The Data Privacy Paradox and Digital Demand[*]

Long Chen, Yadong Huang, Shumiao Ouyang, Wei Xiong

March 2022

Abstract

A central issue in privacy governance is understanding how consumers balance their privacy concerns and data sharing to satisfy service demands. We combine survey and behavioral data of a sample of Alipay users to examine how data privacy concerns affect their data sharing with third-party mini-programs on the Alipay platform. We find a positive relationship between the respondents' self-stated privacy concerns and their number of data-sharing authorizations, confirming the puzzling data privacy paradox. Instead of attributing this paradox to the respondents' unreliable survey responses, resignation from active protection of their data privacy, or behavioral factors in making their data-sharing choices, we show that this phenomenon can be explained by a positive relationship between consumers' privacy concerns and demands for digital applications.

[*] Long Chen and Yadong Huang are affiliated with Luohan Academy, and Shumiao Ouyang and Wei Xiong are affiliated with Princeton University. We appreciate comments and suggestions made by Alessandro Acquisti, Cameron Peng, Avanidhar Subrahmanyam, Huan Tang, Liad Wagman, David Yang, Liyan Yang, and seminar participants at American Finance Association Meetings, Boston College, the CBDC Series, CUHK Shenzhen, the IMF, the Kansas City Fed, Peking University, Princeton, SAIF, and UCLA. We are grateful for the help provided by Yong Li and Dongyu Wei at Luohan Academy, and Xiaopeng Wang and Xiaoqiang Yang at Ant Group. This study has received exemption from the Institutional Review Board (IRB) of Princeton University.

Sharing of personal data by consumers empowers the booming digital economy. However, there are growing concerns about data privacy protections across the world, as reflected by the enactments of the General Data Privacy Regulation (GDPR) by the European Union in 2018 and the California Consumer Privacy Act (CCPA) by the state of California in 2020. Despite the importance of data privacy and protections, there are many open questions regarding consumers' data privacy preferences and how their privacy preferences affect their data-sharing choices, as discussed by the Luohan Academy Report of Chen et al. (2021). This lack of knowledge is reflected by the "privacy paradox," a term used by policy makers and commentators to loosely describe a general disconnect between consumers' self-stated privacy preferences and their actual privacy-seeking behavior. As summarized by Acquisti, Brandimarte, and Loewenstein (2020), consumers in a wide range of survey and experimental studies often say they care about privacy but at same time choose to share their personal data either freely or for small rewards.

The presence of this disconnect is often used as evidence to argue either that consumers' privacy concerns are not credible or that privacy is no longer achievable in the age of data economy, motivating a systematic examination of consumers' privacy preferences and data-sharing choices. Does the privacy paradox exist in realistic settings when consumers are faced with choices to share personal data with digital service providers? If so, what causes consumers to ignore their privacy concerns in data sharing? Unless we can understand how consumers trade off their privacy preferences with data-sharing needs to satisfy their service demands, privacy governance will not have a solid foundation.

We aim to address these issues in this study by conducting a survey of Alipay users about their data privacy concerns and then matching their survey responses with rich administrative data about their data-sharing choices on the Alipay platform to analyze how their data-sharing choices are related to their stated privacy concerns. Alipay is a highly popular payment and lifestyle platform with more than 900 million active users in China. In addition to its widely used payment system, it also hosts over two million third-party mini-programs, which are lightweight apps that run inside Alipay to offer a variety of digital services to Alipay users. To use a mini-program, a user must first authorize sharing of certain personal data with the mini-program. The requested data sharing varies across mini-programs from innocuous information, such as nickname, to highly sensitive information, such as the national ID number and Sesame credit score.

In policy discussions, a widely-held view is that the privacy paradox exists because users simply cannot afford not to use popular digital applications. Because mini-programs on Alipay vary substantially in the importance of the provided services and the sensitivity of the requested information, this setting provides an ideal opportunity to study how different users, when given the options, balance their privacy concerns with their demand for digital services. Our rich administrative data allow us to examine each user's data-sharing choices along multiple dimensions (initial authorization and later cancellation) and connect these choices to the user's use of each specific mini-program.

In July 2020, we worked with Alipay to conduct a survey of Alipay users, which included 12 questions about their preferences and concerns regarding data sharing with Alipay's mini-programs. We received survey responses from 14,250 Alipay users. In response to a question that explicitly asked whether they are concerned about their data privacy when sharing personal data with mini-programs, 46% said they are very concerned, 39% are concerned, and only 15% are not concerned. During the 13-month period before the survey from July 2019 to July 2020, the "unconcerned" users on average initially visited 14.3 mini-programs and authorized data sharing with 11.2 of them, the "concerned" users initially visited 15.5 mini-programs and authorized 11.5, and the "very concerned" users initially visited 16.3 mini-programs and authorized 11.3. During the 17-month period after the survey from August 2020 to December 2021, the "unconcerned" users initially visited 27.8 mini-programs and authorized 22.5, the "concerned" users visited 32.8 and authorized 24.6, and the "very concerned" users visited 33.4 and authorized 23.8. To the extent that the "very concerned" users rejected nearly 25% of data-sharing requests, they did not resign from active protection of their data privacy by blindly authorizing all requests.

Even though one would expect users with stronger privacy concerns to be more reluctant to share personal data, "concerned" and "very concerned" users, on average, authorized data sharing with almost the same number of mini-programs as "unconcerned" users in the pre-survey period, and even authorized a greater number in the post-survey period. The lack of difference in the pre-survey period and the greater number of data sharing authorization in the post-survey period hold even after controlling for user characteristics such as digital experience, age, gender, and city, as well as mini-program fixed effects. These patterns are puzzling and confirm the data privacy paradox in a setting that is highly relevant to the digital economy. Our study is immune from the critique of Solove (2021) that the behavior involved in privacy paradox studies involves people making decisions about risk in very specific contexts while their self-reported privacy concerns are much more general in nature. Our survey questions specifically target the respondents' concerns about data sharing with Alipay's mini-programs, and are matched by our administrative data specifically about their data sharing with Alipay's mini-programs.

It is tempting to attribute the data privacy paradox to noisiness and unreliability of survey responses. While survey responses are indeed noisy at the individual level, we find that at the group level, the privacy concerns stated in survey responses are positively associated with the respondents' propensity to take two privacy-seeking actions in Alipay: one is canceling previously authorized data sharing with mini-programs, and the other is changing Alipay's default privacy settings, which tend to make a user's information visible to other Alipay users. These findings thus validate the survey-based measure of privacy concerns.

What causes privacy-concerned Alipay users to ignore their privacy concerns in authorizing data sharing? The privacy literature has suggested a number of psychological and behavioral factors to explain the privacy paradox, including consumers' ignorance about the consequences of data sharing (Pew, 2019), present bias which causes consumers to overweight immediate convenience from using digital applications and underweight future cost of sharing personal data (Acquisti, 2004), and illusion of control which causes consumers to feel more in control when making data-sharing choices (Brandimarte, Acquisti and Loewenstein, 2013). In contrast to the focus of this literature, we find that the data privacy paradox is present beyond users without college education (and thus likely less knowledge about data privacy) and users with weak self-controls. Instead, our analysis uncovers a curious, *positive* correlation between Alipay users' data privacy concerns and digital demands—that is, users with stronger privacy concerns also use their authorized mini-programs more frequently and more extensively. As the greater demands of privacy-concerned users for digital services may offset or even dominate their privacy concerns about sharing personal data with mini-programs, this correlation helps to explain the data privacy paradox.

The positive correlation between privacy concerns and digital demands is a new finding to the literature and connects privacy concerns directly to demands for digital services, albeit in an unexpected way. A common wisdom suggests that privacy concerns would deter users from extensive use of digital services that usually require sharing of personal data, leading to a negative correlation between privacy concerns and digital demands. Instead, our finding suggests that privacy concerns are possibly developed through the process of using digital services. That is, as some users gradually develop enjoyment from using the powerful and convenient services offered by mini-programs, they may also become more concerned about the potential risks from their extensive data sharing with those programs.

To further explore this relationship, we examine a hypothesis that more-active users of mini-programs are more likely to cancel their data-sharing authorizations with mini-programs. One cannot take this hypothesis for granted as it counters our usual intuition that more-active users incur greater costs from canceling a mini-program. To our pleasant surprise, by using two different measures of user activeness and after controlling for various user characteristics and mini-program fixed effects, we find that more-active users of mini-programs in our survey sample are indeed more likely to cancel data sharing with mini-programs in a one-year period. We also take advantage of a salient incident made by Alipay on January 3, 2018, which greatly stimulated Alipay users' awareness of the need to protect their data privacy, to compare the responses of different Alipay users. Interestingly, in response to the incident, heavy users in a representative sample of 100,000 users—randomly drawn from the full set of active Alipay users—are again more likely than light users to cancel data sharing with mini-programs. Taken together, we find evidence in two different samples by using both unconditional and conditional tests to highlight the positive relationship between Alipay users' privacy concerns and digital demands.

As more-active Alipay users were more likely to complete the survey, our sample of the survey respondents is biased toward more-active users. The representative sample of Alipay users also allows us to verify robustness. By using an alternative, behavior-based measure of privacy concerns through users' changes of their Alipay default privacy settings, we find results that are fully consistent with those found from using the survey sample and the survey-based privacy measure.

Our paper adds to the literature on the data privacy paradox, including, Gross and Acquisti (2005), Goldfarb and Tucker (2012), and Athey, Catalini and Tucker (2017). These studies have designed creative surveys and experiments to measure individuals' privacy preferences. See Acquisti, Brandimarte, and Loewenstein (2020) for a recent review of this literature. By combining survey data with extremely extensive administrative data, our study not only confirms the paradox in a highly relevant setting but also uses the paradox as an entry to analyze the nature of data privacy concerns. We uncover data privacy concerns as a preference developed through the use of digital applications, which, to our knowledge, is a new dimension not previously explored by the literature. A question closely related to the data privacy paradox is how much a consumer values her data privacy, as addressed by Acquisti, John and Lowenstein (2013) and Tang (2020). Our analysis of data-sharing choices faced by consumers on a highly popular digital platform shows that the value of data privacy crucially depends on the two sides of data sharing, and, interestingly, is likely to increase over time with the deepening of the data economy.

This finding also adds to the literature on privacy preferences. See Acquisti, Taylor and Wagman (2016) for an extensive review. This literature has pointed to several sources of consumers' privacy concerns. For example, while data sharing allows sellers to better match consumers with their preferred products, it may also expose consumers to potential price discrimination by sellers, e.g., Taylor (2004) and Acquisti and Varian (2005). Data sharing also exposes consumers to greater risk that their personal data might be hacked or leaked (Fainmesser, Galeotti and Momot, 2019). Data sharing may also expose vulnerable consumers to targeted advertising by temptation goods sellers (Liu, Sockin and Xiong, 2020). Our survey also shows supportive evidence for these arguments. More importantly, our paper shows that regardless of the sources of privacy concerns, they are likely to grow with the use of digital applications and the accumulation of personal data shared with digital service providers.

The emerging literature on the data economy has emphasized two important features of data sharing—nonrivalry and increasing returns to scale, e.g., Jones and Tonetti (2020), Farboodi and Veldkamp (2020), and Cong, Xie and Zhang (2020). Empirically, Ouyang (2021) evaluates the value of Alipay data and shows that data sharing can facilitate credit provision, especially to the underserved. Considering the implication of our analysis that consumers' privacy concerns may grow with their data accumulated with digital service providers, consumers may become more restrictive with their data sharing over time, preventing the economy from realizing the full promise of data sharing and thus making privacy protection even more important. This importance has motivated a growing body of literature to empirically examine the impact of data privacy regulations, e.g., Goldberg, Johnson and Shriver (2019) and Aridor, Che and Salz (2020). It has also motivated innovative designs of decentralized digital platforms that are based on cryptographic technologies to prevent digital platforms' potential abuse of their control of extensive consumer data, as argued by Sockin and Xiong (2022).

The paper is organized as follows. Section I provides some institutional background about the data-sharing arrangement between users and mini-programs in Alipay. Section II describes the survey of Alipay users and reports some summary statistics of the data used in our analysis. We analyze the data privacy paradox in Section III and further examine the relationship between Alipay users' privacy concerns and demands for the digital services provided by the mini-programs in Sections IV and V. We conclude in Section VI.

I. Institutional Background

As this paper studies data sharing of Alipay users with third-party mini-programs on the Alipay platform, this section provides some background information about the Alipay platform and the data-sharing arrangement between users and mini-programs in Alipay.

