Data Privacy and Digital Demand*

Long Chen, Yadong Huang, Shumiao Ouyang, Wei Xiong

June 2022

Abstract

We combine survey and behavioral data to analyze consumers' data-sharing choices in a realistic setting in which they exchange personal data for digital services. We find that respondents with stronger privacy concerns authorize more, rather than less, data sharing, confirming the data privacy paradox. Instead of attributing this paradox to the respondents' unreliable survey responses, resignation from privacy, or behavioral biases, we uncover that privacy-concerned respondents have greater demands for digital services, which offset their privacy concerns. Our findings highlight a key tension for the data economy—privacy concerns and digital demands both grow with the deepening of digital services.

* Long Chen and Yadong Huang are affiliated with the Luohan Academy, and Shumiao Ouyang and Wei Xiong are affiliated with Princeton University. We appreciate comments and suggestions made by Alessandro Acquisti, Laura Brandimarte, Avi Goldfarb, Cameron Peng, Avanidhar Subrahmanyam, Huan Tang, Catherine Tucker, Liad Wagman, David Yang, Liyan Yang, and seminar participants at the American Finance Association Meetings, Boston College, the CBDC Series, CUHK Shenzhen, EIEF, the IMF, the Kansas City Fed, the NBER Conference on Economics of Privacy, Peking University, Princeton, SAIF, and UCLA. We are grateful for the help provided by Yong Li and Dongyu Wei at the Luohan Academy and Xiaopeng Wang and Xiaoqiang Yang at Ant Group. Wei Xiong also thanks the Keynes Fund of the University of Cambridge for financial support. 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, which may become a driver of the macroeconomy, as recognized by the recent theoretical models of Jones and Tonetti (2020), Farboodi and Veldkamp (2020), and Cong, Xie and Zhang (2020).[1] However, there are also growing concerns about data privacy protections across the world, as reflected by changes in consumer altitudes, see for example, Goldfarb and Tucker (2012), and the enactments of the General Data Privacy Regulation (GDPR) by the European Union in 2018, the California Consumer Privacy Act (CCPA) by the state of California in 2020, and the Personal Information Protection Law (PIPL) by China in 2021. Despite the importance of data privacy and protections, there is little agreement among academics and policy makers regarding how consumers' privacy concerns affect their data-sharing choices, as reflected by the controversies regarding the "data privacy paradox," a term used by commentators to describe a disconnect between consumers' self-stated privacy concerns and their actual privacy-seeking behavior. A wide range of surveys and experimental studies (e.g., Spiekermann, Grossklags and Berendt, 2001; Gross and Acquisti, 2005; Norberg, Horne and Horne, 2007; Athey, Catalini and Tucker, 2017), have documented that consumers often say they care about privacy but at same time choose to share their personal data either freely or for small rewards. See Acquisti, Brandimarte and Loewenstein (2020) for a recent review. However, Solove (2021) challenges whether this paradox exists based on a critique of privacy paradox studies, which involve behaviors in very specific contexts, whereas self-reported privacy concerns are much more general in nature. Even those who support the presence of the paradox may attribute it to different mechanisms and thus draw different policy implications. A direct interpretation of the paradox is that consumers' privacy concerns are not credible, thus discrediting arguments for regulating data privacy,[2] while another argument attributes the paradox to consumers' resignation from privacy because privacy self-management is too overwhelming a task for consumers in the data economy age.[3]

These controversies motivate several key questions for this paper: 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? An even more

[1] See Chen et al. (2021) for an extensive report of data sharing in the booming digital economy. Ouyang (2021) provides empirical evidence for data sharing facilitating credit provision to the underserved.

[2] See Ben-Shahar (2016), Cooper and Wright (2018), and Fuller (2019) for extensive discussions in the legal literature.

[3] See Hoffmann, Lutz and Giulia Ranzini (2016).

demanding issue regards the implications of the privacy paradox for consumer welfare. Addressing this issue requires an analysis of not only consumers' privacy costs in data sharing but also their benefits. Such a joint evaluation is still missing from the literature. The data privacy paradox thus provides an entry to understanding how consumers trade off their privacy concerns with data-sharing needs to satisfy their digital demands, the foundation of the data economy.

We address these issues by analyzing data-sharing choices and use of digital services by a sample of Alipay users in a very relevant setting for the digital economy. Alipay is a highly popular payment and lifestyle platform in Chinas that has more than 900 million active users. In addition to its widely used payment system, it 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 authorize, at the initial entry, sharing of certain personal data with the mini-program. The requested data sharing is usually justified by the provided services. Thus, it varies across mini-programs from innocuous information, such as the user's nickname on the Alipay platform, to highly sensitive information, such as the national ID number and credit score. This exchange of personal data for digital services is typical on digital platforms.

