MONASH University

MONASH BUSINESS SCHOOL

Department of Economics
ISSN number 1441-5429

Digital Privacy: GDPR and Its Lessons for Australia

Discussion Paper no. 2022-19

Ratul Das Chaudhury and Chongwoo Choe

Abstract:

Australia's Privacy Act 1988 is under review with a view to bringing Australia's privacy laws into the digital era, more in line with the European Union's General Data Protection Regulation (GDPR). This article discusses how the GDPR can be refined and standardized to be more effective in protecting privacy in the digital era while not adversely affecting the digital economy that relies heavily on data. We argue that an ideal data policy should be informative and transparent about the potential privacy costs while giving consumers a menu of opt-in choices into which they can self-select themselves.

Keywords: digital privacy, GDPR, opt-in

JEL Classification: K24

Ratul Das Chaudhury: Postdoctoral Research Fellow, Centre for Global Business, Monash Business School (email: ratul.daschaudhury@monash.edu); Chongwoo Choe: Director, Centre for Global Business, and Professor, Department of Economics, Monash Business School (email: chongwoo.choe@monash.edu).

Digital Privacy: GDPR and Its Lessons for Australia*

Ratul Das Chaudhury† and Chongwoo Choe‡

September 10, 2022

Abstract

Australia's Privacy Act 1988 is under review with a view to bringing Australia's privacy laws into the digital era, more in line with the European Union's General Data Protection Regulation (GDPR). This article discusses how the GDPR can be refined and standardized to be more effective in protecting privacy in the digital era while not adversely affecting the digital economy that relies heavily on data. We argue that an ideal data policy should be informative and transparent about the potential privacy costs while giving consumers a menu of opt-in choices into which they can self-select themselves.

Keywords: digital privacy, GDPR, opt-in
JEL Classification: D21, K24, L51

1 Introduction

If this is the age of information, then privacy is the issue of our times. Activities that were once private or shared with the few now leave trails of data that expose our interests, traits, beliefs, and intentions. (Acquisti et al. 2015).

A popular business model adopted by many of the world's largest tech platforms is the so-called broadcasting model where services are provided free in return for advertising revenue. These firms collect and process consumer data generated from the use of their "free" services. Consumer data thus gathered can help create value for the business in various ways: it can be used for product improvement, for developing new business models, or for general management purposes; the data can be also monetized through

*This article is partly based on the authors' submission to the Australian Competition and Consumer Commission's Digital Platform Services Inquiry. We thank Zhijun Chen, Stephen King, and Chengsi Wang for useful comments. We gratefully acknowledge financial support from the Australian Research Council (grant number DP210102015). The usual disclaimer applies.

†The corresponding author. Postdoctoral Research Fellow, Centre for Global Business, Monash Business School, Clayton, Victoria 3800, Australia; Email: ratul.daschaudhury@monash.edu.

‡Director, Centre for Global Business, and Professor, Department of Economics, Monash Business School. Email: chongwoo.choe@monash.edu.

sales of data-based services or even by direct sales of data to third parties.[1] A famous quote dating back to the 1970s in relation to advertising in commercial broadcasting resonates even louder in the digital era: *if you are not paying for the product, then you are the product.* But the key difference between the traditional broadcasting model and the business model in the digital era is the role played by consumer data.

In the age of digital transformation, buyers are not only consumers but also producers of data, which in turn becomes a valuable input to production of goods and services. Indeed, data is the new oil in the digital era, as famously declared by the *Economist* in 2017.[2] Consumer-generated data analyzed with powerful machine-learning tools can enable firms to offer new or improved products, develop more target-oriented business models, and venture into new business opportunities (Hagiu and Wright, 2020). Online recommendation systems and targeted advertising have become the cornerstone of modern-day marketing. The availability of big data and finer-grained analysis has enabled firms in some industries to exercise personalized pricing, once considered only a theoretical possibility (Choe et al. 2018; Chen et al. 2020).

Consumer data is collected not only by tech platforms that consumers directly interact with but also by data brokers who collect and sell data to third parties. There are about 4,000 data brokers globally, including companies such as Acxiom and Oracle, who keep an enormous amount of data about individual consumers, ranging from relatively harmless data such as the city of residence to more sensitive data such as health issues or police record. A recent estimate suggests that the global data broker market is worth approximately US$250 billion in 2020 and is expected to grow to US$365 billion in 2027.[3]

Given the stratospheric rise of large tech platforms, the expansion of data brokerage industry, and the rapid growth in online activities, consumers are increasingly concerned about the privacy risks associated with how their personal data is collected and shared. According to the Australian Community Attitudes to Privacy Survey 2020 (OAIC, 2020), privacy is a major concern for 70 percent of Australians, and almost 9 in 10 want more choice and control over their personal information. The survey also finds that 84 percent of Australians perceive identity fraud and data breaches as the biggest risk to data privacy. Such a concern is well justified: from January to June 2021, the Office of the Australian Information Commissioner received 446 data breach notifications, with about half of these breaches resulting from cyber security incidents.[4] Similar sentiments

[1] See, for example, "Data monetization: New value streams you need right now," *Forbes*, June 9, 2020.

[2] "The world's most valuable resource is no longer oil, but data," *The Economist*, May 16, 2017. For a comprehensive review of the literature on the digital economy, see Goldfarb and Tucker (2019).

[3] https://www.maximizemarketresearch.com/market-report/global-data-broker-market/55670/

[4] https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-

towards online privacy are observed in the US: a study by Pew Research Center reports that about 80 percent of Americans think their personal data is less secure now and that data collection poses more risks than benefits (Auxier et al. 2019).

