring consents (Custers, 2016; Mayer-Schönberger, 2011), but would likely also exhibit similar problems; for some, they would likely be just another forced click of an 'agree' button without much thought (Custers, 2016).

Non-negotiability. Broad, non-negotiable consents make sense to many companies, as their business models drive them to make privacy policies as general as possible in terms of the quality, quantity and possible uses of personal data (Custers, 2016; Srnicek, 2017). Implementing negotiability of privacy policies, for example, by using smart contracts, may be costly in service design sense. Also, tweaking privacy policies before accepting them increases the decision-making effort required from individuals, and customised consents lead to the production of additional metadata and complicate data management (Custers, 2016). Despite the incentives for non-negotiable consent, there is nothing which fundamentally prohibits negotiations. To the extent that the lack of negotiations is attributable to each individual having low bargaining power against data collectors, individuals could organise as a collective entity to leverage the dependence of organizations on them as data sources. We argue that such collective action is difficult to achieve without some kind of coordinating entity, and the CI could act as one. Introducing an intermediary between individuals and data-collecting entities would, in any case, affect the power balance of the situation, and in the best case this would help individuals have a say over the terms under which their data is used. However, it seems safe to assume that the intermediary might leverage its position also for its own benefit, which may or may not align with the interests of the individuals.

Scale. Given the amount of effort expected from individuals, the scale problem seems difficult to overcome. However, we argue that it is not a problem of principle but is largely due to how privacy self-management is implemented in practice. Making each decision simple enough, or gathering many small but similar decisions under one higher-level decision, would make the whole decision-making effort more manageable. Automation and aides which help identify important decisions would work in this manner. Practical questions include whether or not it is possible to simplify decisions enough while fulfilling both the expectations individuals have of the ability to affect each decision, and the requirements data protection regulations place on the way consent is provided.

Aggregation. Data aggregation is an obstacle which is difficult to tackle in principle, as latent data emerges only ex post, after decisions have been made. The relationship between disclosed data and the consequences of this disclosure are therefore obscured, and latent data can make meaningful analysis of costs and benefits impossible. While it is likely impossible to fully overcome this obstacle, some improvements to the current situation can be envisioned. Making people aware of known outcomes of data aggregation, based not only on their own but also on others' past decisions, could help them become better informed about potential latent data. Consent metadata can, for example, be aggregated across individuals and used to provide information on what data others have chosen to disclose. Simply explaining the likely purpose of a mobile application's permission request has been found to play an important role in privacy decisions (Liu et al., 2016). Metadata can also be used to form predictions of likely consequences of disclosure decisions without access to the actual disclosed data, for example, predicting the potential revealing of contextual data based on location data alone, if others have already provided contextual and location data (Bellovin et al., 2013). Such consequences are, of course, a moving target, and predictions

would necessarily be coarse, but they might still be better than the heuristics individuals currently have to rely on.

Downstream uses. To the extent that downstream uses of data happen with user's consent, it is possible in principle to make it easier to take these uses into account in privacy decisions. For example, by combining consent metadata with other data sources, the network of data flows originating from the initial data collectors could be tracked in a manner similar to tracing the relationships of online ad platforms (Helmond and van der Vlist 2016), and visualising them might make sense-making easier. Naturally, the possibility of increasing transparency is limited to data flows which are consented to and trackable – excluding, for example, downstream uses through surveillance or data leaks. In addition, structural constraints of the data industry, such as opaque business practices and analytical layers which separate data sources from data uses, limit efforts to increase transparency (Crain, 2016).

Cognitive demands. The cognitive limits of human decision-making fundamentally restrict cost–benefit analyses. An obvious way to tackle this problem is to make decisions less demanding. On-time provision of relevant information could make it less demanding to be informed, but all efforts to make now-opaque consequences of data disclosure transparent run the risk of making each decision even more complex. Here, as in the context of the scale problem, one solution facilitated by a CI is to change the nature of the decisions; instead of considering each decision separately, an overall decision could be made on privacy management principles. The CI would offer a limited choice of more or less conservative privacy profiles, and then would recommend actions based on those profiles. At the extreme end, technically nothing prevents totally automated consenting, so that the intermediary would automatically provide consent on behalf of the individual or revoke consent from services no longer in use. Such solutions, however, may be at odds with current privacy regulations.