Specifically, we conduct a survey of Alipay users about their data privacy concerns and then match their survey responses with rich administrative data about their data-sharing choices with mini-programs on the Alipay platform to analyze how their choices are related to their stated privacy concerns and their actual use of these mini-programs. 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 conducted 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 authorized data sharing with 11.2 mini-programs, the "concerned" users authorized sharing with 11.5, and the "very concerned" users authorized sharing with 11.3. During the 17-month period after the survey, from August 2020 to December 2021, the "unconcerned" users authorized data sharing with 22.5 mini-programs, the "concerned" users authorized data sharing with 24.6, and the "very concerned" users authorized sharing with 23.8.

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 authorizations 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 puzzling patterns confirm the data privacy paradox in a setting that is highly relevant to the digital economy. To make the paradox even more puzzling, the "concerned" and "very concerned" users, each month, authorized data sharing with more mini-programs in the post-survey period than in the pre-survey period.

Our study is immune from the aforementioned critique of Solove (2021) because our survey questions specifically targeted the respondents' concerns about data sharing with Alipay's mini-programs and are matched by administrative data specific about their data sharing with those mini-programs. To further validate the self-stated privacy concerns from the survey, we also examine the respondents' propensity to take two privacy-seeking actions in Alipay: 1) canceling previously authorized data sharing with mini-programs, and 2) changing Alipay's default privacy settings, which tend to make a user's information visible to other Alipay users. We find that the privacy concerns stated in the survey responses are positively associated with these privacy-seeking actions, thus confirming that the survey responses capture the respondents' specific privacy concerns about data sharing with mini-programs on Alipay.

What causes privacy-concerned Alipay users to ignore their privacy concerns in authorizing data sharing? In our sample, the respondents, on average, rejected 26.5% of the data-sharing requests from mini-programs. This nontrivial rejection rate indicates that they have not resigned from privacy. The privacy literature has also suggested a number of psychological and behavioral

factors to explain the privacy paradox, including consumers' ignorance about the consequences of data sharing (Kesan, Hayes and Bashir, 2015), 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 these studies, our analysis uncovers a new finding, which connects privacy concerns to demands for digital services—users with stronger privacy concerns also use their authorized mini-programs more frequently and more extensively. This curious finding suggests that privacy-concerned users have greater demands for digital services, which in turn may offset or even dominate their privacy concerns. Thus, the observed data sharing by privacy-concerned users may not contradict their altitudes, but rather reflect the trade-off between data privacy and digital demand.

To further establish the positive relationship between privacy concerns and digital demands, we also examine a hypothesis that more-active users of mini-programs are more likely to cancel their data sharing 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. Nevertheless, by using two different measures of user activeness and controlling for various user characteristics and mini-program fixed effects, we confirm that more-active users of mini-programs in our survey sample are indeed more likely to cancel data sharing with mini-programs.

To analyze how privacy concerns may grow over time across different consumers, we also examine a salient incident on January 3, 2018, triggered by Alipay, which greatly stimulated Alipay users' awareness of the need to protect their data privacy. In response to this incident, heavy users in a representative sample of 100,000 users—randomly drawn from the full set of active Alipay users—are more likely than light users to cancel data sharing with mini-programs.[4] Thus, this event study establishes a dynamic effect that heavy use of digital services makes users more susceptible to privacy concerns.

[4] This representative sample, along with an alternative, behavior-based measure of privacy concerns through users' changes to their Alipay default privacy settings, also allow us to show the robustness of our key findings to the survey sample, which are biased toward more-active users.

Overall, our study not only confirms the data privacy paradox in a highly relevant setting for the digital economy, but also uses the paradox as an entry to highlight a key tension of the data economy—both consumers' privacy concerns and digital demands grow with the deepening of the digital economy. Thus, our study fills an important gap in the literature by establishing the intricate, joint dynamics of data privacy and digital demand, which represent the cost and benefit sides of consumers' data sharing. Analyzing either side in isolation may generate misleading implications. For example, considering consumers' privacy concerns in isolation from their digital demands may lead to a misleading interpretation of the data privacy paradox as a contradiction between self-stated altitudes and actual behaviors.

Our study contributes to a better understanding of both the costs and benefits of data sharing. On the cost side of data sharing, our analysis highlights that consumers' privacy concerns grow with their use of digital applications and the accumulation of their personal data shared with digital service providers. This finding not only confirms an upward shift in privacy concerns, (e.g., Goldfarb and Tucker, 2012), but more importantly highlights an essential characteristic of data privacy—it is not simply an isolated preference as sometimes suggested in policy discussions, but rather a type of risk induced by data sharing in the process of using digital applications. Economists have long emphasized that the value of privacy is associated with economic consequences of hiding one's private type (Stigler, 1980; Posner, 1981). Such economic consequences depend on the contexts in which specific consumer data are shared with specific firms or parties. While data sharing allows sellers to better match consumers with their preferred products, it may also expose consumers to potential price discrimination by sellers (Taylor, 2004; 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).