Before the advent of the digital era, privacy was not viewed as something that needed regulatory protection, an argument put forward most notably by the Chicago School. For example, Posner (1981) argued that fully-informed individuals with control over their personal information would make rational decisions about disclosing or withholding information, so that regulatory intervention would interfere with efficient flow of information.[5] As the opening quote suggests, however, the monitoring of personal information is ubiquitous in the digital era. Moreover, a recent study finds that 80 percent of the data collected by online service providers through mobile apps is not related to the direct performance of the app, but is primarily shared with data brokers or third parties for analytics, advertisement, etc. (Bian et al. 2022). In short, people are hardly aware of the extent to which their personal information is collected and shared; nor do they have full control over their personal information.[6]

The European Union's General Data Protection Regulation (GDPR) that came into effect in 2018 is a response to the growing privacy concern in the digital era. It was followed by similar privacy laws and regulations around the world. In Australia, the Attorney-General announced in 2019 that the Australian Government would conduct a review of the Privacy Act 1988, as part of the government's response to the Australian Competition and Consumer Commission's Digital Platforms Inquiry.[7] The review seeks to bring Australia's privacy laws into the digital era and, therefore, the GDPR assumes critical relevance in the review.

The GDPR is a good starting point in protecting privacy in the digital era. But it is not without problems. As we argue in this article, the GDPR needs to be improved and refined not to stifle competition and investment in data-driven businesses while protecting privacy more effectively. The purpose of this article is to critically assess the GDPR with a view to offering some recommendations as to how the GDPR can be modified to balance the tradeoff between the benefits from data and privacy. The GDPR is quite comprehensive covering all aspects relevant to the collection and processing of

statistics/notifiable-data-breaches-report-january-june-2021

[5] Acquisti and Grossklags (2005) challenge this view by arguing that people are not informed enough to make privacy-sensitive decisions and, even when they are sufficiently informed, they trade off long-term privacy for short-term benefits. The latter is also related to the so-called the privacy paradox whereby people relinquish privacy in exchange for small incentives, even though their stated preferences for privacy may be strong (Berendt et al. 2005; Athey et al. 2017).

[6] For an excellent review of the economics literature on privacy, see Acquisti et al. (2019).

[7] https://www.ag.gov.au/integrity/consultations/review-privacy-act-1988

personal data including data sharing. To keep our discussion focused and at manageable length, we mainly discuss data collection in this article.

The rest of the article is organized as follows. In Section 2, we discuss the key elements of the GDPR, provide evidence on the effects of the GDPR, and document the problems identified with the GDPR. Section 3 discusses how the data policy under the GDPR can be modified to better manage the tradeoff between privacy and the benefits from data. This is followed by a brief conclusion in Section 4.

2 The GDPR and its effects

Privacy concerns arise when personal data is collected and shared without the knowledge or consent by the data subjects.[8] On the other hand, consumer data is a valuable input in the digital era, as explained previously. Moreover, data sharing can be vital for innovations leading to new products and services. For example, sharing health data can be instrumental in innovations in health care (or tackling the global pandemic such as the COVID-19), and sharing detailed automation data is required for developing the technology for safer self-driving cars. In addition, given the data advantages enjoyed by large digital platforms that may also work as barriers to entry, data sharing can be considered necessary in promoting competition, open banking being a prime example.[9] Strict privacy laws may help protect privacy but stifle innovation and competition, and harm data-driven businesses. Thus, the key question is how to balance the benefits from data and the costs from privacy breach.

Before the GDPR, the Data Protection Directive (Directive 95/46/EC) adopted by the European Union in 1995 specified a number of guidelines for the collection and use of personal data. The Data Protection Directive builds on the principles such as notice, purpose, consent, security, disclosure, access, and accountability. But these guidelines were non-binding and not specific enough. Consequently, online businesses often relied on opaque processes in collecting data, rendering consumers little or no control over how their data is collected and used. In this section, we first discuss how firms collected consumer data before the GDPR. Then we discuss how the GDPR tried to rectify the problem, after which we document studies that report 'unintended' consequences of the GDPR.

[8] One may ask if the problem can be solved by anonymizing personal data before processing it, which is indeed a requirement in various data protection laws. However, Rocher et al. (2019) demonstrate that anonymizing personal data through deidentification is not a fail-safe way to protect privacy, showing that almost all Americans can be correctly re-identified in any dataset using 15 demographic attributes.

[9] For discussions on the costs and benefits of customer data sharing in the digital era, see, for example, Liu and Serfes (2006) or Choe et al. (2022).

2.1 Data collection

There are various ways digital businesses collect user data, use of internet cookies being one of the most popular methods. Cookies (HTTP cookies, internet cookies, web cookies, or browser cookies) are small text files that are downloaded into the user's device by a web browser when the user visits a particular website. They are browser- and site-specific.[10] Cookies were initially designed to enhance user experience, reduce network traffic, and lower server storage costs by enabling web servers to store useful user information including the browsing activity and retrieve this information during subsequent page visits. Over time, however, third-party cookies have become commonly used by analytics firms and advertisers primarily to gather user data. These are cookies issued by an external domain and not by the website a user is browsing, and can track a user across websites.