Alipay is a mobile application, owned by Ant Group, which has grown from offering online payment services into the world's largest payment and lifestyle platform. Alipay has more than 900 million active users in China, which is more than 70% of the population. In addition to providing a wide range of financial services, such as digital payments, micro-loans, credit cards, insurance, and wealth management, Alipay is also an ecosystem that enables third parties to offer mini-programs inside Alipay. These mini-programs are "subapplications" within the Alipay application that provide users with advanced and extensive digital services, such as bike-sharing, on-demand logistics, and food ordering, without requiring users to download or install separate applications. By June 2020, over two million mini-programs had emerged on Alipay. The number of mini-program users has increased from 21% of Alipay users in 2015Q4 to 49% in 2019Q2 (Chen et al., 2021).

To use a mini-program in Alipay, users must authorize sharing of certain personal data with the mini-program. When a user first visits the mini-program, the mini-program will ask the user to authorize sharing of certain information necessary for its service, including but not limited to nickname, gender, phone number, national ID number, and Sesame credit score. The requested information varies across mini-programs. Some information is innocuous, such as a nickname, while other information is more sensitive, such as one's national ID number and Sesame credit score. A user has two possible choices, either agree to or reject the data-sharing request. Only after the user authorizes the request, is she allowed to use the service offered by the mini-program. This setting makes the data-sharing authorization an explicit exchange of the user's data for the mini-program's service. This data-sharing authorization lasts for a certain time period; at the expiration of that time, the mini-program will ask the user to authorize the data sharing again at her next entry to the mini-program. After a user authorizes data sharing with a mini-program, the user also has the option to cancel the data-sharing authorization at any time before the end of the authorization period. We will examine both the authorization and cancellation decisions through a sample of Alipay users in our study.

For example, Hellobike is a widely used mini-program that offers a bike-sharing service. Users can access Hellobike through either the separate Hellobike application or the Hellobike mini-program inside the Alipay application. There were over 230 million registered users of Hellobike in mid-2019 from the Hellobike application and from mini-programs inside other applications. The Hellobike mini-program in Alipay requests three types of information at a user's initial visit: 1) basic information, such as nickname, profile picture, gender, and location; 2) Sesame credit score, which helps to evaluate trustworthiness of the user and determine whether to require a deposit; and 3) identification information, such as real name, phone number, and national ID number. After the user authorizes the sharing of the requested information, the user can use Hellobike's shared bikes.[1]

Also relevant to our study are Alipay's default settings for each user's data sharing with other users; these settings allow users to take advantage of Alipay's social media functions. Alipay allows each user to choose from a variety of privacy settings, such as whether to show one's real name to friends in Alipay, whether to make ten recent posts visible to strangers, whether to allow connection without permission, and whether to be searchable by phone number. These settings enable users to personalize privacy preferences. The default privacy settings tend to make users

[1] Figure A1 in the Online Appendix provides three additional examples to illustrate the variety of data-sharing requests by mini-programs in Alipay. The first one is a mini-program that searches for part-time jobs. It requests the user to share a mobile number. The second one relates to social connections and requires users to share their nickname, profile, gender, and location. The third one provides legal consulting services and requires sharing of the user's location.

visible and easy to connect with. Some users have chosen to change the default settings, which is an action that reflects privacy concerns about revealing their information to other Alipay users. In our analysis, we use changing the default privacy settings as a behavior-based measure of a user's privacy concerns as an alternative to our main survey-based measure. This measure is appealing because these default settings are not directly related to services provided by mini-programs. On the other hand, changing the default settings requires a user to have the necessary knowledge about how to do so in the Alipay app. Thus, we also need to control for the user's knowledge and digital experience in using this measure.

II. Survey and Administrative Data

We conducted a survey of Alipay users about their privacy preferences and then combined the survey responses with the respondents' administrative data inside the Alipay application to study how their stated privacy preferences are related to actual data-sharing choices. In this section, we first describe the survey and then report summary statistics of data-sharing authorizations and other administrative data of the survey respondents as well as a representative sample of Alipay users for comparison.

A. The Survey

In July 2020, we worked with Alipay to conduct a survey of Alipay users. The survey consisted of 12 questions about Alipay users' preferences regarding data sharing with third-party mini-programs in Alipay. The survey was distributed through the message box at the center of the front page of the Alipay application, a highly visible channel,[2] to a random sample of 2.5 million active Alipay users. In total, 27,597 users opened the survey link and 14,250 completed the survey. In the middle of the survey, there is a question: "*Have you ever used mini-programs in Alipay?*" Only those respondents who answered *yes* to this question would advance to see the rest of the survey questions specifically related to privacy concerns about data sharing with mini-programs. In the collected survey responses, 10,875 respondents indicated that they had used mini-programs in

[2] See Figure A2 in the Online Appendix for a picture of the Alipay front page, which highlights the distribution channel for the survey.

Alipay, accounting for 76% of all respondents.[3] These 10,875 respondents are the main sample for our analysis.

Due to the natural tendency that more-active users are more likely to pay attention to the message box in the Alipay application and thus to open the survey link, this sample of survey respondents is representative of more-active Alipay users rather than the whole population of Alipay users. To focus on the data privacy paradox, a phenomenon that is revealed by survey studies, we use this sample of survey respondents as the main sample of our analysis. For robustness and comparison, we have also examined a representative sample of 100,000 Alipay users, who were randomly drawn from the whole population of Alipay users.

The survey was in Chinese; we provide an English translation of the survey questions in the Appendix. Table 1 summarizes the responses to seven of the questions in the survey. In response to a general question "*Are you concerned about privacy issues while using digital services?*" 93% of the respondents were very concerned, 6% were concerned, and only 1% were not concerned. The very high percentage of respondents either very concerned or concerned with data privacy is consistent with other surveys regarding general privacy attitudes, which also find strong concerns about data privacy. [4]

In response to a question specific to data sharing with mini-programs in Alipay, "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*" 46% of the respondents were very concerned, 39% were concerned, and 15% were not concerned. Relative to the earlier question about general concerns about data privacy, the respondents were less concerned by data sharing with mini-programs in Alipay. The large difference between the responses to these two questions confirms a concern raised by Solove (2021) about the importance of closely matching consumers' privacy concerns with their data-sharing choices in analyzing the data privacy paradox. As this latter question is directly related to our analysis of data sharing with

[3] Figures A3–A6 in the Online Appendix provide some characteristics of the survey respondents. It took most respondents more than sixty seconds to complete the survey, indicating that they answered the questions in a serious way (Figure A3). The geographical distribution of the respondents across the provinces in China lines up well with the distribution of the population (see Figure A5), except that the share of respondents from the most populated Guangdong province is about 17%, substantially higher than its population share of about 8.2%.

[4] See Special Eurobarometer 431 (2015), available at https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf; Pew Research (2015), available at https://www.pewresearch.org/internet/2015/05/20/americans-attitudes-about-privacy-securityand-surveillance/; The Chinese Consumer Association (2018); Global Privacy Enforcement Network (2018).

mini-programs, we will use the respondents' answers to this question as a key measure of their privacy concerns in our later analysis. Specifically, we will compare the data-sharing authorizations among respondents with different levels of privacy concerns about data sharing with mini-programs.

We also asked the respondents this specific question: "*What privacy issues are you concerned about when using mini-programs in Alipay?*" This question allowed each respondent to select more than one option from a list of four, including: 1) data leakage and security, 2) price discrimination by merchants, 3) seductive advertising and temptation consumption, and 4) others. The first choice represents potential concerns about insufficient protections provided by mini-programs to secure user data and prevent hacking and other data leakage, as modeled by Fainmesser, Galeotti and Momot (2019). The second choice represents a concern that extensive data sharing by consumers may allow merchants to infer consumers' reservation prices and thus employ price discrimination. There is a large body of economics literature analyzing this concern in the digital economy, as reviewed by Acquisti, Taylor and Wagman (2016), Bergemann and Morris (2019), and Goldfarb and Tucker (2019). The third choice represents a new concern that in the booming digital economy, extensive data sharing by consumers may expose consumers' personal weaknesses, such as a lack of self-control, to online advertisers and sellers, as recently emphasized by Liu, Sockin and Xiong (2020). Interestingly, 86% of the respondents selected data leakage and security, 49% selected seductive advertising and temptation consumption, and 21% selected price discrimination by merchants. To the extent that only 5% of the respondents selected "others," the first three concerns well captured the main privacy concerns of the respondents.

In response to two related questions "*Do you know how to change privacy settings in Alipay?*" and "*Have you ever changed your privacy settings in Alipay?*" 60% of the respondents indicated they knew how to change privacy settings, and 39% of the respondents say they had changed their privacy settings. We will use changing Alipay's default privacy settings as a behavior-based measure of privacy concerns for users in our representative random sample.

B. Administrative Data

A key strength of our study is that we are able to link the survey responses with the respondents' extensive administrative data inside the Alipay application, which allow us to examine how respondents' privacy preferences are related to their actual data-sharing choices and use of the authorized mini-programs. For each Alipay user in our sample, we have access to three sets of administrative data in Alipay: general information, information related to data-sharing authorizations and cancellations with mini-programs, and the use of mini-programs. Table 2 reports summary statistics of these three sets of variables for the survey sample and the random sample, both of which are used in our analysis.

The Survey Sample

Table 2 reports summary statistics of key variables for our survey sample. Panel A covers three sets of user information: general profile, data sharing with mini-programs, and monthly use of mini-programs. For the general information, also known as user profile, we have access to information on the gender, age, and city of each user. We also include their digital experience, which is measured by the number of months since a user first registered on Alipay. The average user age is 32.82 years and the average digital experience is 74.97 months. We also construct dummy variables to measure a respondent's privacy concerns based on the answer to the following survey question: "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*" The possible responses were "*not concerned,*" "*concerned,*" or "*very concerned.*" We define the *Concerned Dummy* variable as 1 if the answer was "*concerned,*" and 0 otherwise; we define the *Very Concerned Dummy* variable as 1 if the answer was "*very concerned*", and 0 otherwise.

The information on data sharing with mini-programs consists of five variables at the user level. The first two variables measure how users share their data with mini-programs over the period from July 2019 to December 2021, which covers the time of the survey in July 2020. First, we count the number of initial visits by a user to mini-programs; this is when a data-sharing request pops up. Second, we count how many times the user authorizes the data-sharing requests. The other three variables measure a user's cancellations of previously authorized data sharing with mini-programs. As we mentioned earlier, an Alipay user can actively terminate data sharing with a mini-program at any time. We define a dummy variable, *has canceled*, which takes a value of 1 if the user has ever canceled data sharing with at least one mini-program during the measurement period of January 2013 to July 2020 (a seven-year period before the survey), and 0 otherwise. The measure # *Cancellations* is defined as the number of active mini-programs that a user canceled between January 2013 to July 2020. We count a mini-program as active if the user has used it at least once, which implies an outstanding data-sharing authorization by the user. The *Cancellation Rate* is the number of canceled authorizations from January 2013 to July 2020 divided by the total number of active mini-programs.

In our survey sample, a respondent, on average, initially visited 46.57 mini-programs with a standard deviation of 55.45 and a maximum of 1609 from July 2019 to December 2021. The number of data-sharing authorizations has a mean of 34.22, a standard deviation of 22.78, and a maximum value of 422. From January 2013 to July 2020, 48% of the respondents canceled at least one data-sharing authorization. Despite that almost half of the respondents actively canceled data sharing, the average number of cancellations is 2.66, and the average cancellation rate is 0.05. This low cancellation rate shows that it is still relatively infrequent for Alipay users to cancel data-sharing authorizations.

The information on mini-program use includes monthly use of each pair of user and mini-program (user × mini-program × month level) from July 2019 to July 2020.[5] It comprises four variables: 1) the number of active days; 2) the number of sessions; 3) the number of launches; and 4) the number of page visits. These variables are different from each other by construction. A user might use a mini-program for several sessions in a day. In each session, she might launch the mini-program multiple times. In each launch, she might visit several pages inside the mini-program. We find that, on average, in each month, a user in our survey sample is active in a mini-program on 0.57 days, with 0.81 sessions, 2.29 launches, and 5.20 pageviews.

Panel B of Table 2 further compares three groups of users: "unconcerned", "concerned", and "very concerned", sorted by their responses to the survey question "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*". Even though there is not any significant difference in age, "concerned" and "very concerned" users have longer digital experience, are more likely female, and are more likely to have a college degree or above.

The Random Sample

[5] Alipay did not systematically record data on users' activities related to mini-programs before 2019. As a result, we cannot expand all of these variables to cover the period since 2013.

For comparison, we have also constructed a random sample of 100,000 Alipay users, who we randomly selected from all active Alipay users. We report their summary statistics in Table 3. The users in this random sample have an average age of 36.6 years and an average digital experience of 60.7 months, confirming that our main survey sample tends to be younger people with longer digital experience. During the period between July 2019 and July 2020, users in the random sample initially visited, on average, 3.02 mini-programs and authorized data sharing with 2.4 of them. Furthermore, in each month, a user in the random sample was active in a mini-program on 0.27 days, with 0.34 sessions, 1.10 launces, and 3.06 pageviews. As expected, the survey sample indeed covers more-active users than the random sample, as reflected by their greater number of data-sharing authorizations with mini-programs and more extensive use of these mini-programs. Despite the difference in these samples, as we will show, the main results of our analysis are robust across these samples.