Several studies estimate how much consumers value their data privacy. Acquisti, John and Lowenstein (2013) adopt a field experiment to show that consumers' privacy valuations are sensitive to contextual and nonnormative factors. Tang (2020) uses a natural experiment through consumers' fintech loan applications, which require loan applicants to provide certain personal information. Lin (2022) uses an experimental setting to differentiate instrumental privacy preferences, which are generated from payoffs related to a consumer's type being revealed, from

intrinsic privacy preferences, which are independent of any economic payoffs. Our analysis not only suggests that a consumer's privacy valuation depends on the context of data sharing, but more importantly highlights a sharp characteristic that it increases with accumulated data sharing.

On the benefit side of data sharing, the literature on the data economy (e.g., Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020; Cong, Xie and Zhang, 2020), has highlighted two important features of data sharing—nonrivalry and increasing returns to scale, which imply that data shared by consumers allow digital service providers to provide more powerful services and thus further increase consumers' digital demands. The increasing trends in both costs and benefits of data sharing make it possible to explain the rising trend in Alipay users' data-sharing authorizations in our sample. Nevertheless, if privacy concerns rise more rapidly than digital demands in the future, privacy concerns may eventually limit the growth of the data-sharing economy. It is thus vital to strengthen privacy protections for ensuring the full promise of the data-sharing economy.[5]

The paper is organized as follows. Section I provides the institutional background of Alipay users' data sharing with mini-programs. Section II describes the survey of Alipay users and reports summary statistics. We analyze the data privacy paradox in Section III and further examine the relationship between privacy concerns and digital demands in Sections IV and V. Section VI reports robustness analysis, and Section VII concludes the paper. We also provide an Online Appendix for additional analysis.

I. Institutional Background

This section provides background information about the Alipay platform and the data-sharing arrangement between Alipay users and third-party mini-programs in Alipay. Alipay is a mobile application, which started by offering online payment services and has grown 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 Chinese population. In addition to providing a wide range of financial services, such as digital payments, micro-loans, credit cards, insurance, and wealth

[5] 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; 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).

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 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 the sharing of certain information necessary for its service. The requested information varies across mini-programs.[6] Some information is innocuous, such as the user's nickname, while other information is more sensitive, such as one's national ID number or credit score. A user has two choices: agree to or reject the data-sharing request. Only after the user authorizes the request is she allowed to use the services offered by the mini-program. This setting makes the data-sharing authorization an explicit exchange of personal data for digital services.[7] This data-sharing authorization lasts for a certain period; at the expiration of the period, the mini-program asks the user to reauthorize the data sharing at her next entry into 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 of a sample of Alipay users.

[6] 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. 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) credit score, which helps to evaluate the trustworthiness of the user and determine whether a deposit is required; and 3) identification information, such as real name, phone number, and national ID number. After a user authorizes sharing of the requested data, the user can use Hellobike's shared bikes. Figure A1 in the Online Appendix provides three additional examples. 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.

[7] Our setting provides a simpler trade-off than the data-sharing decisions faced by consumers with many public websites. As a mandate of the GDPR, public websites give users an option to opt in or out of their collection of user data. In a typical arrangement, if a user allows a website to collect her data, the website can use the user data to provide personalized services. Even if the user opts out of the data collection, she may be still able to use the website, but the services are not personalized. Thus, for the user, sharing personal data brings the gain of personalized services as opposed to nonpersonalized services. In our setting, an Alipay user cannot use any service from a mini-program unless she authorizes data sharing.

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 the public, whether to allow connections 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 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 privacy-seeking action to validate our survey-based measure of privacy concerns.

II. Survey and Administrative Data

In this section, we first describe the survey of Alipay users about their privacy concerns and then report summary statistics of data-sharing authorizations and other administrative data of the survey respondents.

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, 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, a question asked, "*Have you ever used mini-programs in Alipay?*" Only those respondents who answered "yes" to this question advanced 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 Alipay, accounting for 76% of all respondents.[8] These 10,875 respondents are the main sample for our analysis.

[8] Figures A2–A5 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

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 analyze 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 Online 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. 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 specific data-sharing choices in analyzing the data privacy paradox. As this latter survey question is directly related to our analysis of data sharing with 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

way (Figure A2). The geographical distribution of the respondents across the provinces in China lines up well with the distribution of the population (see Figure A4), 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%.

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. As only 5% of the respondents selected "others," it appears that 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.