A website's cookie policy can be based on users' opt-in consent or opt-out consent. In the former, no data collection is set by default and the website can collect data only when the user explicitly opts in to data collection by agreeing to accept the website's cookies. In the latter, data collection is set by default and the user has to act proactively to opt out if she does not want to accept the website's cookies. Studies show that default settings matter for individual decisions in various contexts (Acquisti et al. 2015) including online privacy policies. For example, Johnson et al. (2002) provide evidence from online privacy experiments showing that opt-in results in much lower levels of participation (20%) than opt-out (75%). In a similar vein, Johnson et al. (2020) examine the AdChoices program in the US and its opt-out mechanism for data consent, and report that only a small fraction of consumers opt out of online behavioral advertising: only 0.23% of ad impressions are from opt-out consumers. Consequently, websites can collect more data when they rely on opt-out consent than opt-in consent.

Prior to the GDPR or in jurisdictions without GDPR-style data protection laws such as the California Consumer Privacy Act, digital businesses tried to keep users in the dark regarding how their data is collected. Their websites typically detailed their privacy policies in a long and complex legal language, but without much information on their cookie policies. Even when the website provides information on their cookie policies, opt-out consent was a dominant form of data collection. It is conceivable that some tech-savvy, privacy-conscious consumers may proactively choose to opt out of the

[10] As an example, Google Analytics uses a cookie named '_ga cookie' to assign a client ID to a user, which can be used to track the user in subsequent visits to the website. The _ga cookie comprises four distinct values (version, domain, random unique ID, and the first visit time stamp), and is used to uniquely identify the user. Each time the user takes action on a website or an app (called a 'hit'), the data and the user's client ID are sent back to Google Analytics.

website's data collection, or delete cookies after each session. Nonetheless, the absence of clear opt-in choice resulted in unregulated collection of personal data with potential privacy breach, prompting various privacy regulations and laws around the world.

2.2 The GDPR

The stated purpose of the GDPR is to protect natural persons with regard to the processing of personal data, to promote the free movement of such data, and to repeal Directive 95/46/EC. The GDPR is the most stringent law governing personal data protection. It was adopted on April 14, 2016 and became enforceable on May 25, 2018. The GDPR builds on the same principles as its predecessor, the Data Protection Directive 95/46/EC, but it superseded the Directive with more specific data protection requirements, stiffer enforcements, and penalties for non-compliance. Importantly, the GDPR enhances individuals' control over data by stipulating the right to explicit consent, the right to data erasure, and the right to data portability. The GDPR's consent requirement stipulates that consumers be allowed to make an informed, specific, and unambiguous consent to processing their data. Thus, it requires in principle opt-in consent to data collection, which essentially bans data controllers from using opt-out options, a predominant way to obtain consent prior to the GDPR.

The GDPR became a blueprint for various privacy regulations in countries such as Chile, Brazil, Japan, New Zealand, Singapore, South Korea, etc. The US does not have a federal-level law on consumer privacy like the GDPR. Instead, a few US states have laws to protect consumer privacy, with California taking the lead. The California Consumer Privacy Act (CCPA) was signed into law in 2018 and took effect from 2020. The CCPA secures new privacy rights for consumers in California by providing them with the rights to know, to delete, to opt-out, and to non-discrimination in regards to their personal information. It is worth noting that the CCPA requires opt-out rather than opt-in as in the GDPR, but it limits selling of personal information, requiring a "Do Not Sell My Personal Information" link to be included by businesses on their homepage.[11] Other states such as Maryland (Maryland Online Consumer Protection Act) and New York (New York Privacy Act) also require firms to inform consumers about the broad categories of information shared with third parties, but without giving consumers an opportunity to opt-out.

[11] As an example, see https://privacy.thewaltdisneycompany.com/en/current-privacy-policy/.

2.3 The effect of the GDPR on cookie policies

Following the enactment of the GDPR and its commencement in 2018, there have been a significant increase in the use of cookie consent notices, transparent display of privacy policy and opt-in consent, and some decrease in the use of cookies. At the same time, there have been numerous reported cases of GDPR data breaches and fines for non-compliance.

First, Degeling et al. (2019) examined 500 most popular websites for each EU country - 6,579 websites in total - between December 2017 and October 2018. They found a significant increase in the display of cookie consent notices, or cookie banners, which inform users about a site's cookie use and user tracking practices. There was a 16% rise in the implementation of cookie consent notices among these websites, from 46.1% in January 2018 to 62.1% in May 2018. The websites displaying cookie banners increased by 43% in Ireland and 45.4% in Italy.

Second, according to Degeling et al. (2019), the majority of websites they examined had some form of privacy policies in January 2018, which rose to 84.5% after May 2018. Countries with a lower rate of privacy policies (e.g., Latvia) added more privacy policies than those where privacy policies were already common (e.g., Germany, Spain). As for industries, the availability of privacy policies in the EU increased by 9.7% in education, 7.1% in health, and 6.8% in government websites, to name but a few.

Third, during the past decades, the use of third-party cookies had been increasing, largely due to the increased use of web analytics, targeted advertising, and marketing campaigns. For example, as of 2014, many websites set over 100 third-party cookies, with a maximum number of cookies (both first and third-party) reaching over 800.[12] A month after the GDPR took effect, Degeling et al. (2019) found no significant change in the use of third-party cookies although the number of first-party cookies decreased from 22 to 18 on average. On the other hand, Libert et al. (2018) found from popular news websites in seven EU countries that the average count of third-party cookies per page has gone down by 22% following the GDPR, 45% in the UK, 33% in Spain and 32% in Italy and France. They also found that the GDPR led to a reduction in advertising and marketing cookies by 14%, and social media cookies by 9%.