Social norms. The adherence to social norms makes an individual's privacy decisions dependent not only on their private costs and benefits, but also on others' expectations about those decisions. Tools which help evaluate the consequences of disclosure affect the private aspects of decisions to disclose data, and to the extent that social norms regulate those decisions, tools do not help. The obstacle that norms place in the way of privacy self-management is, therefore, insuperable within the individuated model, regardless of the privacy management tools developed. Of course, it should be kept in mind that norms with respect to the disclosure of data are not fixed, and they may change over time.

Social data. There is a fundamental inconsistency between privacy self-management and the social nature of personal data. Social data makes my privacy dependent on the choices of others (and vice versa). My goals and privacy preferences might be contradictory to those of others, and the private benefits someone draws from disclosing social data might, from their perspective, overcome the private costs imposed on others. While the interdependencies of decisions and the consequences of my decisions on others can be made more visible by using tools similar to those discussed above, no amount of awareness will solve this fundamental tension.

How difficult are the obstacles?

A key dilemma in all the discussed improvements to privacy self-management is that tools should help individuals take more complexity into account, and at the same time render decision-making easier. By revealing more of the consequences of data processing, we make the individual better informed, but this also makes decisions cognitively more demanding. Therefore, it seems to us that progress could best be made if CIs provided privacy management features on all of the three fronts described above. While not all the privacy self-management obstacles can be overcome, evaluation aides, decision automation and collective action have the potential to lead to better privacy self-management.

Based on our analysis, we can also deduce something about the nature of the privacy self-management obstacles. Some of them seem to be more practical in nature, and potentially solvable by developing tools for privacy management. Timing and duration, non-negotiability and the scale problem can, in principle, be solved by rethinking practices and providing new kinds of privacy management tools. While we consider these problems to be solvable in the sense that they are practical, it does not mean that they are easy to solve. At the other end of the spectrum, obstacles which feature social dimensions exhibit fundamental tensions with the individuated privacy self-management model and are therefore insuperable, unless the individuated principle of the model itself is changed. In between are the cognitive demands of decision-making, aggregation, and downstream uses of data. Privacy management tools can help to mitigate them, but they are challenging issues and exhibit aspects which we consider likely to be unsolvable. Table 2 presents this rough, by necessity, categorisation.

Table 2. Potential to overcome obstacles with privacy self-management tools.

Solvable	
Timing and duration	Practical problem of making it feasible to revisit decisions and revoke consent.
Non-negotiability	Practical problem of negotiating power.
Scale	Practical problem of making each decision easy enough.
Challenging	
Aggregation	Possible to mitigate by increasing awareness of latent data.
Downstream uses	Possible to mitigate by providing information on consented and traceable data flows.
Cognitive demands	Possible to mitigate by changing the nature of decisions.
Insuperable	
Social norms	Cannot be overcome within the individuated model.
Social data	Cannot be overcome within the individuated model.

Discussion

Our overview of the potential of the CI approach was largely positive in nature, in that we looked at the possibilities of developing features which are potentially beneficial for individuals. It is clear that the restrictions of the approach, the implications of the CIs, and the limitations of privacy self-management also merit discussion.

To begin with, the CI approach rest on the assumption that people are inclined to manage privacy. While the experienced lack of control and implicit means used to gain some control exhibit a demand for better privacy self-management tools, some might simply be happy with current services. It is likely that new tools for privacy management alone will not overcome disinterest.