B. Administrative Data

A key strength of our study is that we have access to the respondents' extensive administrative data inside Alipay, which allows us to examine how their privacy concerns are related to their actual data-sharing choices and use of the authorized mini-programs. Table 2 reports summary statistics of the key variables. Panel A covers three sets of user information: general profile, data sharing with mini-programs, and monthly use of mini-programs.

For general information, also known as user profile, we have access to information on 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 (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 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. 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 value 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. These statistics imply the respondents, on average, rejected 26.5% of the data-sharing requests. This nontrivial rejection rate shows that the respondents have not resigned from privacy by simply accepting all data-sharing requests.

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 Alipay users cancel data-sharing authorizations relatively infrequently.

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.[9] The information has 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 higher.

III. The Data Privacy Paradox

By combining the respondents' survey responses and administrative data, we examine how their data-sharing choices are related to their privacy concerns. Specifically, we test 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 then 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.

[9] Alipay did not systematically record data on users' activities related to mini-programs before 2019. As a result, we cannot cover these variables before 2019.

For simplicity, we suppose that the cost for user i to share the requested 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 captures the contribution of the mini-program, and ϵ_{ij} is a noise component independent across the user and mini-program pair. The user component c_i is larger if the user is more vulnerable to targeted advertising or more sensitive to price discrimination by firms. The mini-program component c_j is larger if the mini-program requests more-sensitive data and is less reputable in privacy protection.

Similarly, we linearly decompose the benefit to the user from using the mini-program, b_{ij}, 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 independent across the user and mini-program pair. The user component b_i is higher if the user is more receptive to digital services, and the mini-program component b_j is larger if the mini-program offers more powerful services.

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

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

After controlling for the mini-program's characteristics, the authorization choice is driven by the user's characteristics through the term $b_i - c_i$. We start with a baseline case, in which b_i and c_i are independent. That is, the user's privacy concerns are not related to her appreciation for digital services. Consistent with this case, commentators in policy discussions of data privacy often view privacy concerns in isolation of consumers' demands for digital services. Consequently, a user with stronger privacy concerns (i.e., larger c_i) is less likely to authorize data sharing, as summarized by the following hypothesis:

Hypothesis 1: All 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, a user's digital demands may offset her privacy concerns, thus making her data-sharing choices insensitive to her privacy concerns. We will also examine this possibility in our later analysis.

B. Privacy Concerns and Data Sharing

In Figure 1, we compare the number of data-sharing authorizations by Alipay users who 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 data sharing with 22.5, "concerned" users visited 32.8 and authorized date sharing with 24.6, while "very concerned" users visited 33.4 and authorized date sharing with 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 the pre- and post-survey periods both contradict Hypothesis 1 that privacy-concerned users are more reluctant to authorize data sharing.

As users also differ in other dimensions beyond privacy concerns, we adopt a cross-sectional regression at the user level to control for various user characteristics:

$$Y_i = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i + a_4\, Digital\ Experience_i + \delta_i + \epsilon_i\,, \quad (1)$$

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 0 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 3 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, although a_2 becomes 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 data sharing 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. Different from the user-level regression, this regression allows us to include mini-program fixed effects γ_j, which control for the heterogeneity across mini-programs.

Table 4 reports the analysis at the user-mini–program level, with Panel A covering the pre-survey sample and Panel B covering the post-survey sample. Even after controlling for mini-program fixed effects, the results are very similar to that from the user-level analysis. 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. 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 3 and 4 reject Hypothesis 1 and instead confirm the data privacy paradox that the respondents' data-sharing authorizations are not negatively related to their privacy concerns. This finding contradicts the common wisdom that privacy-concerned users are more reluctant to share personal data.

We have also explored how the data privacy paradox may vary across users with different characteristics. In Table A1 of the Online Appendix, we expand the regression at the user-mini–program level 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 higher. 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. [10] Interestingly, the data privacy paradox is not simply a phenomenon among users with

[10] We classify a mini-programs as seemingly addictive if its description contains relevant key words, such as "game," "lottery," or "red envelope."

low education and thus insufficient knowledge of data privacy. To the contrary, it is more severe among more educated users. There is also no evidence for the data privacy paradox being more severe among users with weaker self-controls, suggesting that it is a general phenomenon beyond a particular group with insufficient digital knowledge or behavioral weaknesses.

In Figure 2, we also depict the monthly time series of the average monthly data-sharing authorizations by the three groups of Alipay users with different levels of privacy concerns. Despite the substantial fluctuations from month to month, there is a visible increasing trend across the three groups. The gaps among the three groups are small in the pre-survey period, before July 2020. In the post-survey period, after July 2020, the increases in the number of authorizations by the "concerned" and "very concerned" groups become even more pronounced than that by the "unconcerned" group. As we have shown in our previous regression analysis in Tables 3 and 4, these differences in the post-survey period are statistically significant. Taken together, Figure 2 shows that the concerned and very concerned groups experienced larger increases in data-sharing authorizations in our sample. These greater increases over time are even more puzzling, adding a time-series dimension to the data privacy paradox.