Finally, there have been numerous cases of GDPR non-complaince and attendant fines. The GDPR Art. 83 and 84 stipulate that relevant national authorities must assess and impose fines for data protection violations. The fines could be up to 20 million euros or 4% of the total global turnover of the preceding fiscal year, whichever is higher.

[12] https://en.wikipedia.org/wiki/HTTP_cookie

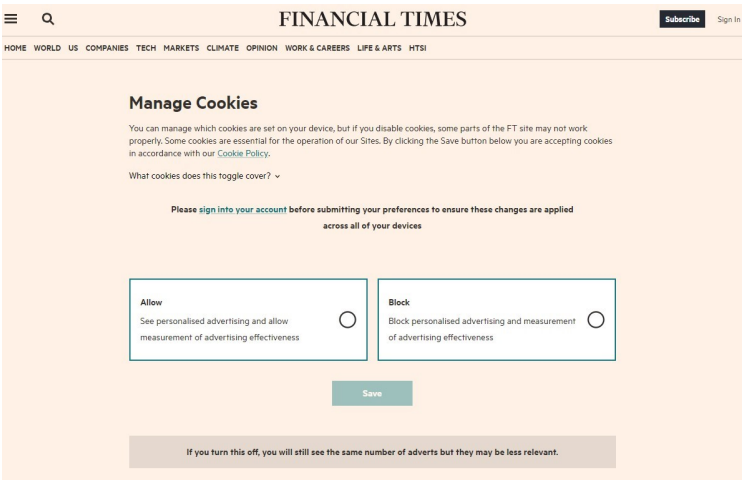
Nonetheless, many businesses were slow in getting their websites GDPR-compliant. By May 2018, over 800 fines were issued for GDPR non-compliance. The biggest fine to date is 746 million euros that the Luxembourg National Commission for Data Protection imposed on Amazon on 16 July, 2021 for violating data processing guidelines and forcing users to comply with cookie policies. Other examples of large fines include 225 million euros for WhatsApp, 90 million euros for Google Ireland, 60 million euros for Facebook, 20 million pounds for British Airways, and 20.4 million euros for Marriott.[13] More recent cases are the fines France's privacy watchdog (CNIL) levied on Google (150 million euros) and Facebook (60 million euros) in January 2022 for making it difficult for users to reject cookies,[14] and the fine Ireland's Data Protection Commission issued in September 2022 to Instagram (405 million euros) over children's data privacy.[15]

2.4 The 'unintended' consequences of the GDPR

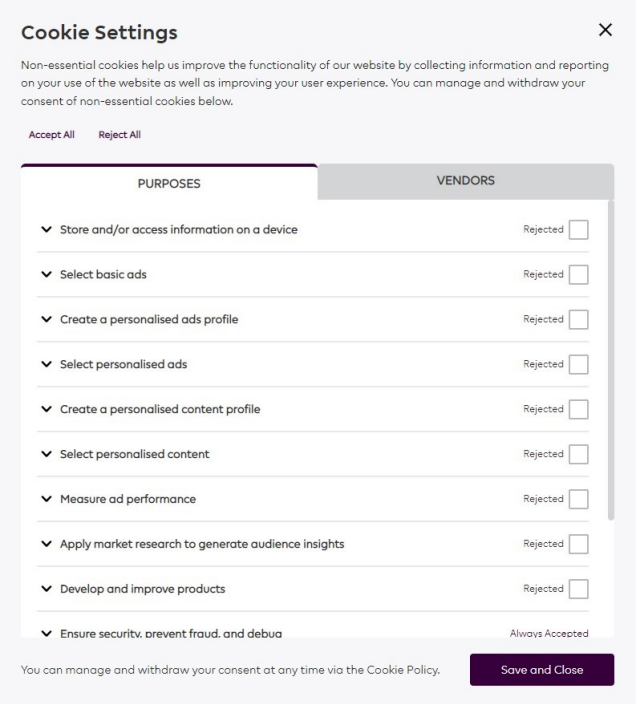
In the assessment of the GDPR two years after it took effect, the European Commission hailed it as an overall success, in particular by empowering citizens through enhanced transparency and privacy rights, and by providing businesses with a harmonized framework for the protection of personal data.[16] Although the information provided in the previous section lends some support to this assessment, our view is that such an assessment needs to be taken with a grain of salt.

First, the GDPR's cookie rules do not go beyond requiring opt-in consent and, therefore, are not refined enough for consumers to make an informed opt-in choice. Indeed, GDPR-compliant cookie policies can take different forms as long as they are largely consistent with the GDPR's principle of opt-in consent for data collection. For example, a website may have a simple binary opt-in policy as in the Financial Times where a user can allow or block all non-essential cookies, as shown below.[17] In this case, consumers may not understand the full implications of opt-in.

[13] https://termly.io/resources/articles/biggest-gdpr-fines/
[14] https://www.reuters.com/world/europe/france-imposes-fines-facebook-ireland-google-2022-01-06/
[15] https://www.bbc.com/news/technology-62800884
[16] https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_1166
[17] https://www.ft.com/preferences/manage-cookies



In case of the English Premier League Football website, non-essential and third-party cookies are further divided into nine different groups with a brief description of their purposes, and users can opt in to each of them separately, as shown below.[18]



Although the English Premier League Football website's cookie policy is more infor-

[18] https://www.premierleague.com/cookie-policy

mative than the one with binary choice, consumers who are not tech-savvy may find it difficult to make an informed choice. Given that consumers' main concern in data collection is privacy, it would be better if information is given on what type of data is collected and how privacy-invasive it is. Consequently, a more careful study is needed to examine whether the GDPR has empowered European citizens through enhanced transparency and data privacy. But the available evidence does not appear to support the European Commission's assessment. For example, Nouwens et al. (2020) scraped the designs of the five most popular consent management platforms introduced after the GDPR on the top 10,000 websites in the UK, and find that dark patterns and implied consent are ubiquitous.[19] In addition, Obar and Oeldorf-Hirsch (2020) provide experimental evidence showing that participants demonstrate general apathy towards privacy and select the 'quick join' clickwrap to simply access the website while ignoring the website's privacy policy and terms of service.[20] If a user agrees to the consent notices to save time and get past the large banners to get access to the website's content, then it defeats the purpose of 'informed consent' stipulated in the GDPR guidelines.