As the users in this random sample did not participate in our survey, we cannot use their survey responses to measure their privacy concerns about data sharing with mini-programs. Instead, we use whether a user has changed Alipay's default privacy settings as a behavior-based measure of privacy concerns. Gross and Acquisti (2005) have used whether a Facebook user changes the default data-sharing settings in Facebook as a key indicator of the user's privacy concerns. Note that this measure is not directly comparable with the survey-based measure of privacy concerns because the behavior-based privacy concerns are specifically related to sharing personal data with other users in Alipay, while the survey-based privacy concerns are specifically related to sharing data with mini-programs.

Furthermore, as discussed by Liu et al. (2022) in a study of stock trading motives, behavior-based measures also face another complication in that they are often related to multiple factors, beyond the particular preference or bias that a behavior-based measure is intended to capture. In our context, as some users may not know how to change Alipay's privacy settings, the behavior-based privacy concern measure is also affected by a user's knowledge of and familiarity with the Alipay application. In the survey sample, 49% of respondents had changed their Alipay privacy settings, while only 9% of the random sample had done so. This contrast is likely because users in the random sample tend to be less active and most of them may not know how to change the privacy settings. Nevertheless, this alternative, behavior-based measure of privacy concerns allows us to examine how privacy concerns are related to the users' data-sharing choices in the random sample, after controlling for the users' digital experience and knowledge.

III. The Data Privacy Paradox

By combining the survey responses of Alipay users with their administrative data in Alipay, we can directly examine how their data-sharing choices are related to their privacy concerns. Specifically, we examine whether users with stronger privacy concerns are more reluctant to share personal data with mini-programs. In this section, we first describe a simple conceptual framework to anchor our analysis and then present some empirical results, which confirm the data privacy paradox. We also validate the survey-based measure of privacy concerns and discuss potential explanations of the data privacy paradox indicated by the respondents in the survey.

A. Conceptual Framework

To decide whether to share the requested personal data with a mini-program, an Alipay user needs to compare the benefits from using the mini-program with the privacy costs of sharing the requested data. Both the benefits and the costs may depend on both the user and the mini-program. For simplicity, we suppose that the cost for user i to share personal data with mini-program j, c_{ij}, can be linearly decomposed as

$$c_{ij} = c_i + c_j + \epsilon_{ij},$$

where c_i represents the user's privacy concerns, c_j measures the sensitivity of the data requested by the mini-program and its privacy protection practice, and ϵ_{ij} is a noise component independent across the user and mini-program pair. By imposing this form, we assume that there is no privacy cost specific to the user and mini-program pair. Similarly, we assume that the benefit for the user to use the mini-program, b_{ij}, can be linearly decomposed as

$$b_{ij} = b_i + b_j + \varepsilon_{ij},$$

where b_i is the user component, b_j is the mini-program component, and ε_{ij} is a noise component, which is independent across the user and mini-program pair.

The user chooses to authorize the data sharing if the benefit is greater than the cost:

$$b_{ij} - c_{ij} = b_i - c_i + b_j - c_j + \varepsilon_{ij} - \epsilon_{ij} > 0.$$

This condition is driven by the characteristics of the user and the mini-program. After controlling for the mini-program's characteristics, the authorization choice is driven by the user's characteristics through the term $b_i - c_i$. If b_i and c_i are independent, a user with stronger privacy concerns (i.e., larger c_i) is less likely to authorize data sharing, while a user with a greater benefit b_i is more likely to authorize it. This implies the following hypothesis:

Hypothesis 1: Everything else being equal, privacy-concerned users are more reluctant to authorize data sharing with mini-programs.

This hypothesis is consistent with the common wisdom reflected by the discussions of the data privacy paradox. We will start our empirical analysis by testing this hypothesis. Alternatively, the benefit b_i and the privacy concern c_i may be positively correlated across users. If so, the users' data-sharing choices are not necessarily sensitive to their privacy concerns. We will also examine this possibility in our later analysis.

B. Privacy Concerns and Data Sharing

We now compare in Figure 1 the number of data-sharing authorizations by Alipay users who have expressed different levels of concern about data sharing in their responses to the survey question "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*". Panel A shows that during the pre-survey period of July 2019 to July 2020, "unconcerned" users on average initially visited 14.3 mini-programs and authorized data sharing with 11.2 of them, "concerned" users visited 15.5 mini-programs and authorized 11.5, and "very concerned" users visited 16.3 mini-programs and authorized 11.3. There is an interesting pattern that "concerned" and "very concerned" users tend to open more new mini-programs than "unconcerned" users and eventually authorize data sharing with almost the same number of mini-programs. The pattern becomes even more striking in the post-survey period from August 2020 to December 2021. Panel B shows that during the post-survey period, "unconcerned" users initially visited 27.8 mini-programs and authorized 22.5, "concerned" users visited 32.8 and authorized 24.6, while "very concerned" users visited 33.4 and authorized 23.8. There is a clear trend that users across all groups visited and authorized more mini-programs in the post-survey period than in the pre-survey period, even after adjusting for the slightly longer post-survey period. More surprisingly, “concerned” and “very concerned” users authorized even more data sharing than “unconcerned” users in the post-survey period. These patterns in pre- and post-survey periods both contradict Hypothesis 1 that privacy-concerned users are more reluctant to authorize data sharing.

As users differ not only in their privacy concerns but also in other dimensions, we first adopt a cross-sectional regression at the user level to control for various user characteristics:

$$\begin{aligned} Y_i = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i \\ +a_4\, Digital\ Experience_i + \delta_i + \epsilon_i \end{aligned} \tag{1}$$

where the dependent variable Y_i is a measure of certain behavior (either the number of data-sharing authorizations or initial visits to mini-programs) by user i; the dummy variable $Concerned_i$ is defined to be 1 if user i answers “concerned” to the question about sharing data with mini-programs in the survey, and 0 otherwise; the dummy variable $Very\ Concerned_i$ is defined to be 1 if user i answers “very concerned” in the corresponding question, and zero otherwise; Age_i and $Digital\ Experience_i$ are two control variables; and δ_i represents fixed effects related to other user characteristics, including gender and city. Without including the controls, the sample size is 10,875. As the characteristics of some users are missing, including the control variables slightly reduces the sample size to 10,858.

Table 4 reports the regression results. Panel A uses the pre-survey sample from July 2019 to July 2020, while Panel B uses the post-survey sample from August 2020 to December 2021. In Panel A, columns (1) and (2) show that the estimates of a_1 and a_2 are both insignificant, with or without the controls, confirming that “concerned” and “very concerned” users do not authorize data sharing with fewer mini-programs than “unconcerned” users in the pre-survey sample. Furthermore, columns (3) and (4) show that the level of privacy concerns is positively correlated with the number of initially visited mini-programs, even though it is uncorrelated with the number of data-sharing authorizations. Specifically, privacy-concerned users, on average, initially visit 1.24 more mini-programs, and “very concerned” users, on average, have 1.97 more initial visits; the coefficients are both highly significant.

In Panel B, column (1) shows that the estimates of a_1 and a_2 are both positive and significant without the controls, while column (2) shows that a_1 and a_2 remain positive, albeit a_2 becoming insignificant, after including the controls. These results confirm that in the post-survey period "concerned" and "very concerned" users authorize more, rather than less, data sharing with mini-programs than unconcerned users.

As highlighted by our conceptual framework, a user's data-sharing authorization with a mini-program may also depend on the services offered and the information requested by the mini-program. To control for mini-program characteristics, we further expand our regression analysis to the user-mini–program level for all possible pairs of users and mini-programs in our sample:

$$Y_{ij} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + \delta_i + \gamma_j + \epsilon_{ij} \qquad (2)$$

For every possible pair of user i and mini-program j, the dependent variable Y_{ij} equals 1 if the user authorizes data sharing with or initially visits the mini-program, and 0 otherwise. Like the user-level regression specified in Equation (1), Age_i, $Digital\ Experience_i$, and δ_i represent controls for user characteristics. We also add γ_j as mini-program fixed effects, which control for the possible heterogeneity in the services offered and information requested by mini-programs.

Table 5 reports the user-mini–program level analysis, with Panel A for the pre-survey sample and Panel B for the post-survey sample. Even after controlling for mini-program fixed effects, the results are very similar to that from the user-level analysis. Panel A shows that in the pre-survey sample, without and with the controls for user and mini-program characteristics, there is no significant difference in the number of data-sharing authorizations across "concerned," "very concerned," and "unconcerned" users, even though the level of privacy concerns is positively correlated with the propensity to have an initial visit to a mini-program. Panel B shows that in the post-survey sample, "concerned" and "very concerned" users authorize more, rather than less, data sharing with mini-programs even after controlling for user and mini-program characteristics.

Overall, Tables 4 and 5 reject Hypothesis 1 and instead confirm the data privacy paradox—that there is no relationship between the level of privacy concerns and the number of data-sharing authorizations in the pre-survey period and a positive relationship in the post-survey period. These findings contradict the common wisdom that privacy-concerned users should be more reluctant to share personal data.

In Table 6, we further explore how the data privacy paradox may vary across users with different characteristics. We expand the user-mini–program level regression specified in Equation (2) by interacting the dummy variables $Concerned_i$ and $Very\ Concerned_i$ with other user characteristics. We focus on two characteristics: education and self-control. We define $Education_i$ as a dummy variable that indicates whether a user has a college degree or above. Interestingly, the interaction terms of $Education_i$ with $Concerned_i$ and $Very\ Concerned_i$ are both significantly positively in column (1) for the pre-survey sample, and are positive, albeit insignificant, in column (3) for the post-survey sample. These results suggest that the data privacy paradox is not simply a phenomenon among users with low education and thus insufficient knowledge of data privacy. To the contrary, it is more severe among more educated users.

We measure $Self\ Control_i$ by whether a user's opt-in rate of seemingly addictive mini-programs is higher than the opt-in rate of other mini-programs in the pre-survey period. The interaction terms of $Self\ Control_i$ with $Concerned_i$ and $Very\ Concerned_i$ are both negative, albeit insignificant, in column (2) for the pre-survey sample. In column (4) for the post-survey sample, the interaction term of $Self\ Control_i$ with $Concerned_i$ is significantly negative, while the interaction term of $Self\ Control_i$ with $Very\ Concerned_i$ is negative, albeit insignificant. Taken together, there is no evidence for the data privacy paradox being more severe among users with weaker self-controls. Thus, the data privacy paradox is a general phenomenon beyond a particular group with behavioral weaknesses.

C. Validating Survey-Based Privacy Concerns

It is tempting to argue that the data privacy paradox may simply reflect unreliability of survey responses. That is, the survey responses may not truthfully or reliably reflect the respondents' privacy preferences. This is a common concern about survey-based measures (see, e.g., Bertrand and Mullainathan 2001). This argument also reflects a widely held suspicion that consumers may not truly be concerned about their data privacy despite the commonly documented privacy concerns in surveys of individuals across the world.

To validate the survey-based measure of privacy concerns, we take advantage of our extensive administrative data about the survey respondents to examine whether the survey-based measure of privacy concerns is positively correlated with actions taken by users to protect their data privacy other than the initial authorization of data sharing with mini-programs. We can observe two such actions: canceling previously authorized data sharing with mini-programs and changing Alipay's default privacy settings. Conceptually, we expect a more privacy-concerned user to be more likely to cancel data sharing and change the default privacy settings.

We again organize our analysis at both the use level and user-mini–program level. For the user-level analysis, we adopt the regression specified in Equation (1) but replace the dependent variable by a dummy variable that indicates whether a user has ever canceled any data-sharing authorization in the period of January 2013 to July 2020 or whether the user ever changed Alipay's default privacy settings between May 2017 and April 2020. Note that both actions require the user to not only have privacy concerns but to have the knowledge necessary to cancel a data-sharing authorization or to change Alipay's default privacy settings. As shown by Table 1, only 60% of the respondents in our survey sample indicated that they knew how to change the default privacy settings in Alipay. We include in the regression extensive controls, including the user's digital experience and age, as well as city and gender fixed effects. These variables serve to control for the user's digital knowledge.

Panel A of Table 7 reports the results from the user-level regressions. In columns (1)–(2), the dependent variable is the *Has Canceled* dummy. All else being equal, the respondents who indicated they are "very concerned" or "concerned" about data sharing with mini-programs have a significantly higher probability of having canceled data sharing with at least one mini-program than "unconcerned" respondents under different regression specifications, with or without including digital experience and age as control variables and including gender and city fixed effects. Furthermore, the probability of having canceled data sharing is also higher in the "very concerned" group than in the "concerned" group.