C. Validating Survey-Based Privacy Concerns

It is tempting to argue that the data privacy paradox may simply reflect the 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 (e.g., Bertrand and Mullainathan, 2001). This argument also reflects the critique made by Solove (2021) that the self-reported privacy concerns from the surveys in privacy paradox studies may not correspond to the observed behaviors.

To validate the survey-based measure of privacy concerns, we take advantage of our extensive administrative data to examine whether the survey-based measure is positively correlated with actions taken by the respondents to protect their data privacy other than the initial authorization of data sharing with mini-programs. We 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 take these actions to protect their privacy.

We again organize our analysis at both the user level and user-mini–program level. For the user-level analysis, we adopt the regression specified in Equation (1) but replace the dependent variable with 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.[11] 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 5 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

[11] Alipay started to record these variables at different points of time, leading to their different periods of measurement.

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 5, we further expand the analysis to the user-mini–program level for cancellation of data sharing. The advantage of the analysis at the user-mini–program level is that we can control for mini-program fixed effects, which allows 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 user and mini-program pair 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, Table 5 confirms that the survey-based measure of privacy concerns is positively related to actions taken by Alipay users to protect their data privacy, thus validating the survey-based measure of privacy concerns. In particular, it shows that the critique of Solove (2021) does not apply to our analysis.

D. Determinants of Data Sharing in Survey

In the survey, we also 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 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.*

Each of these statements presents a potential mechanism that helps Alipay users overcome their privacy concerns in data sharing. The first statement considers that users' trust of Alipay's privacy protection might dominate their privacy concerns. 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. To some extent, this statement reflects a general argument that privacy might be impossible under the attack of increasingly powerful digital technologies in the data economy age. The third statement represents a key consideration for our analysis that the decision to authorize 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 considers 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).

To save space, we report the responses to these statements in Table A2 of the Online Appendix. We split the respondents into two groups, one with "concerned" and "very concerned" respondents and the other with "unconcerned" respondents. For a statement to explain the lack of any difference in the observed data-sharing authorizations between privacy-concerned and unconcerned respondents, we expect the statement to be more agreed to by "concerned" users than "unconcerned" users. Interestingly, the survey responses show that only the third statement, "*I have to share my personal data in exchange for digital services even though I am concerned by my data privacy,*" is agreed to more often by the concerned group (64%) than the unconcerned group (55%). Thus, 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

How shall we explain the lack of a negative relationship between privacy concerns and the number of data-sharing authorizations? Recall the conceptual framework described in Section III.A: it is possible to explain this paradox if a user's privacy concerns about sharing personal data with a mini-program are positively correlated with the benefits from using it. In this section, we examine how privacy concerns are related to digital demands.

A. Privacy Concerns and Use of Digital Services

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, as implied by an intuitive argument that a user with greater demands for digital services is likely to use their authorized mini-programs more intensively and more frequently. Common wisdom suggests that privacy concerns may deter users from digital applications and thus motivates the following hypothesis:

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

We examine this hypothesis by using the following regression specification:

$$Y_{ijt} = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_{it} + a_4\, Digital\ Experience_{it} + \delta_i + \mu_j + \theta_t + \varepsilon_{ijt}, \quad (3)$$

where 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; Age_{it} and $\text{Digital Experience}_{it}$ 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 6 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 intensively. Taken together, the regression results show a positive and robust relationship between digital demands and privacy concerns, firmly rejecting Hypothesis 2.

This finding of privacy-concerned respondents also having greater digital demands implies that their larger number of data-sharing authorizations does not necessarily imply that their self-stated privacy concerns are inconsistent with their actual behaviors, as is often attributed to the data privacy paradox. Instead, it suggests that the observed data-sharing choices may reflect a trade-off between the respondents' privacy concerns and digital demands. This trade-off makes their data sharing insensitive or even positively correlated to their privacy concerns.[12]

B. Digital Demand and Cancellation

How can privacy-concerned users have greater demands for digital services? The economic literature has long emphasized that privacy may not be a primitive preference that is independent of economic contexts, and, instead, is associated with economic consequences of keeping one's private type from being revealed to others (e.g., Stigler, 1980; Posner, 1981). Such economic consequences depend on the contexts through which consumer data are shared with firms and service providers. It is particularly important to recognize that consumers' privacy concerns interact with their demands for digital services.