Second, research shows that the GDPR has had an adverse effect on data-driven businesses and innovation. The GDPR's opt-in policy can be too blunt a tool in balancing the trade-off between privacy and the benefits from data. As mentioned previously, opt-in consent results in less participation than opt-out consent, implying a decrease in data collection, which in turn can harm businesses and innovations that rely heavily on data. Indeed, the GDPR has been shown to significantly reduce the number of visits to a website (Aridor et al. 2020; Schmitt et al. 2020). For example, Aridor et al. (2020) report about 12.5% reduction in total cookies after the GDPR. In addition, Jia et al. (2021) report that the GDPR has dampened incentives to invest in data-related B2C ventures while Janßen et al. (2022) show that the GDPR has induced exit of about 1/3 of available apps at the Google Play Store. Finally, strict privacy laws can tilt the playing field in favor of large firms (Campbell et al. 2015). This is supported by several studies that report evidence that the GDPR increased market concentration in websites (Schmitt et al. 2020) and web technology services (Johnson et al. 2022; Peukert et al. 2022). Somewhat related, Apple's release of privacy label requirements in 2020 is shown to have resulted in decrease in iOS app downloads and app developers' revenue, but the smaller firms are more adversely affected than larger firms (Bian et al. 2022).

[19] A dark pattern is a deceptive user interface that is designed to trick users into doing things that they did not intend to.

[20] Related evidence on websites' nudging consumers into making specific choices is reported in Machuletz and Böhme (2020), and Matte et al. (2020). Utz et al. (2019) provide experimental evidence in support of this.

Put together, one may question if the GDPR is effective in managing the trade-off between privacy and the benefits from data. It could well be that the GDPR's focus was too much on privacy without fully taking into account the benefits from data and the implications for competition. Even on the privacy side, however, it is questionable if the GDPR enabled consumers to make an informed choice in agreeing to the processing of their data. In addition, the case is rather clear, and the evidence is accumulating, that the GDPR has harmed data-driven businesses.

3 Lessons from the GDPR

Given that Australian businesses operating outside the European Union are not subject to the GDPR, we observe that Australian websites vary widely in the way they display cookie banners or provide information on their privacy policies. Many websites do not display cookie banners that allow users to opt in or out of their cookie policies. For example, the Age provides a long and detailed privacy policy statement without giving clear opt-out or opt-in choices on its website; in order to opt out, users are asked to send an email.[21] As another example, Commonwealth Bank of Australia describes the types of cookies they use along with an instruction of how users can delete cookies from their browsers, but not the GDPR-style opt-in boxes that users can tick.[22] In contrast, the Australian Broadcasting Corporation displays a cookie banner that gives users an option to accept only required cookies or all cookies including performance and marketing cookies, hence is GDPR-compliant, albeit in the simplest way.[23]

In this section, we discuss how the data collection under the GDPR can be modified to better manage the trade-off between privacy and the benefits from data. Specifically, we focus on how the various cookie policies described above can be refined and standardized in a way that is more informative to consumers while not leading to unnecessary loss of valuable data. The key starting point is to recognize the fact that there are different types of data with different benefits and privacy costs, and consumers' attitudes towards privacy are also different across individuals. We discuss this below, followed by suggestions as to how the cookie policy under the GDPR can be modified. We then provide an illustrative example to clarify our main point.

[21] https://login.nine.com.au/privacy?client_id=theage
[22] https://www.commbank.com.au/important-info/cookies.html?ei=CB-footer_cookies
[23] https://help.abc.net.au/hc/en-us/articles/4447588409871

3.1 Data types and consumer heterogeneity

The GDPR Art. 4(1) defines personal data as "any information relating to an identified or identifiable natural person ('data subject'); [...] such as a name, an identification number, location data, an online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

Clearly, personal data includes a host of information, some of which may be more valuable to the firm than others. For example, a consumer's income level would be more valuable information than the gender information to the firm that uses the information for targeted promotion. Likewise, some personal data may cause more privacy concerns than others when shared with the firm. In an experimental study, Lin (2022) estimates consumers' intrinsic preferences for privacy based on their willingness to accept (WTA) a monetary compensation in exchange for their personal data. She reports the estimated WTA of \$0.14 for gender information and \$3.82 for income information. These observations suggest that one needs to classify personal data based on at least two attributes, (benefits to the data collector, privacy costs to the data subject). A logical conclusion is that a desirable cookie policy is the one that leads to the collection of more data types with larger benefits and lower privacy costs.