In columns (3)–(4), the dependent variable is the dummy for *Privacy Setting Changed*. Without including the controls, the respondents who indicate they are "very concerned" or "concerned" about data sharing with mini-programs have a higher probability of having changed their Alipay default privacy settings than "unconcerned" respondents. Interestingly, column (4) shows that this higher probability remains highly significant among "very concerned" respondents, albeit not among "concerned" respondents after including the extensive controls.

Furthermore, across both cancellation of data sharing in column (2) and change of default privacy settings in column (4), the probability of taking these protective actions significantly increases with digital experience and decreases with age, consistent with a knowledge effect that more-experienced users and younger users are more likely to have the knowledge necessary to take these actions to protect their data privacy. These results thus confirm that digital experience and age are useful controls for digital knowledge in these user-level regressions.

In Panel B of Table 7, we further expand the analysis to the user-mini–program level for cancellation of data sharing. The advantage of the user-mini–program level analysis is that we can control for mini-program fixed effects, which allow us to compare the propensity to cancel data sharing with the same mini-program by users with different privacy concerns. We adopt the regression specification in Equation (2) for the sample of all existing data-sharing authorizations between any pair of user and mini-program during the July 2019 to July 2020 period. The sample size is 481,143. The dependent variable is a dummy that equals 1 if the user ever canceled the data-sharing authorization, and 0 otherwise. The coefficients of *Concerned* and *Very Concerned* measure the greater propensity of "concerned" and "very concerned" respondents, respectively, to cancel an existing data authorization. We find that the coefficient is especially large and significant for "very concerned" users. Thus, Panel B again confirms that users who are "very concerned" about data privacy are more likely to cancel data sharing with a given mini-program than "unconcerned" users.

Overall, while survey responses are noisy at the individual level, Table 7 reports regression results at both the user level and user-mini–program level to confirm that the survey-based measure of privacy concerns is positively related to actions taken by Alipay users to protect their own data privacy, thus validating the survey-based measure of privacy concerns at the group level. This finding also confirms the recent studies of Liu et al. (2022) and Giglio et al. (2020), which show that survey responses about trading motives and expectations are consistent with stock investment behaviors.

D. What Determines Data Sharing Authorizations?

In the survey, we asked the respondents whether they agreed with each of the following five statements, which were motivated by public and policy discussions of consumers' data sharing:

1. *I agree to authorize data sharing with mini-programs since it is safe in Alipay.*
2. *I agree to authorize data sharing with mini-programs since my information has already been shared in many platforms.*
3. *I have to share my personal data in exchange for digital services even though I am concerned by my data privacy.*
4. *I only authorize data sharing with a mini-program only when the requested information is not important.*
5. *I tend to authorize data sharing with mini-programs that are used by my friends.*

The first statement considers that users' trust of Alipay's privacy protection might dominate their decisions about privacy concerns. Interestingly, as shown earlier in Table 1, 48% of the respondents in our survey sample regarded Alipay's privacy protection as "very good." The second statement is motivated by the concern that users' extensive data sharing with many digital platforms might substantially reduce the marginal concern of sharing data with another mini-program. This statement is particularly relevant for heavy users of digital applications, who need to share their personal data with many digital service providers. To some extent, this statement reflects a general argument that privacy might be impossible under the attack of increasingly powerful digital technologies in the age of data economy.

The third statement represents a key consideration for our analysis that the decision to authorizing data sharing with a mini-program involves a trade-off between the benefits from using the services and the privacy costs of sharing the requested personal data. The fourth statement addresses the concern that users might be ignorant about the consequences of sharing the requested personal data with mini-programs and such ignorance might influence their data-sharing authorizations. Finally, the fifth statement considered whether social influence, an important mechanism in the digital economy, might induce herding behavior among privacy-concerned users and lead them to authorize data sharing, e.g., Acquisti, Brandimarte, and Loewenstein (2020).

Each of these statements present a potential mechanism that helps Alipay users overcome their privacy concerns when asked to authorize data sharing with mini-programs. For a statement to explain the lack of any difference in the observed data-sharing authorizations between privacy-concerned and unconcerned users, we expect the statement to be more agreeable for "concerned" users than for "unconcerned" users.

Table 8 summarizes the responses to these statements. We split the respondents into two groups, one with "concerned" and "very concerned" respondents and the other with "unconcerned" respondents. Panel A reports the percentage of the respondents in each group that agree and disagree with each of the five statements. In response to the first statement, 80% of "unconcerned" respondents agree, while only 42% of "concerned" or "very concerned" respondents agree. That is, "concerned" or "very concerned" respondents are less likely to agree with Alipay being safe than "unconcerned" respondents. As such, one cannot attribute the similar number of data-sharing authorizations by these two groups to the greater confidence of "concerned" and "very concerned" respondents in Alipay's privacy protection.

The panel also shows that only 12% of "concerned" or "very concerned" respondents and 30% of "unconcerned" respondents agree with the second statement that they choose to authorize data sharing with mini-programs because their information has already been shared in many platforms. These low fractions of endorsement indicate that these respondents are not yet frustrated with the challenges in protecting their data privacy. The lower fraction of endorsement by "concerned" or "very concerned" respondents than "unconcerned" respondents also invalidates this statement as a possible explanation for the data privacy paradox.

Similarly, neither ignorance about the consequences of data sharing nor social influence is an likely explanation. Panel A of Table 5 shows that only 20% of "concerned" or "very concerned" respondents and 30% of "unconcerned" respondents agree with the fourth statement that they choose to authorize data sharing with mini-programs when the requested data are unimportant. Furthermore, 44% of "concerned" or "very concerned" respondents and 58% of "unconcerned" respondents agree with the fifth statement that they choose to authorize data sharing with mini-programs that are used by their friends.

The only exception is statement 3: "*I have to share my personal data in exchange for digital services even though I am concerned by my data privacy*." About 64% of "concerned" or "very concerned" respondents agree with this statement, higher than the 55% of "unconcerned" respondents. This difference indicates that the data privacy paradox might be driven by a trade-off between the costs and benefits of data sharing.

Panel B of Table 8 further examines the relationship between the respondents' agreement with each of these statements and their privacy concerns in a regression with digital experience and age as control variables, along with gender and city fixed effects. The regression results further confirm the summary statistics in Panel A. In particular, even after including the control variables and the user-characteristics fixed effects, “concerned” or “very concerned” respondents are 8.9% more likely than unconcerned respondents to agree with statement 3, and this difference is highly significant.

Taken together, the responses from the survey point to a trade-off between the costs and benefits of data sharing as a possible explanation for the puzzling data privacy paradox.

IV. Digital Demands

According to the conceptual framework presented in Section III.A, it is possible to use the trade-off between costs and benefits to explain the data privacy paradox if the privacy concerns of sharing personal data with a mini-program are positively correlated with the benefits from using it.[6] In this section, we examine how privacy concerns are related to digital demands.

A. Privacy Concerns and Digital Demands

We first analyze the relationship between the respondents’ privacy concerns and demands for digital services provided by mini-programs. As it is difficult to directly measure digital demands, we use the respondents’ actual use of the mini-programs they authorize in Alipay as a proxy, motivated by an intuitive argument that a user with greater demand for digital services is likely to more extensively use their authorized mini-programs. A common wisdom suggests that privacy concerns may deter users from digital applications and thus motivates the following hypothesis:

Hypothesis 2: Everything else being equal, privacy-concerned users use their authorized mini-programs less intensively.

We examine this hypothesis by using the following regression specification:

$$\mathrm{Y_{ijt}} = \mathrm{a}_1\, Concerned_i + \mathrm{a}_2\, Very\ Concerned_i + \mathrm{a}_3\, Age_{it} + \ \mathrm{a}_4\, Digital\ Experience_{it}$$

[6] While our analysis focuses on the relationship between digital demands and privacy concerns, another possible explanation to the observed data privacy paradox is present bias. As suggested by Acquisti (2004), present bias causes consumers to overweight the benefit in the present and underweight the privacy cost in the future. Such present bias may provide an orthogonal mechanism to operate in parallel to the mechanism highlighted by our analysis.

$$+\delta_{\mathrm{i}} + \mu_{\mathrm{j}} + \theta_t + \varepsilon_{\mathrm{ijt}}, \qquad (3)$$

where $\mathrm{Y_{ijt}}$ is a measure of user i's use of mini-program j in month t; the dummy variables $Concerned_i$ and $Very\ Concerned_i$ are defined as before; $\mathrm{Age_{it}}$ and $\mathrm{Digital\ Experience_{it}}$ are two control variables; and δ_{i}, μ_{j}, and θ_t represent fixed effects related to user characteristics, mini-program, and time, respectively. This regression allows us to compare the use of the same mini-program in the same month by respondents with different levels of privacy concerns.

Table 9 reports regression results from using four different measures of a respondent's use of a mini-program in a month: the number of active days, the number of sessions, the number of launches, and the number of visited pages. Column (1) shows that without including the controls, a user "unconcerned" about privacy, on average, uses a mini-program on 0.468 days in a month, while a user "concerned" about privacy uses it on 0.102 more days per month than "unconcerned" users, and a "very concerned" user uses it on 0.126 more days per month than an "unconcerned" user, which represents a gap of 27% between "very concerned" and "unconcerned" users. After including the controls in column (2), the difference between "concerned" and "unconcerned" users remain positive and significant, and "very concerned" users also use the applications more than "concerned" users. The results from the other three measures show the same monotonic pattern—users with strong privacy concerns tend to use their authorized mini-program more frequently and more extensively. Taken together, the regression results show a positive and robust relationship between digital demands and privacy concerns, firmly rejecting Hypothesis 2.

How can privacy-concerned users have greater demands for digital services? This question may appear puzzling because we tend to think of privacy concerns as an innate preference that is independent of an individual's consumption and demand. If privacy concerns are like risk aversion, it is difficult to perceive that individuals with strong privacy concerns will become intensive users of digital services, similar to the logic that investors with greater risk aversion cannot have more risky investments. However, as the digital economy is new and still undergoing rapid developments, many consumers are still in the process of learning about their own demands for digital services and concerns about data privacy. It is possible that during this learning process, consumers gradually develop greater demands for digital services and stronger concerns about data privacy at the same time.

This learning process is likely to accompany a user's digital experience. It is easy to believe that as users gain more digital experience, they develop more demand for digital services, even though it may be less clear whether they also develop more concerns about data privacy. Figure 2 illustrates how privacy concerns vary across respondents in our survey sample with different digital experience. Specifically, it sorts all respondents into 12 groups, with the length of digital experience varying from one to 12 years. We measure the privacy concerns of each group by the fraction of the respondents who indicate they are "concerned" or "very concerned" about data sharing with mini-programs. The figure shows that privacy concerns indeed increase with digital experience. Also note that digital experience is unlikely the only factor that drives the users' process of learning about their digital demands and privacy concerns. To the extent that our regression analysis reported in Table 9 has controlled for digital experience, the positive relationship between digital demands and privacy concerns arises from learning beyond digital experience.

Another possibility is that privacy concerns grow with the accumulation of the data shared by a user with digital service providers. The accumulation of the shared data exposes the user to greater privacy risks. For example, the data might be hacked by or leaked to unauthorized parties; the data allow digital service providers to infer the user's reservation utility for different products and thus implement more-effective price-discrimination strategies; and the data allow firms to analyze the user's behavioral weakness such as lack of self-control to certain temptation goods and thus more effectively target their advertisements to the user. Regardless of the specific forms of privacy concerns, the user's privacy concerns are likely to grow with the accumulation of the shared data.

B. Activeness and Cancellation

To firmly establish the notion that privacy concerns grow with digital demands, we examine a further implication of this relationship. If individuals with greater digital demands are also more concerned by data privacy, we would expect more-active users of mini-programs to have a greater propensity to cancel previously authorized data sharing with mini-programs:

Hypothesis 3: Everything else being equal, more-active users of mini-programs are more likely to cancel data sharing with mini-programs.

One cannot take this hypothesis for granted as it counters our usual intuition that active users should be more reluctant to cancel data-sharing authorizations, which would prevent themselves from using those mini-programs. In our analysis, we focus on active cancellations by the users, rather than passive cancellations induced by authorization expirations.

To test this hypothesis, we use two measures of a user's overall activeness in mini-programs. The first is the *Active-Month Ratio,* which is defined as the weighted average fraction of months that the user uses each of the authorized mini-programs, where the weight for a mini-program is the number of months the user has authorized data sharing with the mini-program. The second measure is *log(1+ # Avg. Monthly Active Sessions)*, which is the user-level average of the number of active sessions in a mini-program in each month. *Cancellation Rate* is the number of canceled active authorizations from July 2019 to July 2020 (a one-year period before the survey) divided by the total number of outstanding authorized mini-programs during the period.

Panel A of Table 10 reports the user-level regression results. Due to missing data of some of the survey respondents, the sample size is 9,860. Column (1) shows that when *Active-Month Ratio* increases by 1%, the cancellation rate increases by 0.04%. Column (2) shows that when *log(1+ # Avg. Monthly Active Sessions)* increases by 1, the cancellation rate increases by 0.5%. These two regressions both confirm that more-active users are more likely to cancel previously authorized data sharing with mini-programs.