The existence of economic incentives to maintain the current state of affairs should not be overlooked. The production of latent data is ingrained in the business models of many online companies (Srnicek, 2017; Zuboff, 2015); therefore, it is one underlying reason for the extensive collection of personal data in the first place. If the privileged position of organisations which collect and use data is an outcome of privacy self-management (Coll, 2014), then the current scattering of consents and the associated difficulties in privacy self-management serve existing business interests. Attempts to change the existing consent practices are, therefore, likely met with resistance. Here, legal developments such as GDPR can have a significant impact.

For several of the privacy management obstacles, it is evident that the problem is connected to the fundamental assumption of the privacy self-management model: that individuals themselves consider costs and benefits of data disclosure case by case. As outlined above, these problems could be managed by automating or delegating decisions. The level of abstraction can be increased, and the decision would then concern the rules of automation or to whom the decision would be delegated. Therefore, automation and delegation of consenting decisions could well lead to a better outcome, on the whole. The extent to which these actions are possible within current regulatory contexts is an object of research in its own right.

This assumption also makes privacy self-management an inherently individuated model. The social dimensions of personal data are hidden by framing the issue as 'my data' which is 'about me' (Crabtree and Mortier, 2015). This framing leaves open the question of social data which is not only about me, and focusing on individual cost–benefit analysis downplays the role of the norms which affect decisions. While it can be possible to make the social and societal consequences of data disclosure decisions more transparent, private cost–benefit analyses can still fail to take common good into account, which has the risk of leading to only locally optimal solutions; sometimes taking a broader societal or collective view may lead to a better, globally optimal solution. It would therefore be misleading to think about the social obstacles to privacy management as 'problems' which new consent practices can 'solve'. Rather, they are features of the privacy self-management model and present an inherent tension which cannot be overcome without changing the underlying individuated principle of the model. The obvious question, then, becomes how to take the collective aspect into account in the relationships between individuals and data collectors. Concepts such as networked privacy (Lampinen, 2015) and focusing on the group rather than the individual as the starting point of privacy (Taylor et al., 2017) pave the way towards such alternative models. While this conceptual discussion is ongoing, practical experiments in collective privacy management are underway in more limited contexts, for example, developing an extended notion of ownership of digital content and providing tools which help in reaching collective decisions regarding such content (Squicciarini et al., 2009). It might be possible to extend solutions like this to the more general context of consent as well, which would amount to developing a model to achieve common good from an individual privacy management starting point. CIs can likely function as platforms which facilitate the building of tools from these wider points of view as well. However, the individuated approach required by privacy regulations might render such solutions non-compliant.

While forms of automation could lead to easing the cognitive load of decision-making, automation does not come without its own trade-off: decision-making power is transferred to those forming profiles and recommendations, such as algorithm designers. More generally, given the power invested in intermediary positions between user and data-using entities, we should pay close attention to how intermediaries make use of this power. Intermediaries could channel collective action and set up governance mechanisms which participating organisations are expected to follow, which could well work favourably for individuals. But we should not assume this is the only possible outcome. The intermediary also has the capacity to affect the behaviour of its users, for example through discreet nudges or outright limits to choices. This leads to the possibility of coaxing users towards behaviours which serve its own ends. This also renders the intermediaries tempting targets for attacks, both for the troves of information they contain about the individuals in the form of consents, profiles, and policies, and for the power of influencing the individuals in their privacy decisions.

Conclusions

As we will live with the consent-based privacy self-management model for some time, it pays to investigate ways to make it better. From the recent developments of personal data services, we identified the concept of CIs which gather privacy decisions under a single control point. Based on our analysis, this provides only some direct remedies to the obstacles which currently hinder privacy self-management. However, intermediaries could be leveraged to develop tools to mitigate obstacles, helping people understand the decisions they make, better evaluate their consequences, and simplify the decisions themselves. We conclude that it is indeed possible to make privacy self-management work better, and some of its obstacles seem to be even solvable with new tools. However, not all of the obstacles can be tackled this way. Some obstacles seem challenging in the sense that they could be only mitigated but likely not solved. Finally the inherent problems related to individual-centricity of the model lead to insuperable problems that could be better approached if its individuated assumptions were relaxed.

Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their constructive feedback on this article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Abiteboul S, André B and Kaplan D (2015) Managing your digital life. *Communications of the ACM* 58(5): 32–35.

Acquisti A, Brandimarte L and Loewenstein G (2015) Privacy and human behavior in the age of information. *Science* 347(6221): 509–514.

Agarwal Y and Hall M (2013) ProtectMyPrivacy: Detecting and mitigating privacy leaks on iOS devices using crowdsourcing. In: *Proceeding of the 11th annual international conference on mobile systems, applications, and services*, Taipei, Taiwan, 25–28 June, pp.97–110. New York: ACM.

Alba D (2015) 23andMe teams with Big Pharma to find treatments hidden in our DNA. *Wired*.

Altman I (1975) *The Environment and Social Behavior. Privacy–Personal Space–Territory–Crowding*. Monterey: Brooks-Cole Publishing Company.

Andrejevic M (2014) The Big Data divide. *International Journal of Communication* 8: 1673–1689.

Anthes G (2015) Data brokers are watching you. *Communications of the ACM* 58(1): 28–30.

Bellovin SM, Hutchins RM, Jebara T, et al. (2013) When enough is enough: Location tracking, mosaic theory, and machine learning. *NYU Journal of Law & Liberty* 8: 555–628.

Böhme R and Köpsell S (2010) Trained to accept? A field experiment on consent dialogs. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Atlanta, USA, 10–15 April, pp.2403–2406. New York: ACM.

Chaudhry A, Crowcroft J, Howard H, et al. (2015) Personal data: Thinking inside the box. *Aarhus Series on Human Centered Computing* 1(1): 29–32.

Coll S (2014) Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. *Information, Communication & Society* 17(10): 1250–1263.

Cozy Cloud (2017) Cozy cloud website. Available at: https://cozy.io/en/ (accessed 20 April 2017).

Crabtree A and Mortier R (2015) Human data interaction: Historical lessons from social studies and CSCW. In: *ECSCW 2015: Proceedings of the 14th European conference on computer supported cooperative work*, Oslo, Norway, 19–23 September, pp.3–21. Cham: Springer.

Crain M (2016) The limits of transparency: Data brokers and commodification. *New Media & Society*. Epub ahead of print 7 July 2016. DOI:10.1177/1461444816657096.

Custers B (2016) Click here to consent forever: Expiry dates for informed consent. *Big Data & Society* 3(1): 1–6.

de Montjoye YA, Shmueli E, Wang SS, et al. (2014) openPDS: Protecting the privacy of metadata through SafeAnswers. *PLoS one* 9(7): 1–9.

Digi.me (2017). Digi.me website. Available at: https://digi.me (accessed 20 April 2017).

European Commission (2015) Data protection. Special Eurobarometer (431).

European Commission (2016) An emerging offer of 'Personal Information Management Services' – Current state of service offers and challenges. European Commission Report.

European Data Protection Supervisor (2016) EDPS opinion on personal information management systems. Towards more user empowerment in managing and processing personal data. Opinion 9/2016.

European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council. *Official Journal of the European Union* L119: 1–88.

Fang L and LeFevre K (2010) Privacy wizards for social networking sites. In: *Proceedings of the 19th international conference on world wide web*, Raleigh, USA, 26–30 April, pp. 351–360. New York: ACM.

Gigerenzer G and Selten R (2001) Rethinking Rationality. In: Gigerenzer G and Selten R (eds) *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press, pp. 1–12.

Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91(3): 481–510.

Harbach M, Hettig M, Weber S, et al. (2014) Using personal examples to improve risk communication for security & privacy decisions. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Toronto, Canada, 26 April–1 May, pp.2647–2656. New York: ACM.

Helmond A and van der Vlist FN (2016) Big Data advertising infrastructures: A comparative study of social media Ad platforms. In: *Internet, Politics & Policy 2016*. 22–23 September, Oxford, UK.

Hoofnagle CJ and Urban JM (2014) Alan Westin's privacy homo economicus. *Wake Forest Law Review* 49(261): 261–317.