On one hand, consumers' privacy concerns intensify with the quantity of personal data shared with digital service providers, as implied by various theories of privacy concerns. The privacy cost of personal data being hacked by or leaked to unauthorized parties (e.g., Fainmesser, Galeotti and Momot, 2019) is increasing with the shared data. More data being shared also allows digital service providers to more effectively price discriminate users (e.g., Taylor, 2004; Acquisti and Varian,

[12] Similarly, in a study of stock trading motives based on both survey and behavioral data, Liu et al. (2022) find that behavior-based measures of trading motives are also related to multiple factors, which may complicate any test of a specific trading motive.

2005), and more effectively target users' personal vulnerabilities (e.g., Liu, Sockin and Xiong, 2020). On the other hand, more shared data by a consumer allows digital service providers to better infer the consumer's preferences and thus provide more powerful personalized services, as implied by increasing returns to scale of data sharing (e.g., Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020; and Cong, Xie and Zhang, 2020). Thus, despite that privacy costs are increasing with shared data, consumers may continue to share their personal data because the benefits from data sharing also grow with shared data.

To firmly establish the notion that privacy concerns grow with digital demands, we further examine 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: All 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 them 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 7 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 the measurement period of the cancelation rate that starts 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% increase 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 7 shows the relationship between the user's activeness and the propensity to cancel a mini-program at 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 previously canceled at least one data-sharing authorization.

Taken together, Table 7 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 thus confirms the key notion that users with greater digital demands tend to be more concerned about data privacy. We establish this notion without using the survey-based privacy concerns.

V. Growing Privacy Concerns

The notion that consumers' privacy concerns intensify with the data that they share with digital service providers implies that their privacy concerns grow over time with their use of digital applications. Figure 3 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.

How do privacy concerns grow across users with different digital demands? We take advantage of a salient incident to examine this question. 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 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 public awareness of data privacy issues and led to a spike in Alipay users' cancellation of data sharing with mini-programs, as shown by Figure A6. 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 were more likely to cancel data sharing with mini-programs.

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

$$Daily\ Cancellation\ Dummy_{i,t} = \alpha_0 + \sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{H,\tau} \cdot Heavy\ User_i \cdot \mathbb{I}(t=\tau)$$

$$+ \beta_{H,6} \cdot Heavy\ User_i \cdot \mathbb{I}(t \geq 6) + \sum_{\substack{\tau=-5, \\ \tau\neq-1}}^{5} \beta_{L,\tau} \cdot Light\ User_i \cdot \mathbb{I}(t=\tau)$$

$$+ \beta_{L,6} \cdot Light\ User_i \cdot \mathbb{I}(t \geq 6) + \delta_i + \varepsilon_{i,t}, \quad (4)$$

where t corresponds to the number of days after the incident on January 3, 2018, $Daily\ Cancellation\ Dummy_{i,t}$ is a dummy variable indicating whether user i has canceled at least one mini-program during the day t, $Heavy\ User_i$ 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 represents individual fixed effects, and $\varepsilon_{i,t}$ is random error that varies across individuals and over time.

This event occurred before our main survey sample. To avoid any potential survival bias, we have constructed a random sample of 100,000 Alipay users, who are randomly selected from all active Alipay users. We report their summary statistics in Table A3 of the Online Appendix. The users in this random sample have an average age of 36.6 years and an average digital experience of 60.7 months, suggesting that this random sample tends to be older and have shorter digital experience. Users in this random sample also authorized data sharing with fewer mini-programs and were less active in using their authorized mini-programs relative to users in the survey sample.

We use this random sample to estimate the regression specified in Equation (4). Panel A of Figure 4 depicts the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients. Consistent with Hypothesis 4, heavy users of mini-programs are significantly more responsive to the incident, showing stronger privacy concerns through their greater propensity to cancel data sharing with mini-programs. This response is temporary, possibly due to the quick actions taken by Alipay and the incident eventually going off social 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 A7.

Like before, one might argue that the greater propensity of heavy users to cancel data sharing reflects their better knowledge of how to cancel authorizations in the Alipay application rather than their stronger privacy concerns stimulated by the incident. To address this argument, 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 4 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. 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 A7. 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 confirms that users with greater digital demands become more concerned about data privacy after the incident. This evidence reinforces the notion that concerns about data privacy are positively correlated with demands for digital services. In the process of using digital applications, a consumer gradually accumulates personal data with digital service providers. The accumulated data expose the consumer to greater privacy risks in that the data might be hacked by or leaked to unauthorized parties and the consumer may face more severe price discrimination or targeted advertising by sellers.