One might argue that cancellation of data sharing requires knowledge of how to cancel a data-sharing authorization and as a result, the positive relationship between cancellation and activeness may reflect active users' being more knowledgeable about cancellation rather than their privacy concerns. To address this argument, we restrict our sample to the respondents with at least one cancellation between January 2013 and June 2019, which is right before our main sample period started in July 2019. To the extent that these respondents all know how to cancel, the differential cancellation rate among them reflects the difference in privacy concerns rather than knowledge. In columns (3) and (4), we focus on this subsample of respondents with at least one cancellation before the sample period. The sample size drops from 9,860 to 3,916. Despite the smaller sample, the coefficients of the two activeness measures remain highly significant, with a 1% increases in *Active-Month Ratio* leading to a 0.08% increases in the cancellation rate, and an increase of 1 in *log(1+ # Avg. Monthly Active Sessions)* leading to a 1.2% increase in the cancellation rate.

Panel B of Table 10 shows the relationship between the user's activeness and the propensity to cancel a mini-program in the user-mini–program level. The activeness measures are still at the user level, and we control for mini-program fixed effects in all the regressions in addition to the previously used control variables. The strong positive relationship between user activeness and the propensity to cancel data-sharing authorization remains robust and highly significant, across the two measures of user activeness and across either the full sample of all survey respondents or the subsample of respondents who have previously canceled at least one data-sharing authorization.

Taken together, Table 10 shows that more-active users are more likely to cancel data sharing with mini-programs, and this positive relationship is not driven simply by active users being more knowledgeable about how to cancel a data-sharing authorization. Instead, this positive relationship between user activeness and the propensity to cancel data sharing supports Hypothesis 3 and confirms the key notion that users with greater digital demands tend to be more concerned about data privacy.

V. Analysis of the Random Sample

We acknowledge that our survey sample tends to include more-active users of Alipay, as they are more likely to complete the survey. This bias raises a natural concern that our findings may not hold in the general population of Alipay users. To address this concern, we also analyze a random sample of all Alipay users. This random sample contains 100,000 users that were drawn randomly from the whole population of all active Alipay users. As we already summarized in Table 3, the random sample is indeed less active in using mini-programs than the survey sample. The numbers of visited and authorized mini-programs in the random sample are only about one-third of those in the survey sample. Of the users in the random sample, 12% canceled data sharing with at least one mini-program, in contrast to 48% in the survey sample. As to the use of mini-programs, the average values of the four measures in the random sample reduce to less than one-half of those in the survey sample. These differences motivate us to examine whether our key findings remain robust in the random sample.

Before we show the robustness results, we also use the random sample to study the heterogenous reactions of Alipay users to a privacy-related incident in early 2018. This event study provides additional evidence on how these users' privacy concerns are related to their digital demands. We prefer the random sample for this event study because the analysis does not require the users' survey responses and the random sample is larger and more representative.

A. An Event Study

Even though the positive relationship between digital demands and the propensity to cancel mini-programs suggests that privacy concerns grow with digital demands, one may be concerned by the omitted variable problem that some unobservable, time-varying factors might have caused both privacy concerns and digital demands to grow over time. To address this concern, we take advantage of a salient event that had led to greater awareness of data privacy issues among the Alipay users.

On January 3, 2018, Alipay launched its Annual User Footprint Report within the mobile wallet app, allowing users to get an idea of how frequently and for what purposes they had used Alipay in 2017. By default, a box consenting to the "Sesame Credit Service Agreement" was checked on the report's landing page. Users who failed to notice the checked box would have unintentionally agreed to use Alipay's Sesame credit score service. Some internet users quickly discovered this misleading design, and this incident went viral on Chinese social media. On the same day, Alipay removed this particular default feature from the report and issued a statement to explain and apologize to the public, stating that it would not enroll users who had accidently consented to the agreement into its Sesame credit service. Despite these fixes, this incident sharply increased the public awareness of data privacy issues and led to a spike in Alipay users' cancellation of their data sharing with mini-programs, as shown by Figure A8. Thus, this incident provides an exogenous event for us to examine the heterogeneity in the reactions of Alipay users.

Specifically, we examine whether heavy users of mini-programs showed stronger reactions, which possibly reflect their stronger privacy concerns stimulated by the incident:

Hypothesis 4: In response to the incident, heavy users of mini-programs are likely to cancel data sharing with mini-programs.

To test this hypothesis, we follow an event study framework to analyze the following regression:

$$\begin{aligned} Daily\ Cancellation\ Dummy_{i,t} = \alpha_0 &+ \textstyle\sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{H,\tau} \cdot Heavy\ User_i \cdot \mathbb{1}(t=\tau) \\ &+ \beta_{H,6} \cdot Heavy\ User_i \cdot \mathbb{1}(t \geq 6) + \textstyle\sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{L,\tau} \cdot Light\ User_i \cdot \mathbb{1}(t=\tau) \\ &+ \beta_{L,6} \cdot Light\ User_i \cdot \mathbb{1}(t \geq 6) + \delta_i + \varepsilon_{i,t}, \end{aligned} \tag{4}$$

where $Daily\ Cancellation\ Dummy_{i,t}$ is a dummy variable indicating whether user i has cancelled at least one mini-program during the day t, t corresponds to the number of days after the incident on January 3, 2018, $Heavy\ User_i$ is a dummy indicating whether user i has more extensive use of mini-programs than 75% of the users in the sample as of November 30, 2017, $Light\ User_i$ is a dummy that equals $1 - Heavy\ User_i$, δ_i is the individual fixed effects, and $\varepsilon_{i,t}$ is the error term that varies across individuals and over time.

Panel A of Figure 3 depicts the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients estimated from the regression with the random sample. Consistent with our expectation, those heavy users of mini-programs are significantly more responsive to the incident, showing stronger privacy concerns through their greater propensity to cancel data sharing. However, the response is temporary, possibly due to the quick fixes by Alipay and the incident eventually going off the media. This finding is robust when we directly test the difference between the response of heavy and light users to this incident in Panel A of Figure A9.

Like before, one might argue that the greater propensity of heavy users to cancel data sharing reflects their better knowledge of how to cancel authorization in the Alipay app, rather than their stronger privacy concerns stimulated by the incident. To address this concern, we focus on the subsample of Alipay users in the random sample who had canceled data sharing with at least one mini-program before November 30, 2017. This filter ensures that the remaining users all had the necessary knowledge about data sharing cancellation before the incident. Panel B of Figure 3 depicts the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients estimated from this subsample. Although the behavioral gap between heavy and light users becomes smaller, the gap remains significant, with heavy users being more likely to cancel data sharing with mini-programs during the incident. The smaller gap indicates that knowledge also plays an important role in driving up the greater propensity of heavy users. For this subsample, we also directly test the difference in the response between heavy and light users in Panel B of Figure A9. The difference is significant on days 0, 2 and 3 of the incident.

Taken together, our analysis of the responses of Alipay users to the privacy-related incident on January 3, 2018 supports Hypothesis 4 and thus provides additional evidence that users with greater digital demands become more concerned about data privacy after the incident. This evidence reinforces the notion that data privacy concerns are positively correlated with demands for digital applications.

B. Robustness

We now use the random sample to verify our key findings from the survey sample. Because users in the random sample did not take our survey, we cannot use their responses to the survey questions to measure their privacy concerns. Instead, we use *Privacy Setting Changed*, a dummy indicating whether a user has changed the Alipay's default privacy settings, as a behavior-based measure of the user's privacy concerns. Relative to the survey-based measure, this behavior-based measure is more objective as it is immune to noise in the survey, but it is also affected by the user's knowledge about how to change Alipay's default privacy settings. Despite this potential weakness, we can still use this behavior-based measure, after suitable control for user knowledge, to examine how privacy concerns are related to data-sharing authorization and cancellation.

In Table 11, we report the results from using this behavior-based measure to re-examine the three key results in the random sample. Panel A shows the results from user-level regressions of the number of data-sharing authorizations or initial visits to mini-programs on users' privacy concerns, using similar specifications as Table 4. Interestingly, the more concerned users, as measured by changing their default privacy settings, not only visit significantly more mini-programs but also authorize data sharing with significantly more mini-programs, even after controlling for users' digital experience and age (which are powerful controls for user knowledge) as well as user gender and user city fixed effects. The greater number of data-sharing authorizations makes the data privacy paradox even stronger in the random sample.

Panel B reports how the use of mini-programs is related to privacy concerns by using specifications similar to Table 9, except that we use the privacy setting change dummy as the measure of privacy concerns. We again find that in the random sample, more-concerned users tend to use their authorized mini-programs more frequently and more extensively across the four use measures.

Panel C examines how the cancellation rate of data-sharing authorizations with mini-programs is related to user activeness, using specifications similar to Panel B of Table 10. We again observe that the cancellation rate is significantly and positively correlated with user activeness. Users with higher *Active-Month Ratio* and *log(1+ # Avg. Monthly Active Sessions)* have a higher probability of canceling their data-sharing authorizations. This relationship holds in both the full sample and the subsample of users who had previously canceled at least one mini-program before July 2019.

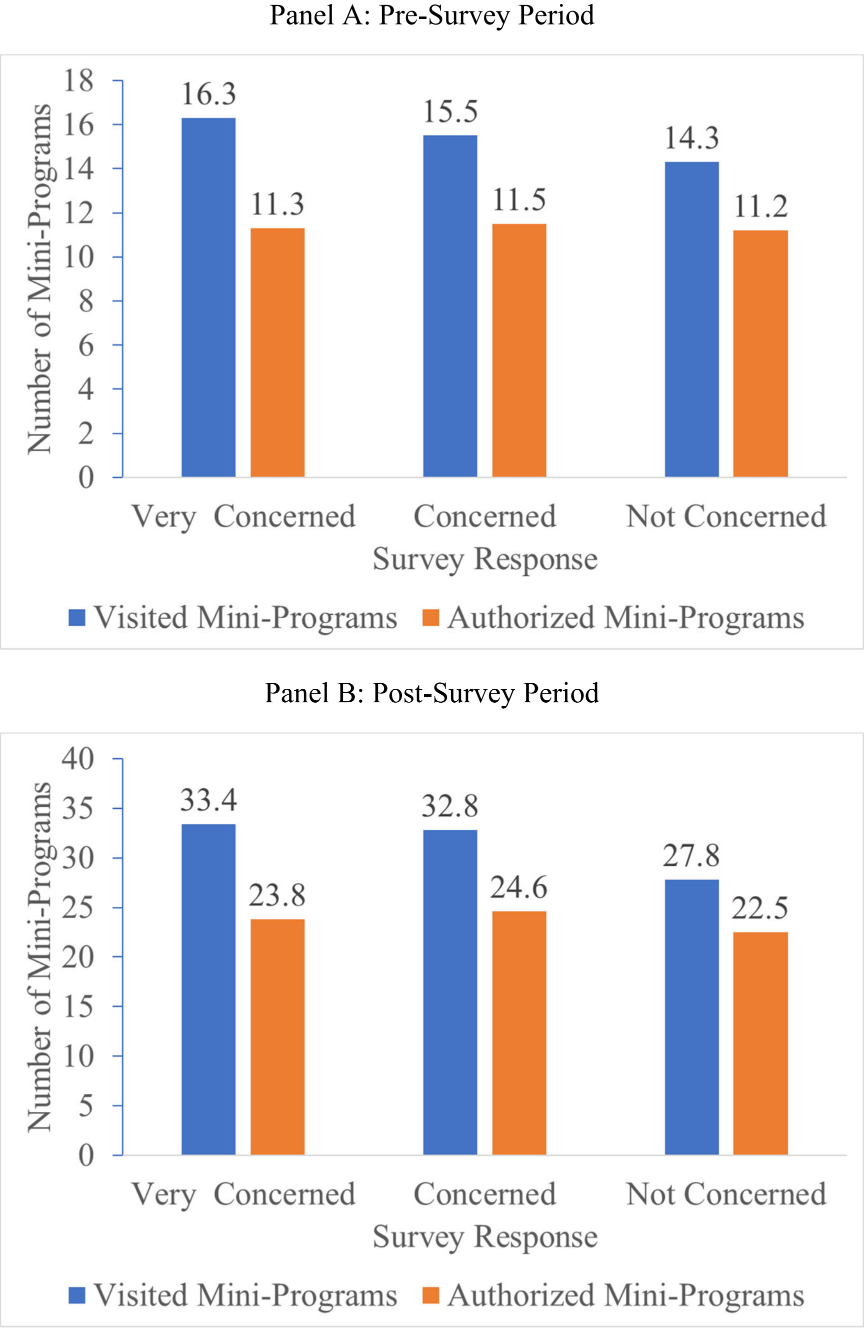
Taken together, Table 11 confirms that the three key results of our analysis are robust in the representative random sample of Alipay users.