Consumers also differ in their attitudes towards privacy because what constitutes sensitive personal information differs across individuals. Aquisti et al. (2015) discuss studies that cluster individuals to three privacy segments: privacy fundamentalists, pragmatists, and unconcerned. Goldfarb and Tucker (2012) report differing preferences for privacy across different age groups. Lin (2022) provides experimental evidence that shows people are highly heterogeneous in their preferences for privacy. For example, more educated and wealthier consumers tend to have stronger preferences for privacy. Recognizing such heterogeneity is important since one-size-fits-all cookie policies can result in either too much or too little data collected. A desirable cookie policy that recognizes consumer heterogeneity is the one that offers a menu of opt-in choices, into which different consumers can self-select themselves.

3.2 Towards a more effective cookie policy

There are two main factors to consider in designing an effective cookie policy. First, it needs to be informative and easy to understand. Research shows that users respond more effectively when privacy notices are concise and displayed in a salient way (Ebert et al. 2021). Apple's Privacy Nutrition Labels introduced in 2020 serve as a good ex-

ample. Their purpose is to provide users with standardized and transparent information regarding the way iOS app developers collect and use consumer data.[24] The privacy labels fall into three categories: Data Used to Track You, Data Linked to You, and Data Not Linked to You. Thus one can say the first category represents the most invasive data collection, and the third, the least invasive.[25] Bian et al. (2022) report that the introduction of privacy labels was conducive to raising privacy awareness, consistent with the experimental evidence in Ebert et al. (2021).

Second, users need to be given clear choices presented in a transparent way. For example, cookies can be divided into several groups, depending on the types of information collected, how invasive the tracking can be, and for what purposes the data is used. Once again, Apple's privacy labels are a good example in this regard. However, Apple does not allow consumers to opt in to only a subset of cookies: the user faces a binary choice of agreeing to all data collection or none.[26] After Apple introduced the binary choice for opt-in consent, just about 4 percent of US users are reported to have opted in, with adverse effects on app developers and advertisers.[27] Giving users more choice may have resulted in more users opting in to a subset of cookies, thereby preventing unnecessary loss of valuable data.

Based on the above discussions, we argue that a desirable cookie policy needs to combine the transparency and informativeness as in Apple's privacy labels with several options to opt in to different sets of cookies. A simple example would be to classify all cookies into three categories depending on how privacy-invasive the data collection can be, as in Apple's Privacy Nutrition Labels. Each category needs to have a clear explanation of possible privacy costs and the expected benefits if consumers opt in. Given this, consumers have the choice to opt in to each category of cookies separately. Consumers choosing to opt in to highly privacy-invasive cookies would do so because they are less privacy-sensitive and/or because they expect extra benefits by opting into that category of cookies, which more than offset their privacy concerns. Consumers with significant privacy concerns may opt in to only the least invasive category of cookies, even though they may expect some reduction in the quality of service. This way, consumers can self-select themselves into different sets of cookies, thereby optimally balancing their

[24] https://www.apple.com/au/privacy/labels/

[25] For more details on the privacy labels, see Bian et al. (2022).

[26] Apple used to allow its app developers to track users' online activities by using Apple's IDFA (identifier for advertisers), a unique ID assigned to an Apple device. Consent to IDFA-tracking was set by default, although users could opt out. After its iOS 14.5 update in 2021, Apple introduced GDPR-style opt-in consent whereby users are given a binary option to click "Allow" button in a pop-up message (https://developer.apple.com/app-store/userprivacy-and-data-use/).

[27] https://mashable.com/article/ios-14-5-users-opt-out-of-ad-tracking

privacy costs and the utility from using the website.[28] As a result, different amounts of data are collected from different types of consumers, which improves upon the case where opt-in choice is binary. In the next section, we illustrate this idea with a simple example.

3.3 An illustrative example

Consider an economy with two consumers, indexed $i = 1, 2$, two types of data for each consumer, denoted by $\theta = a, b$, and one digital business which we simply call firm. Each type of data has value to the firm denoted by $\pi_\theta > 0$. Data also has additional value to the economy as a whole because data can create external benefits beyond the firm that collects it.[29] Thus the social value of data exceeds the value to the firm, which we denote by v_θ where $v_\theta > \pi_\theta$. A consumer agreeing to share her data with the firm incurs expected privacy cost that is consumer- and data-dependent, denoted by $c_{i\theta} > 0$ for $i = 1, 2$ and $\theta = a, b$. For example, the privacy cost incurred by consumer 1 in sharing type-a data with the firm is c_{1a}.[30]

We assume consumer 1 is more privacy-sensitive than consumer 2 in the sense that $c_{1\theta} > c_{2\theta}$ for $\theta = a, b$. In addition, type-a data is less privacy-invasive, hence leads to lower privacy cost to consumers, than type-b data. But it also has lower value to the firm and the economy as a whole.[31] For example, type-a data can be the consumer's city of residence and type-b data can be her health data. Put together, we assume that $c_{1a} < \pi_a < v_a$ and $\pi_b < v_b < c_{1b}$. Thus, it is socially optimal to collect only type-a data from consumer 1. For consumer 2, we assume $c_{2\theta} < \pi_\theta$ for $\theta = a, b$ so that it is socially optimal to allow both types of data to be collected.

Consider first the case where the firm can costlessly collect data and consumers do not make opt-in decisions.[32] This may describe the situation before the GDPR where the traditional broadcasting model applies. Consumers may simply log in to the firm's website to enjoy its 'free' service without knowing that their data is being collected. Since the firm does not need to induce consumers' opt-in, it only cares about π_θ and

[28] This is a direct application of mechanism design under adverse selection, of which classic references are Mussa and Rosen (1978) or Baron and Myerson (1982).

[29] For the discussions on the positive externalities from data, see, for example, Fainmesser et al. (2019) or Bergemann et al. (2022).