The growing privacy concerns, especially among users with greater digital demands, make the increasing trend in data-sharing authorizations shown by Figure 2 even more puzzling. To explain this trend, we again need to recognize the dynamics of consumers' digital demands. As we discussed before, more data sharing allows digital service providers to provide services that are more powerful, leading to increasing returns to scale of data sharing and thus an increasing trend in consumers' digital demands. Even though our data cover only a one-year period of the Alipay users' use of their authorized mini-programs, Figure A8 in the Online Appendix indeed shows a pronounced increasing trend during the sample period. With the costs and benefits of data sharing both increasing over time, consumers may authorize more data sharing over time, despite their growing privacy concerns. However, it is also important to recognize that if privacy concerns rise more rapidly than digital demands in the future, privacy concerns may eventually limit the growth of the data-sharing economy. It is thus vital to ensure privacy protections and manage consumers' privacy concerns below their digital demands.

VI. Robustness

Our survey sample tends to include more-active users, 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 use the random sample of 100,000 Alipay users to verify the key results from our survey sample. As reported in Table A3, the random sample is indeed less active in using mini-programs than the survey sample.[13] 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 Alipay's default privacy settings, as a behavior-based measure of the user's 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.[14]

In Table A4 of the Online Appendix, 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 3. Interestingly, the more concerned users 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, indicating that the data privacy paradox is 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 6. We again find that in the random sample, more-concerned users tend to use their authorized mini-programs more frequently and more intensively 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 7. We again observe that the cancellation rate is significantly and

[13] 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.

[14] 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.

positively correlated with user activeness. Taken together, we confirm that the three key results of our analysis are robust in the representative random sample of Alipay users.

VII. Conclusion

In this paper, we combine both survey and administrative data to examine how data sharing of Alipay users with third-party mini-programs in Alipay is related to their privacy concerns. Even though one would expect users with stronger privacy concerns to be more reluctant to share personal data, we find that privacy-concerned users authorize more, rather than less, data sharing than unconcerned users, thus confirming the data privacy paradox in a setting highly relevant to the booming digital economy.

Instead of attributing this paradox to either an unreliable survey-based measure of privacy concerns, Alipay users' resignation from privacy, or their behavioral biases in making data-sharing choices, we uncover a new finding that privacy-concerned users use their authorized mini-programs more frequently and more intensively than unconcerned users. This finding offers a new explanation to the data privacy paradox through the greater demands of privacy-concerned users for digital services, which may dominate their privacy concerns about data sharing. Furthermore, our analysis highlights the joint dynamics of the users' privacy concerns and digital demands in determining their data sharing—not only do their privacy concerns grow with their use of mini-programs but so do their demands for digital services—leading to more data sharing over time, despite their growing privacy concerns.

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.

Ben-Shahar, O. (2016). Privacy is the New Money, Thanks to Big Data. *Forbes*.

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.

Cooper, J. C., & Wright, J. (2018). The Missing Role of Economics in FTC Privacy Policy. *The Cambridge Handbook of Consumer Privacy*, 465.

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

Farboodi, M., & Veldkamp, L. (2020). A Model of the Data Economy. Working Paper, MIT and Columbia.

Fuller, C.S. (2019). Is the Market for Digital Privacy a Failure? *Public Choice*, 180(3), 353–381.

Hoffmann, C. P., Lutz, C., & Ranzini, G. (2016). Privacy Cynicism: A New Approach to the Privacy Paradox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(4).

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.

Kesan, J. P., Hayes, C. M., & Bashir, M. N. (2015). A Comprehensive Empirical Study of Data Privacy, Trust, and Consumer Autonomy. *Indiana Law Journal*, 91, 267–352.

Lin, T. (2022). Valuing Intrinsic and Instrumental Preferences for Privacy. *Marketing Science*, forthcoming.

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.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors. *Journal of Consumer Affairs,* 41(1), 100–126.

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

Posner, R. A. (1981). The Economics of Privacy. *American Economic Review*, 71(2), 405–409.

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.

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd Generation E-commerce: Privacy Preferences Versus Actual Behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, 38-47.

Stigler, G. J. (1980). An Introduction to Privacy in Economics and Politics. *Journal of Legal Studies*, 9(4), 623-644.

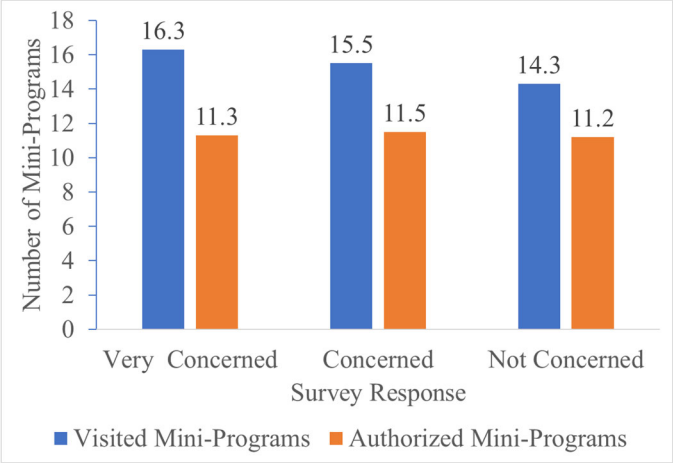
Tang, H. (2020). The Value of Privacy: Evidence from Online Borrowers. Working Paper, London School of Economics.