VI. Conclusion

In this paper, we examine how data privacy preferences affect data sharing of Alipay users with third-party mini-programs in Alipay. Even though one would expect users with stronger privacy concerns to be more reluctant to share data, we find that there is a positive relationship between privacy concerns, measured by either survey responses or observed behaviors, and the number of data-sharing authorizations, confirming the puzzling data privacy paradox. We attribute this paradox to the trade-off faced by users between privacy costs and economic benefits of sharing personal data with mini-programs. In particular, we highlight a positive correlation between privacy concerns and demands for digital applications—that is, users with stronger privacy concerns tend to benefit more from using mini-programs. To the extent that economic benefits overcome consumers' privacy concerns in their decisions to share their personal data with mini-programs in our sample, our analysis confirms that data sharing is beneficial to consumer welfare.

Figure 1. The Data Privacy Paradox

This figure depicts the numbers of initial visits and data sharing authorizations to mini-programs by Alipay users in three groups based on their answers to the question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" Panel A covers the pre-survey period from July 2019 to July 2020, while Panel B covers the post-survey period from August 2020 to December 2021.



Panel A: Pre-Survey Period

Panel B: Post-Survey Period

Figure 2. Digital Experience and Privacy Concerns

This figure depicts the fraction of users indicating that they are “concerned” or “very concerned” about negative impacts caused by information shared with mini-programs in Alipay, across groups with different digital experience, measured by the length of time since a user registered on Alipay. For each group, we also show the 68.3% confidence band of the mean estimate.

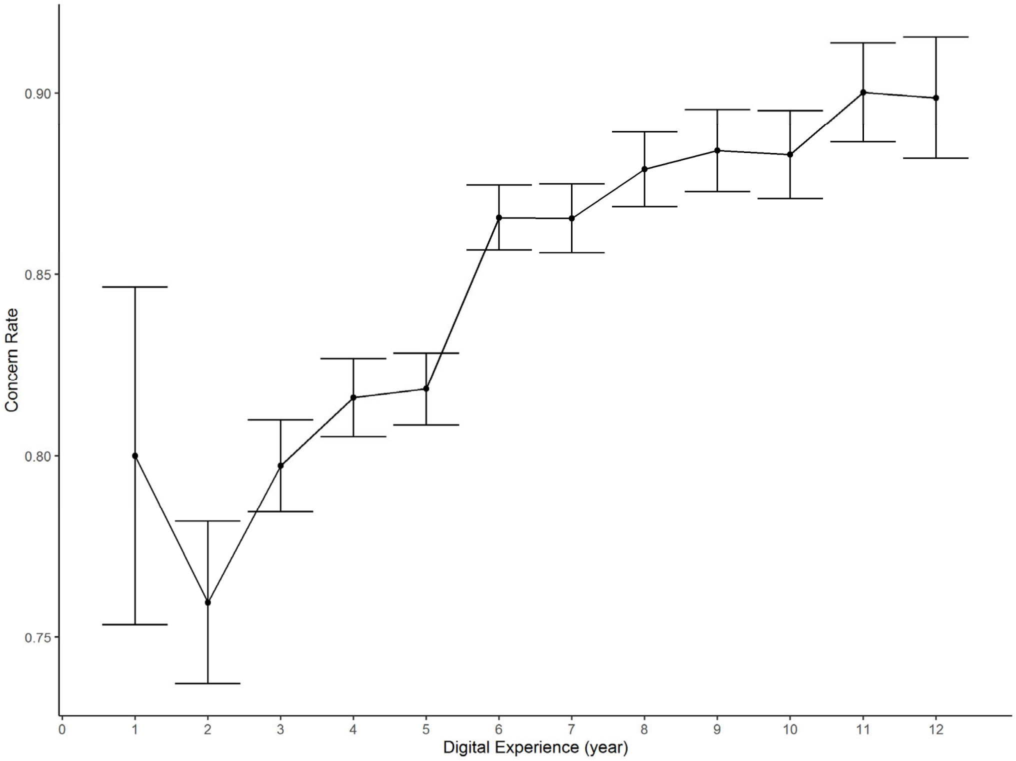
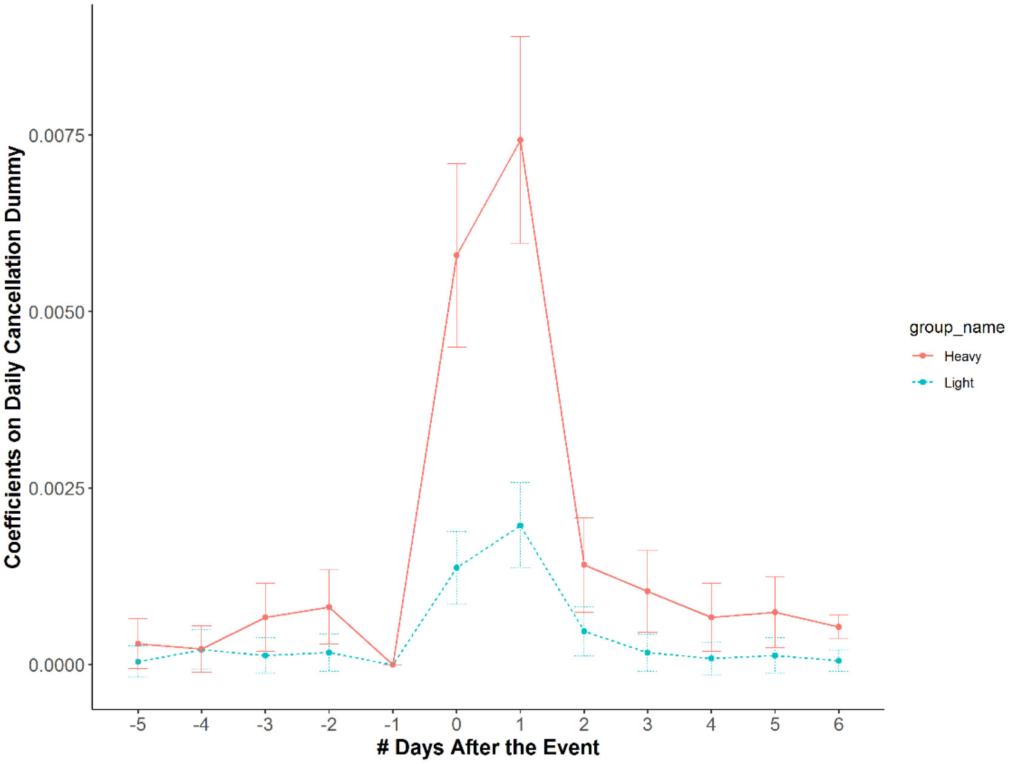


Figure 3. Activeness and Response to the 2017 Footprint Report Incident

The figures plot the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients estimated the regression specified in equation (4), where the bands indicate 95% confidence intervals. Panel A covers the random sample of 100,000 Alipay users without any filtering, and Panel B covers only the users who had cancelled data sharing with at least one mini-program before November 30, 2017 in the random sample. The data is at individual and daily level. The sample period ranges from December 29, 2017 to January 31, 2018.

Panel A: Unfiltered Users



Panel B: Users with Cancellation before November 30, 2017

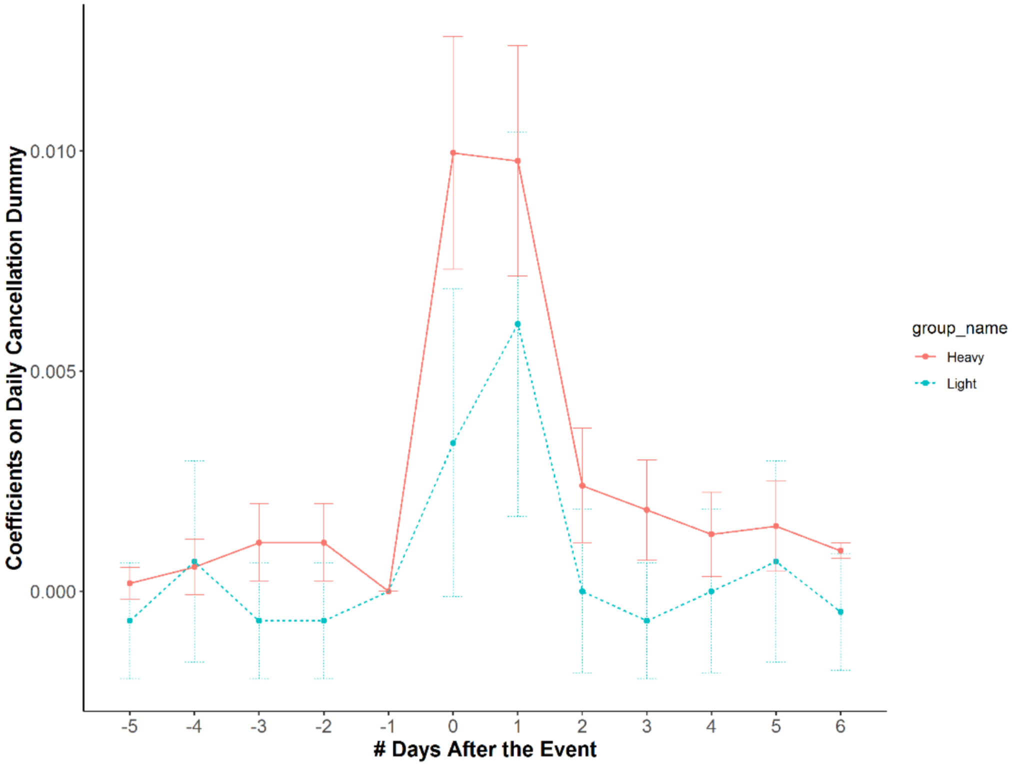


Table 1. Responses to Selected Survey Questions

This table summarizes responses to seven of the survey questions.

	Count	Total	Share
A. Are you concerned about privacy issues while using online services?			
Very concerned	13284	14250	93%
Concerned	882	14250	6%
Not concerned	84	14250	1%
B. What do you think about privacy protection in Alipay?			
Very good	6789	14250	48%
Ordinary	5600	14250	39%
Not good	679	14250	5%
No idea	1182	14250	8%
C. Do you know how to change privacy settings in Alipay?			
Yes	8529	14250	60%
No	5721	14250	40%
D. Have you ever changed your privacy settings in Alipay?			
Yes	5557	14250	39%
No	5025	14250	35%
No idea	3668	14250	26%
E. Have you ever used mini-programs in Alipay?			
Yes	10875	14250	76%
No	3375	14250	24%
F. Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?			
Very concerned	5005	10875	46%
Concerned	4244	10875	39%
Not concerned	1626	10875	15%
G. What privacy issues are you concerned about when using mini-programs in Alipay? (multiple choices)			
Data leakage and security	9377	10875	86%
Price discrimination by merchants	2314	10875	21%
Seductive advertising and temptation consumption	5333	10875	49%
Others	500	10875	5%

Table 2. Summary Statistics of the Survey Sample

This table reports summary statistics of the main sample of 10,875 users who finished the survey in July 2020 and indicated that they had used mini-programs in Alipay. Panels A reports user information in three parts. The first part reports the general information. *Concerned Dummy* and *Very Concerned Dummy* are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." *Privacy Setting Changed*, a proxy measure for privacy concerns, is a dummy variable equal to 1 if a user changed their privacy setting at least once between May 2017 and April 2020, and 0 otherwise. *Digital Experience* is the number of months since the user firstly registered on Alipay, and *Age* is the user's physical age in July 2020. The second part covers data sharing with mini programs, including the number of authorized and entered mini-programs over both the pre-survey period of July 2019 to July 2020 and the post-survey period of August 2020 to December 2021; the *Has Canceled* status, *# Cancellations,* and *Cancellation Rate* of used mini-programs over the pre-survey period of January 2013 to July 2020. The third part reports summary statistics of monthly use variables of Alipay users in each mini-program during the pre-survey period from July 2019 to July 2020, including number of active days, number of uses, number of launches, and number of visited pages. Use variables are winsorized at the 1% and 99% levels. Panel B reports the mean digital experience, age, female dummy, and education dummy for each group. *Female Dummy* equals 1 if a user is female, and 0 otherwise. *Education Dummy* equals 1 if a user has a college degree or above, and 0 otherwise.