[30] Sharing data by one consumer can impose privacy costs on other consumers (Choi et al. 2019; Acemoglu et al. 2019; Ichihashi, 2021). For simplicity, we do not consider such negative data externalities. But the analysis can be extended without difficulty.

[31] Such a correlation between privacy cost and the value to the firm may be reasonable for some data and some industries, while the correlation can be in the other direction in other cases. Although we do not consider other cases, the analysis can be done in an analogous way.

[32] Our main point stays robust when we allow consumers to make opt-in decisions with a small probability, as long as that probability is smaller than that under the GDPR.

ignores the consumer's privacy cost. Consequently, it will collect both types of data from both consumers. This leads to socially suboptimal data collection: too much data is collected.

Suppose now consumers pro-actively make opt-in decisions but the opt-in choice is given in a binary form in which the consumer opts in to both types of data or none. Given the prevalence of binary opt-in consent in GDPR-compliant websites, one can say this is a reasonable description of the situation after the GDPR. Unlike the first case, cookie banners and opt-in consent boxes inform consumers of possible privacy costs that may follow their opt-in decisions. This means that, in order to induce consumers to opt in, the firm needs to compensate consumers for the privacy cost, either by providing improved service to opt-in consumers, or even offering monetary incentives. Denote the value of compensation by γ, which is assumed to be equal to the cost to the firm in providing the compensation. If $\gamma \geq \sum_\theta c_{1\theta} > \sum_\theta c_{2\theta}$, then both consumers opt in. If $\sum_\theta c_{1\theta} > \gamma \geq \sum_\theta c_{2\theta}$, then only consumer 2 opts in. If $\gamma < \sum_\theta c_{2\theta}$, then neither consumer opts in. This leads to the following two cases. First, if $\sum_\theta c_{1\theta} > \sum_\theta \pi_\theta$, then the firm cannot induce consumer 1's opt-in, but can choose $\gamma \in (\sum_\theta \pi_\theta, \sum_\theta c_{2\theta})$ to induce consumer 2's opt-in, the latter because we assumed $\sum_\theta c_{2\theta} < \sum_\theta \pi_\theta$. Second, if $\sum_\theta c_{1\theta} \leq \sum_\theta \pi_\theta$, then the firm can choose $\gamma \in (\sum_\theta c_{1\theta}, \sum_\theta \pi_\theta)$ to induce both consumers' opt-in. Thus, the binary opt-in choice leads to a weakly smaller amount of data collected than when consumers do not make opt-in decisions, but the socially optimal data collection is not possible under the binary opt-in choice, given the consumer heterogeneity.

Finally, consider the case where consumers can choose to opt in to each data type separately. Because consumer type is private information, the firm can offer opt-in compensation for each data type only. We will show that this case can lead an outcome with the socially optimal data collection. Denote by γ_θ the compensation for opt-in to type-θ data. Because $c_{1a} < \pi_a$, the firm can choose $\gamma_a \in (c_{1a}, \pi_a)$ and induce consumer 1 to opt in to type-a data. Such γ_a will also induce consumer 2 to opt in to type-a data because $c_{2a} < c_{1a}$. Next, the firm can induce consumer 2 to opt in to type-b data by choosing $\gamma_b \in (c_{2b}, \pi_b)$. But consumer 1 will not opt in to type-b data because $\gamma_b - c_{1b} < \gamma_b - \pi_b < 0$. In sum, the cookie policy where each consumer is allowed to opt in to each type of data separately with compensation given by $\gamma_a \in (c_{1a}, \pi_a)$ and $\gamma_b \in (c_{2b}, \pi_b)$ implements the socially optimal data collection. Under this cookie policy, consumer 1 self-selects herself into opting in to only type-a data while consumer 2 opts in to both types of data, hence the socially efficient amount of data is collected.

Even though consumers' privacy concerns remain private information which the firm cannot use in designing its cookie policy, a suitable cookie policy that depends on data

types can lead to a separation of consumers with different privacy concerns into different opt-in choices. As is clear from the above example, the necessary condition for efficient separation is clear information about possible privacy costs, i.e., $c_{i\theta}$, and the expected benefits from opting in to different data types, i.e., γ_θ. Although this is a simple example, a general point can be made that a more refined opt-in choice can overcome the inefficiency in data collection associated with a simple, binary opt-in choice.

4 Conclusion

Australia's Privacy Act dates back to 1988, when the digital economy was still in its infancy. As the digital economy grows at breakneck speed and affects every aspect of our daily lives, consumers' digital privacy has become a pressing issue. The GDPR is a good starting point in protecting consumers' privacy in the digital era. But it appears that not enough consideration was given to the adverse effects of the GDPR on data-driven businesses. Nor is it clear if the intended privacy protection achieves the desired result. In this article, we have discussed how the GDPR can be refined and standardized to be more effective in protecting privacy while not stifling the data-based economy. Our main point is that an ideal data policy should inform consumers about possible privacy costs in a transparent way while giving consumers a menu of opt-in choices into which consumers can self-select themselves. Compared to the GDPR's opt-in requirement, such a policy will be more effective in protecting privacy without leading to undesirable loss of valuable data.

References

[1] Acemoglu, D., Makhdoumi, A., Malekian, A. and A. Ozdaglar (2019). Too much data: Prices and inefficiencies in data markets. NBER Working Paper 26296.

[2] Acquisti, A. and J. Grossklags (2005). Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 26-33.

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

[4] Acquisti, A., Taylor, C. and L. Wagman (2016). The economics of privacy. *Journal of Economic Literature*, 54(2): 442-492.