Hub of All Things (2017) Hub of All Things GitHub page. Available at: https://github.com/Hub-of-all-Things (accessed 20 April 2017).

Kastrenakes J (2015) Spotify updates privacy policy with clearer language after backlash. *The Verge*.

Kelley PG, Cranor LF and Sadeh N (2013) Privacy as part of the app decision-making process. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Paris, France, 27 April–2 May, pp.3393–3402. New York: ACM.

Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Lampinen A (2015) Networked privacy beyond the individual: Four perspectives to "Sharing". *Aarhus Series on Human Centered Computing* 1(1): 1–4.

Lampinen A, Lehtinen V, Lehmuskallio A, et al. (2011) We're in it together: Interpersonal management of disclosure in social network services. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Vancouver, Canada, 7–12 May, pp.3217–3226. New York: ACM.

Liu B, Andersen MS, Schaub F, et al. (2016) Follow my recommendations: A personalized privacy assistant for mobile App permissions. In: *Proceedings of the 12th symposium on usable privacy and security*, Denver, USA, 22–24 June, pp.27–41. Berkeley: USENIX.

McDonald AM and Cranor LF (2008) The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4(543): 1–22.

Mai JE (2016) Big data privacy: The datafication of personal information. *Information Society* 32(3): 192–199.

Mayer-Schönberger V (2011) *Delete: The Virtue of Forgetting in the Digital Age*. Princeton: Princeton University Press.

Meeco (2017) Meeco website. Available at https://meeco.me (accessed 20 April 2017).

Norberg PA, Horne DR and Horne DA (2007) The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs* 41(1): 100–127.

Poikola A, Kuikkaniemi K and Honko H (2015) *MyData – A Nordic Model for Human-Centered Personal Data Management and Processing*. Helsinki: Finnish Ministry of Transport and Communications.

Rashidi B, Fung C and Vu T (2015) Dude, ask the experts! Android resource access permission recommendation with RecDroid. In: *2015 IFIP/IEEE international symposium on integrated network management (IM)*, Ottawa, Canada, 11–15 May, pp.296–304. New York: IEEE.

Schneier B (2010) A taxonomy of social networking data. *IEEE Security and Privacy* 8(4): 88.

Seife C (2013) 23andMe is terrifying, but not for the reasons the FDA thinks. *Scientific American*.

Snell K, Starkbaum J, Lauß G, et al. (2012) From protection of privacy to control of data streams: A focus group study on biobanks in the information society. *Public Health Genomics* 15(5): 293–302.

Solove DJ (2013) Privacy self-management and the consent dilemma. *Harvard Law Review* 126(7): 1880–1903.

Squicciarini AC, Shehab M and Paci F (2009) Collective privacy management in social networks. In: *Proceedings of the 18th international conference on world wide web*, Madrid, Spain, 20–24 April, pp.521–530. New York: ACM.

Srnicek N (2017) *Platform Capitalism*. Cambridge, UK: Polity Press.

Taylor L, Floridi L and van der Sloot B (2017) Introduction: A new perspective on privacy. In: Taylor L, Floridi L and van der Sloot B (eds) *Group Privacy. New Challenges of Data Technologies*. Cham: Springer International Publishing, pp. 2–12.

Turow J, Hennessy M and Draper N (2015) The Tradeoff Fallacy. How marketers are misrepresenting American consumers and opening them up to exploitation. A report from the Annenberg School for Communication. University of Pennsylvania.

Wachter S (2017) Privacy: Primus inter pares – Privacy as a precondition for self-development, personal fulfilment and the free enjoyment of fundamental human rights. Available at: https://ssrn.com/abstract=2903514 (accessed 20 April 2017).

Zetter K (2013) Hackers finally post stolen Ashley Madison data. *Wired Magazine*.

Zimmer M (2016) OkCupid study reveals the perils of Big-Data science. *Wired Magazine*.

Zuboff S (2015) Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30: 75–89.