Panel A: User Information

	N	Mean	Std	Min	p25	Median	p75	Max
General information								
Concerned Dummy	10,875	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Very Concerned Dummy	10,875	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Privacy Setting Changed	10,875	0.49	0.5	0.00	0.00	0.00	1.00	1.00
Digital Experience (month)	10,871	74.97	35.07	4.00	48.00	70.00	97.00	190.00
Age (year)	10,858	32.82	10.27	10.00	25.00	31.00	39.00	82.00
Data sharing with mini-programs								
# Authorized Mini-Programs	10,875	34.22	22.78	0.00	19.00	30.00	43.00	422.00
# Entered Mini-Programs	10,875	46.57	55.45	1.00	26.00	38.00	53.00	1609.00
Has Canceled	10,875	0.48	0.50	0.00	0.00	0.00	1.00	1.00
# Cancellations	10,857	2.66	5.54	0.00	0.00	0.00	3.00	80.00
Cancellation Rate	10,857	0.05	0.10	0.00	0.00	0.00	0.06	1.00
Monthly mini-program use								
# Active Days	1,521,645	0.57	2.92	0.00	0.00	0.00	0.00	31.00
# Uses	1,521,645	0.81	5.01	0.00	0.00	0.00	0.00	75.00
# Launches	1,521,645	2.29	15.07	0.00	0.00	0.00	0.00	230.00
# Visited Pages	1,521,645	5.20	33.67	0.00	0.00	0.00	0.00	503.00

Panel B: Privacy Concern and Personal Characteristics

	Not Concerned	Concerned	Very Concerned	Difference	Difference
	(1)	(2)	(3)	(2) – (1)	(3) – (1)
Mean Digital Experience	66.868	75.725	76.961	8.857***	10.093***
				(1.018)	(0.996)
Mean Age	32.873	32.731	32.881	-0.142	0.008
				(0.300)	(0.293)
Mean Female Dummy	0.148	0.282	0.280	0.134***	0.132***
				(0.013)	(0.012)
Mean Education Dummy	0.137	0.221	0.214	0.084***	0.077***
				(0.012)	(0.012)

Table 3. Summary Statistics of the Random Sample

This table reports summary statistics of a representative random sample of 100,000 Alipay users. Panels A reports user information in three parts. The first part reports the general information. *Privacy Setting Changed*, a proxy measure for privacy concerns, is a dummy variable equal to 1 if a user changed their privacy setting at least once between May 2017 and April 2020, and 0 otherwise. *Digital Experience* is the number of months since the user firstly registered on Alipay, and *Age* is the user's physical age in July 2020. The second part covers data sharing with mini programs, including the number of authorized, entered, and canceled mini-programs over the pre-survey period of July 2019 to July 2020; the *Cancellation Rate* of used mini-programs between July 2019 and July 2020; and the *Has Canceled* status over the period of January 2013 to July 2020. The third part reports summary statistics of monthly use variables of Alipay users in each mini-program during the pre-survey period from July 2019 to July 2020, including number of active days, number of uses, number of launches, and number of visited pages. Use variables are winsorized at the 1% and 99% levels.

	N	Mean	Std	Min	p25	Median	p75	Max
General information								
Privacy Setting Changed	98,679	0.09	0.28	0.00	0.00	0.00	0.00	1.00
Digital Experience (month)	99,600	60.69	36.81	0.00	32.00	55.00	82.00	190.00
Age (year)	97,876	36.61	12.89	1.00	27.00	34.00	46.00	120.00
Data sharing with mini programs								
# Authorized Mini-Programs	100,000	2.40	3.52	0.00	0.00	1.00	3.00	136.00
# Entered Mini-Programs	100,000	3.02	4.59	0.00	0.00	2.00	4.00	248.00
Has Canceled	99,995	0.12	0.32	0.00	0.00	0.00	0.00	1.00
# Cancellations	98,674	0.30	1.45	0.00	0.00	0.00	0.00	61.00
Cancellation Rate	98,674	0.01	0.07	0.00	0.00	0.00	0.00	1.00
Monthly mini-program use								
# Active Days	3,036,555	0.27	1.59	0.00	0.00	0.00	0.00	27.00
# Uses	3,036,555	0.34	2.21	0.00	0.00	0.00	0.00	40.00
# Launches	3,036,555	1.10	6.90	0.00	0.00	0.00	0.00	123.00
# Visited Pages	3,036,555	3.06	19.96	0.00	0.00	0.00	0.00	342.00

Table 4. The Data Privacy Paradox at the User Level

This table presents regression analysis of the data privacy paradox at the user level. *Concerned Dummy* and *Very Concerned Dummy* in Panel A are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." Panel A reports results for the pre-survey period from July 2019 to July 2020, while Panel B reports results for the post-survey period from August 2020 to December 2021. Columns (1)–(2) show results for the number of authorized mini-programs, and columns (3)–(4) for the number of initially visited mini-programs. We report standard errors in parentheses. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively.

	# Authorized Mini-programs		# Visited Mini-programs	
	(1)	(2)	(3)	(4)
Panel A: Pre-Survey Period				
Concerned Dummy	0.334	0.207	1.262***	1.243***
	(0.213)	(0.214)	(0.322)	(0.320)
Very Concerned Dummy	0.127	-0.007	1.990***	1.965***
	(0.209)	(0.211)	(0.331)	(0.336)
Constant	11.177***		14.310***	
	(0.178)		(0.274)	
Observations	10,875	10,858	10,875	10,858
Adjusted *R*2	0.0001	0.021	0.003	0.045
Panel B: Post-Survey Period				
Concerned Dummy	2.044***	1.292**	5.007***	4.104***
	(0.534)	(0.541)	(1.124)	(1.122)
Very Concerned Dummy	1.308**	0.632	5.592***	5.003***
	(0.536)	(0.540)	(1.145)	(1.199)
Constant	22.532***		27.790***	
	(0.460)		(0.843)	
Observations	10,875	10,858	10,875	10,858
Adjusted *R*2	0.001	0.050	0.001	0.05
City FE	N	Y	N	Y
Gender FE	N	Y	N	Y
Control Age	N	Y	N	Y
Control Digital Experience	N	Y	N	Y

Table 5. The Data Privacy Paradox at the User-Mini–Program Level

This table presents regression analysis for the data privacy paradox at the User-Mini–Program Level. *Concerned Dummy* and *Very Concerned Dummy* are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." Panel A reports results for the pre-survey period from July 2019 to July 2020, while Panel B reports results for the post-survey period is from August 2020 to December 2021. Columns (1)–(2) show results for the number of authorized mini-programs, and columns (3)–(4) for the number of initially visited mini-programs. We cluster the standard errors at the user level and report them in parentheses. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively.

	Authorized Dummy (0/1)		Visited Dummy (0/1)	
	(1)	(2)	(3)	(4)
Panel A: Pre-Survey Period				
Concerned Dummy (× E-4)	0.862	0.386	2.897***	2.552***
	(0.745)	(0.735)	(0.848)	(0.836)
Very Concerned Dummy (× E-4)	0.028	-0.465	3.755***	3.340***
	(0.736)	(0.728)	(0.846)	(0.840)
Constant	0.004***		0.005***	
	(0.0001)		(0.0001)	
Observations	25,414,875	25,364,288	25,414,875	25,364,288
Adjusted *R*2	0.000	0.105	0.000	0.129
Panel B: Pre-Survey Period				
Concerned Dummy (× E-4)	2.496***	1.667***	3.918***	3.090***
	(0.564)	(0.557)	(0.622)	(0.623)
Very Concerned Dummy (× E-4)	1.452***	0.743	3.367***	2.668***
	(0.558)	(0.548)	(0.616)	(0.617)
Constant	0.003***		0.003***	
	(0.000)		(0.000)	
Observations	64,999,875	64,887,408	64,999,875	64,887,408
Adjusted *R*2	0.000	0.106	0.000	0.121
Mini-program FE	N	Y	N	Y
City FE	N	Y	N	Y
Gender FE	N	Y	N	Y
Control Age	N	Y	N	Y
Control Digital Experience	N	Y	N	Y

Table 6. Heterogeneity Analysis of the Data Privacy Paradox

This table reports heterogeneity analysis of the data privacy paradox. *Concerned Dummy* and *Very Concerned Dummy* are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." We interact privacy concern measures with a user's personal characteristics. *Education* is a dummy indicating whether the user has a college degree or above. *Self Control* is a dummy indicating whether the user's opt-in rate of seemingly addictive mini-programs is higher than the opt-in rate of other mini-programs in the pre-survey period. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Authorized Dummy (0/1)			
	Pre-Survey Period		Post-Survey Period	
	(1)	(2)	(3)	(4)
Concerned Dummy (× E-4)	-0.649	0.600	1.153*	2.127***
	(0.811)	(0.771)	(0.592)	(0.607)
Very Concerned Dummy (× E-4)	-1.096	-0.327	0.417	1.128*
	(0.807)	(0.765)	(0.584)	(0.589)
Concerned Dummy × Characteristics Measure (× E-4)	5.120***	-2.644	0.823	-2.370*
	(1.833)	(1.855)	(1.587)	(1.340)
Very Concerned Dummy × Characteristics Measure (× E-4)	3.329*	-2.501	0.045	-2.161
	(1.800)	(1.827)	(1.572)	(1.347)
Characteristics Measure	-0.000	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Characteristics Measure	Education	Self-Control	Education	Self-Control
Mini-program FE	Y	Y	Y	Y
City FE	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y
Control Age	Y	Y	Y	Y
Control Digital Experience	Y	Y	Y	Y
Observations	25,364,288	25,364,288	64,887,408	64,887,408
Adjusted *R*2	0.105	0.105	0.106	0.106

Table 7. Validating Survey-Based Privacy Concerns

This table reports the relationship between the survey-based measure of privacy concerns and actions taken to protect data privacy, including canceling data-sharing authorizations with mini-programs and changing Alipay's default privacy settings. *Concerned Dummy* and *Very Concerned Dummy* in Panel A are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." Panel A shows results for user-level regressions. In columns (1)–(2), the dependent variable is a dummy that indicates whether a user has canceled at least one data-sharing authorization in the period of January 2013 to July 2020. In columns (3)–(4), the dependent variable is a dummy that indicates whether a user has changed the Alipay's default privacy settings the period of May 2017 to April 2020. Panel B shows results for user-mini–program level regressions, where we cluster the standard errors at the user level. In each pair of user-mini–program with existing data-sharing authorization, the dependent variable is a dummy that indicates whether the user canceled the authorization in the period of July 2019 to July 2020. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User Level Analysis

	Has Canceled (0/1)		Privacy Setting Changed (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy	0.060***	0.033***	0.028*	0.012
	(0.014)	(0.014)	(0.015)	(0.015)
Very Concerned Dummy	0.082***	0.051***	0.060***	0.041***
	(0.014)	(0.014)	(0.014)	(0.015)
Digital Experience		0.004***		0.001***
		(0.0001)		(0.0001)
Age		-0.003***		-0.001***
		(0.0005)		(0.0005)
Constant	0.420***		0.454***	
	(0.012)		(0.012)	
City FE	N	Y	N	Y
Gender FE	N	Y	N	Y
Observations	10,857	10,841	10,875	10,858
Adjusted *R*2	0.003	0.097	0.002	0.011

Panel B: User-Mini–Program Level Analysis

	$Canceled\ Dummy_{ij}$	
	(1)	(2)
Concerned Dummy	-0.001	0.004
	(0.003)	(0.003)
Very Concerned Dummy	0.005	0.011***
	(0.003)	(0.003)
Digital Experience (× E-4)		1.218***
		(0.305)
Age (× E-4)		2.547**
		(1.141)
Constant	0.058***	
	(0.003)	
Mini-program FE	N	Y
City FE	N	Y
Gender FE	N	Y
Observations	481,143	480,542
Adjusted $R2$	0.0001	0.107

Table 8. Determinants of Data-Sharing Authorizations in Survey

Panel A summarizes the responses of the respondents to five statements. The respondents are split into two groups, one for those whose answers to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" are "concerned" or "very concerned," and the other group for those whose answers to this survey question are "not concerned." Panel B shows the regression results. The dependent variable takes a value of 1 if a respondent agrees with a statement. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: Summary of Responses to Survey Statements

	Count	Share	Count	Share	Total
	Agree		Disagree		
Q1: I agree to authorize data sharing with mini-programs because it is safe in Alipay.					
Concerned or very concerned	3,918	42%	5,331	58%	9,249
Not concerned	1,308	80%	318	20%	1,626
Q2: I agree to authorize data sharing with mini-programs because my information has already been shared in many platforms.					
Concerned or very concerned	1,083	12%	8,166	88%	9,249
Not concerned	493	30%	1,133	70%	1,626
Q3: I have to share my information in exchange for digital services even though I have concerns about my data privacy.					
Concerned or very concerned	6,030	65%	3,219	35%	9,249
Not concerned	913	56%	713	44%	1,626
Q4: I only authorize data sharing with mini-programs when the requested data are not important.					
Concerned or very concerned	1,852	20%	7,397	80%	9,249
Not concerned	485	30%	1,141	70%	1,626
Q5: I tend to authorize data sharing with mini-programs that are used by my friends.					
Concerned or very concerned	4,042	44%	5,207	56%	9,249
Not concerned	942	58%	684	42%	1,626

Panel B: Regression Analysis

Agree with	Q1	Q2	Q3	Q4	Q5
	(1)	(2)	(3)	(4)	(5)
Concerned or very concerned	-0.320***	-0.203***	0.083***	-0.096***	-0.158***
	(0.011)	(0.013)	(0.014)	(0.014)	(0.014)
Digital Experience	-0.001***	-0.001***	0.0003**	-0.001***	-0.00001
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Age	0.002***	0.001**	0.0005	0.004***	-0.001
	(0.001)	(0.0004)	(0.0005)	(0.0005)	(0.001)
City FE	Y	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y	Y
Observations	8,658	9,637	9,780	9,356	9,110
Adjusted *R*2	0.070	0.052	0.013	0.019	0.014

Table 9. Demand for Digital Services

This table examines the relationship between privacy concerns and demand for digital services. *Concerned Dummy* and *Very Concerned Dummy* in Panel A are dummy variables that equal 1 if the answer to the survey question "*Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?*" is "concerned" or "very concerned." We use four user-app-month–level variables from July 2019 to July 2020 to capture demand for digital services, namely, number of active days, number of uses, number of launches, and number of visited pages. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We cluster the standard errors at the user level and report standard errors in parentheses.