[5] Aridor, G., Che, Y.-K. and T. Salz (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. NBER Working Paper 26900.

[6] Athey, S., Catalini, C. and C. Tucker (2017). The digital privacy paradox: Small money, small costs, small talk. NBER Working Paper 23488.

[7] Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M. and E. Turner (2019). Americans and privacy: Concerned, confused and feeling lack of control over their personal information. Pew Research Center.

[8] Baron, D. and R. Myerson (1982). Regulating a monopolist with unknown cost. *Econometrica*, 50: 911-930.

[9] Berendt, B., Günther, O. and S. Spiekermann (2005). Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, 48(4): 101–106.

[10] Bergemann, D., Bonatti, A. and T. Gan (2022). The economics of social data. *RAND Journal of Economics*, 53(2): 263-296.

[11] Bian, B., Ma, X. and H. Tang (2022). The supply and demand for data privacy: evidence from mobile apps. Available at https://ssrn.com/abstract=3987541.

[12] Campbell, J, Goldfarb, A. and C. Tucker (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy*, 24(1): 47-73.

[13] Chen, Z., Choe, C. and N. Matsushima (2020). Competitive personalized pricing. *Management Science*, 66(9): 4003-4023.

[14] Choe, C., King. S. and N. Matsushima (2018). Pricing with cookies: behavior-based price discrimination and spatial competition. *Management Science*, 64(12): 5669-5687.

[15] Choe, C., Matsushima, N. and M. J. Tremblay (2022). Behavior-based personalized pricing: When firms can share customer information. *International Journal of Industrial Organization*, 82, 102846.

[16] Choi, J. P., Jeon, D.-S. and B.-C. Kim (2019). Privacy and personal data collection with information externalities. *Journal of Public Economics*, 173: 113-124.

[17] Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F. and T. Holz (2019). We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. *Network and Distributed Systems Security Symposium 2019*.

[18] Ebert, N., Ackermann, K. A. and B. Scheppler (2021). Bolder is better: Raising user awareness through salient and concise privacy notices. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, article no: 67, 1-12.

[19] Fainmesser, I. P., Galeotti, A. and R. Momot (2019). Digital privacy. HEC Paris Research Paper No. MOSI-2019-1351.

[20] Goldfarb, A. and C. E. Tucker (2012). Shifts in privacy concerns. *American Economic Review*, 102(3): 349-353.

[21] Goldfarb, A. and C. E. Tucker (2019). Digital economics. *Journal of Economic Literature*, 57(1): 3-43.

[22] Hagiu, A. and J. Wright (2020). When data creates competitive advantage. *Harvard Business Review*, January-February issue.

[23] Ichihashi, S. (2021). The economics of data externalities. *Journal of Economic Theory*, 196, 105316.

[24] Janßen, R., Kesler, R., Kummer, M. E. and J. Waldfogel (2022). GDPR and the lost generation of innovation apps. NBER Working Paper 30028.

[25] Jia, J., Jin, G. Z. and L. Wagman (2021). The short-run effects of the General Data Protection Regulation on technology venture investment. *Marketing Science*, 40(4): 661-684.

[26] Johnson, E. J., Bellman, S. and G. L. Lohse (2002). Defaults, framing and privacy: Why opting in-opting out. *Marketing Letters*, 13(1): 5-15.

[27] Johnson, G. A., Shriver, S. K. and S. Du (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science*, 39(1): 33-51.

[28] Johnson, G. A., Shriver, S. K. and S. G. Goldberg (2022). Privacy & market concentration: Intended & unintended consequences of the GDPR. Working paper, SSRN: https://ssrn.com/abstract=3477686.

[29] Libert, T., Graves, L., and R.K. Nielsen (2018). Changes in third-party content on European news websites after GDPR. Reuters Institute for the Study of Journalism.

[30] Lin, T. (2022). Valuing intrinsic and instrumental preferences for privacy. *Marketing Science*, 41(4): 235-253.

[31] Liu, Q. and K. Serfes (2006). Consumer information sharing among rival firms. *European Economic Review*, 50(6): 1571-1600.

[32] Machuletz, D. and R. Böhme (2020). Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies*, 481-498.

[33] Matte, C., Bielova, N. and C. Santos (2020). Do cookie banners respect my choice? : Measuring legal compliance of banners from IAB Europe's transparency and consent framework. *IEEE Symposium on Security and Privacy.*

[34] Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory*, 18: 301-317.

[35] Nouwens, M., Liccardi, I., Veale, M., Karger, D. and L. Kagal (2020). Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.*

[36] OAIC (2020). Australian community attitudes to privacy survey 2020. Prepared for the Office of the Australian Information Commissioner by Lonergan Research.

[37] Obar, J. A. and A. Oeldorf-Hirsch (2020). The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1): 128-147.

[38] Peukert, C., Bechtold, S., Batikas, M. and T. Kretschmer (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science*, 41(4): 318-340.

[39] Posner, R. A. (1981). The economics of privacy. *American Economic Review*, 71(2): 405-409.

[40] Rocher, L., Hendrickx, J. M. and Y.-A. de Montjoye (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10:3069. https://doi.org/10.1038/s41467-019-10933-3.

[41] Schmitt, J., Miller, K. M. and B. Skiera (2020). The impact of privacy laws on online user behavior. Working paper, arXiv:2101.11366v2.

[42] Utz, C., Degeling, M., Fahl, S., Schaub, F. and T. Holz (2019). (Un)informed consent: Studying GDPR consent notices in the field. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 973-990.