	# Active Days		# App Uses		# App Launches		# Visited Pages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Concerned Dummy	0.102***	0.088***	0.155***	0.138***	0.434***	0.399***	0.847***	0.772***
	(0.027)	(0.020)	(0.046)	(0.035)	(0.131)	(0.105)	(0.262)	(0.219)
Very Concerned Dummy	0.126***	0.102***	0.206***	0.172***	0.568***	0.490***	1.144***	0.996***
	(0.028)	(0.021)	(0.048)	(0.037)	(0.135)	(0.110)	(0.269)	(0.230)
Digital Experience		-0.0001		-0.0003		-0.001		-0.001
		(0.000)		(0.001)		(0.001)		(0.003)
Age		0.020***		0.033***		0.080***		0.128***
		(0.001)		(0.002)		(0.005)		(0.011)
Constant	0.468***		0.651***		1.864***		4.339***	
	(0.023)		(0.039)		(0.112)		(0.226)	
Mini-program FE	N	Y	N	Y	N	Y	N	Y
Year-Month FE	N	Y	N	Y	N	Y	N	Y
City FE	N	Y	N	Y	N	Y	N	Y
Gender FE	N	Y	N	Y	N	Y	N	Y
Observations	1,521,645	1,519,020	1,521,645	1,519,020	1,521,645	1,519,020	1,521,645	1,519,020
Adjusted *R*2	0.0002	0.119	0.0002	0.096	0.0001	0.086	0.0001	0.078

Table 10. Activeness and Cancellation

This table examines the relationship between user activeness and cancellation of previously authorized mini-programs. The sample covers user-mini–program pairs that had been active between July 2019 and July 2020. *Cancellation Rate* is the number of canceled mini-programs by a user from July 2019 to July 2020 divided by the total number of the user's active mini-programs. We use two user-level measures of activeness. The first one is active-month ratio, which refers to the total number of months a user has been active as a percentage in the total number of months from the beginning to the end of authorizations in all mini-programs. The second one is the logarithm of the average monthly active uses. Panel A shows results for the user-level regression. We use the whole sample in columns (1) and (2) and a subsample with users who canceled at least one mini-program before July 2019 in columns (3) and (4). Panel B reports the results of the user-mini–program level regressions, where we cluster the standard errors at the user level. We use the whole sample in columns (1) and (2) and a subsample with users who canceled at least one mini-program before July 2019 in columns (3) and (4). We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User Level Regression

	$Cancellation\ Rate_i$			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.042***		0.080***	
	(0.008)		(0.016)	
log(1+ # Avg. Monthly Active Sessions)		0.005***		0.012***
		(0.001)		(0.003)
Digital Experience (× E-4)	-0.112	-0.203	-1.834***	-2.000***
	(0.194)	(0.194)	(0.448)	(0.454)
Age (× E-4)	-1.250*	-0.549	-1.666	-0.682
	(0.746)	(0.689)	(1.896)	(1.823)
City FE	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y
Sample	All	All	Has Canceled	Has Canceled
Observations	9,860	9,860	3916	3916
Adjusted *R*2	0.012	0.005	0.027	0.014

Panel B: User-Mini–Program Level Regression

	$Canceled\ Dummy_{ij}$			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.047***		0.081***	
	(0.007)		(0.011)	
log(1+ # Avg. Monthly Active Sessions)		0.003**		0.007***
		(0.001)		(0.003)
Digital Experience (× E-4)	1.557***	1.464***	-2.358***	-2.534***
	(0.218)	(0.217)	(0.410)	(0.409)
Age (× E-4)	-0.284	0.885	3.818**	5.396***
	(0.810)	(0.812)	(1.532)	(1.551)
Mini-program FE	Y	Y	Y	Y
City FE	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y
Sample	All	All	Has Canceled	Has Canceled
Observations	437,521	437,521	231,255	231,255
Adjusted *R*2	0.127	0.127	0.172	0.170

Table 11. Robustness Tests

This table reports three sets of robustness tests from using the representative random sample of 100,000 Alipay users. Panel A presents the robustness test for the digital privacy paradox, where the regressions are at the user level. *Privacy Setting Changed* is a behavior-based measure for privacy concerns, defined as a dummy variable that equals 1 if a user changed the default privacy settings at least once between May 2017 and April 2020, and 0 otherwise. Columns (1) and (2) show results for the number of authorized mini-programs, and columns (3) and (4) show results for the number of initially visited mini-programs. In columns (2) and (4), we control for digital experience and age, along with gender and city fixed effects. Panel B tests the positive relationship between privacy concerns and demand for digital services, where the regressions are in the user-mini–program-month level, and the standard errors are clustered at the user level. We use four variables from July 2019 to July 2020 to capture demand for digital services, namely, number of active days, number of uses, number of launches, and number of visited pages. Columns (1), (3), (5), and (7) show regression results without any controls, while columns (2), (4), (6), and (8) control for digital experience and age, as well as user gender, user city, mini-program, and year-month fixed effects. Panel C examines the positive relationship between user activeness and cancellation of mini-programs, where the regressions are at the user-mini-program level, and the standard errors are clustered at the user level. The sample covers user-mini–program pairs that had been active between July 2019 and July 2020. We use two measures of user activeness. The first one is active-month ratio that refers to the total number of months the user is active as a percentage of the total number of months from the beginning to the end of authorizations in all mini-programs. The second one is the logarithm of the average monthly active uses. We use the whole sample in columns (1) and (2) and a subsample of users who canceled at least one mini-program before July 2019 in columns (3) and (4). In all the regressions, we control for digital experience and age, as well as gender and city fixed effects. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User Level Analysis of the Data Privacy Paradox

	# Authorized Apps		# Visited Apps	
	(1)	(2)	(3)	(4)
Privacy Setting Changed	2.851***	2.443***	3.599***	3.158***
	(0.083)	(0.082)	(0.117)	(0.116)
Controls	N	Y	N	Y
Observations	98,679	96,596	98,679	96,596
Adjusted *R*2	0.023	0.094	0.022	0.068

Panel B: User-Mini–Program-Month Level Analysis of Privacy Concerns and Digital Demand

	# Active Days		# Active Sessions		# App Launches		# Visited Pages	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Privacy Setting Changed	0.032***	0.043***	0.042***	0.059***	0.102***	0.173***	0.301***	0.521***
	(0.009)	(0.007)	(0.012)	(0.010)	(0.034)	(0.031)	(0.086)	(0.081)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	3,021,210	3,007,635	3,021,210	3,007,635	3,021,210	3,007,635	3,021,210	3,007,635
Adjusted $R2$	0.00005	0.061	0.00004	0.052	0.00003	0.046	0.00003	0.045

Panel C: User-Mini–Program Level Analysis of Activeness and Cancellation

	$Canceled\ Dummy_{ij}$			
	(1)	(2)	(3)	(4)
Active-Month Ratio	0.002		0.026***	
	(0.001)		(0.005)	
log(1+ # Avg. Monthly Active Sessions)		0.003***		0.011***
		(0.001)		(0.002)
Controls	Y	Y	Y	Y
Sample	All	All	Has Canceled	Has Canceled
Observations	1,048,150	1,048,150	324,094	324.094
Adjusted $R2$	0.140	0.141	0.205	0.205

References

Acquisti, A. (2004). Privacy in Electronic Commerce and the Economics of Immediate Gratification. *Proceedings of the 5th ACM Conference on Electronic Commerce,* 21-29.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age. *Journal of Consumer Psychology*, 30(4), 736-758.

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What Is Privacy Worth? *Journal of Legal Studies*, 42(2), 249–274.

Acquisti, A., & Varian, H. R. (2005). Conditioning Prices on Purchase History. *Marketing Science*, 24(3), 367–381.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2), 442–492.

Aridor, G., Che, Y. K., & Salz, T. (2020). The Economic Consequences of Data Privacy Regulation: Empirical Evidence From GDPR. National Bureau of Economic Research.

Athey, S., Catalini, C., & Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk (Working Paper No. 23488). National Bureau of Economic Research.

Bergemann, D., & Morris, S. (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, 57(1), 44–95.

Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data, *American Economic Review* 91, 67–72.

Brandimarte, L., Acquisti, A. and Loewenstein, G., (2013). Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4(3), 340-347.

Chen, L., Bolton, P., Holmström, B. R., Maskin, E., Pissarides, C. A., Spence, A. M., Sun, T., Sun, T., Xiong, W., Yang, L., Huang, Y., Li, Y., Luo, X., Ma, Y., Ouyang, S., & Zhu, F. (2021). Understanding Big Data: Data Calculus in the Digital Era. Luohan Academy Report.

Cong, W., Xie, D., & Zhang, L. (2020). Knowledge Accumulation, Privacy, and Growth in a Data Economy. *Management Science*, forthcoming.

Fainmesser, I. P., Galeotti, A., & Momot, R. (2019). Digital Privacy. Social Science Research Network.

Farboodi, M., & Veldkamp, L. (2020). Long-Run Growth of Financial Data Technology. *American Economic Review*, 110(8), 2485–2523.

Giglio, S., Maggiori, M., Stroebel, J., & Utkus, S. (2020). Five Facts about Beliefs and Portfolios, *American Economic Review*, forthcoming.

Johnson, G., Shriver, S., & Goldberg, S. (2019). Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR.

Goldfarb, A., & Tucker, C. (2012). Shifts in Privacy Concerns. *American Economic Review,* 102(3), 349–53.

Goldfarb, A., & Tucker, C. (2019). Digital Economics. *Journal of Economic Literature*, 57(1), 3–43.

Gross, R., & Acquisti, A. (2005). Information Revelation and Privacy in Online Social Networks (The Facebook Case). 11.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9), 2819–2858.

Liu, H., Peng, C., Xiong, W., & Xiong, W. (2022). Taming the Bias Zoo. *Journal of Financial Economics,* 143, 716–741.

Liu, Z., Sockin, M., & Xiong, W. (2020). Data Privacy and Consumer Vulnerabilities. Working Paper, Princeton.

Ouyang, S. (2021). Cashless Payment and Financial Inclusion. Working Paper, Princeton.

Pew Research Center (2019). Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Pew Research Center.

Sockin, M. & Xiong, W. (2022). Decentralization Through Tokenization. *Journal of Finance*, forthcoming.

Solove, D. J. (2021). The Myth of the Privacy Paradox. *George Washington Law Review*, 89, 1-51.

Tang, H. (2020). The Value of Privacy: Evidence from Online Borrowers. Working Paper, London School of Economics.

Taylor, C. (2004). Consumer Privacy and the Market for Customer Information. *RAND Journal of Economics* 35 (4), 631–50.

Appendix: The Survey Questionnaire

Q1. Are you concerned about privacy issues while using online services?

Q2. What do you think about privacy protection in Alipay?

Q3. Are you concerned about negative impacts caused by information shared to mini-programs in Alipay?

Q4. Will you avoid visiting mini-programs in Alipay because of privacy concerns?

Q5. What privacy issues are you concerned about when using mini-programs in Alipay? (You may select multiple choices.)

A. Data leakage and security;

B. Price discrimination by merchants;

C. Seductive advertising and temptation consumption;

D. Others

Q6. How many times will you agree if making authorization decisions for ten mini-programs?

Q7. How often do you regret authorizing information to mini-programs in Alipay?

Q8. Do you agree with the arguments below?

1) I agree to authorize data sharing with mini-programs since it is safe in Alipay.

2) I agree to authorize data sharing with mini-programs since my information has already been shared in many platforms.

3) I have to share my personal data in exchange for digital services even though I am concerned about my data privacy.

4) I authorize data sharing with a mini-program only when the requested information is not important.

5) I tend to authorize data sharing with mini-programs that are used by my friends.

Q9. Do you know how to change privacy settings in Alipay?

Q10. Have you ever changed your privacy settings in Alipay?

Q11. Do you know how to opt out from mini-programs in Alipay?

Q12. Have you ever opted out from mini-programs in Alipay?