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

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 through 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

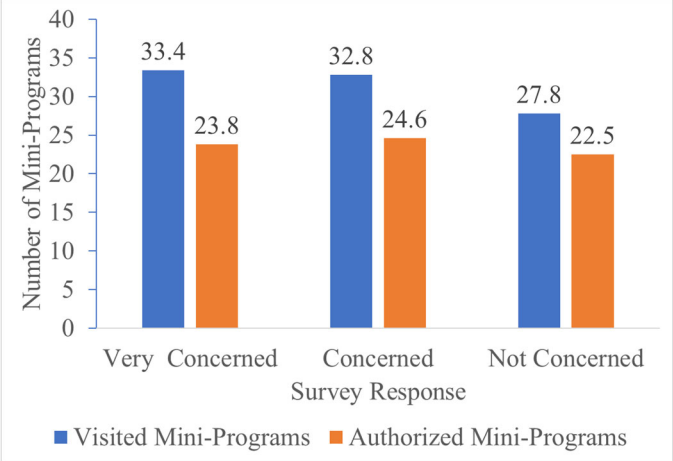


Figure 2: Time Trend in Data-Sharing Authorizations

This figure depicts the monthly time series of the average number of data-sharing authorizations of Alipay users in three groups based on their self-stated privacy concerns. The vertical dash line indicates July 2020, the survey date.

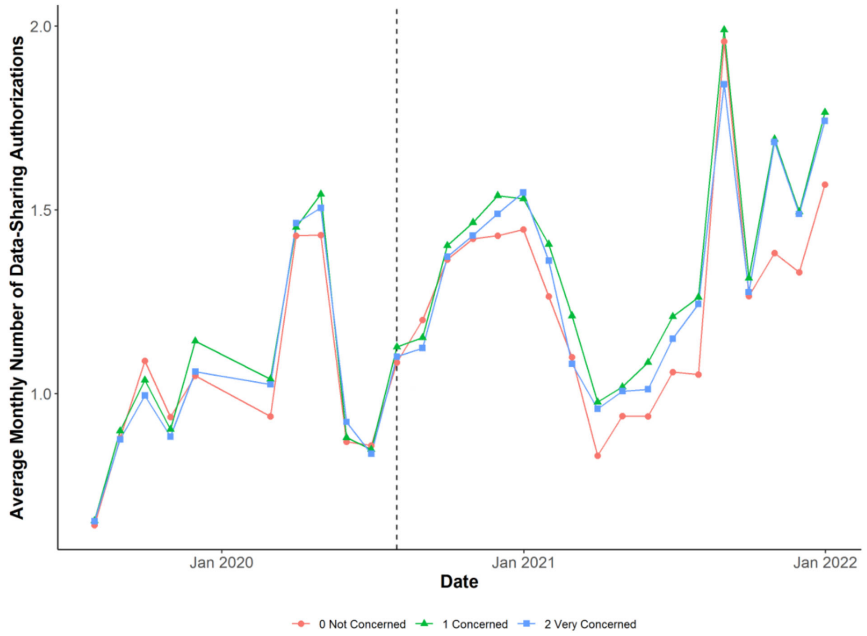


Figure 3: Digital Experience and Privacy Concerns

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 experiences, measured by the length of time since a user registered on Alipay. For each group, we also show the 95% confidence band of the mean estimate.

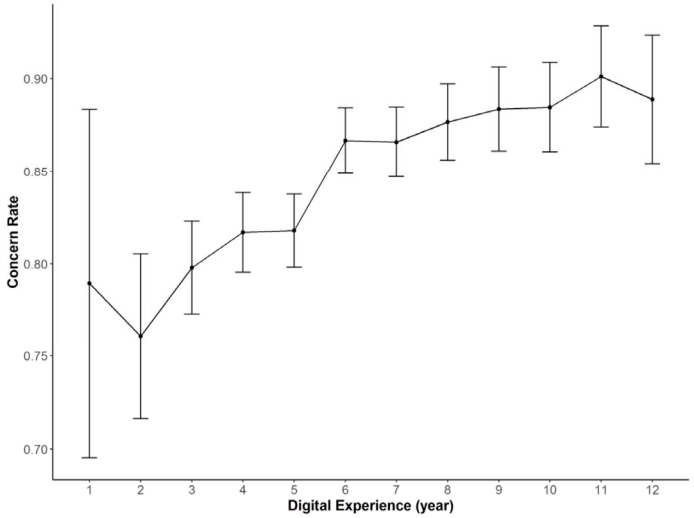
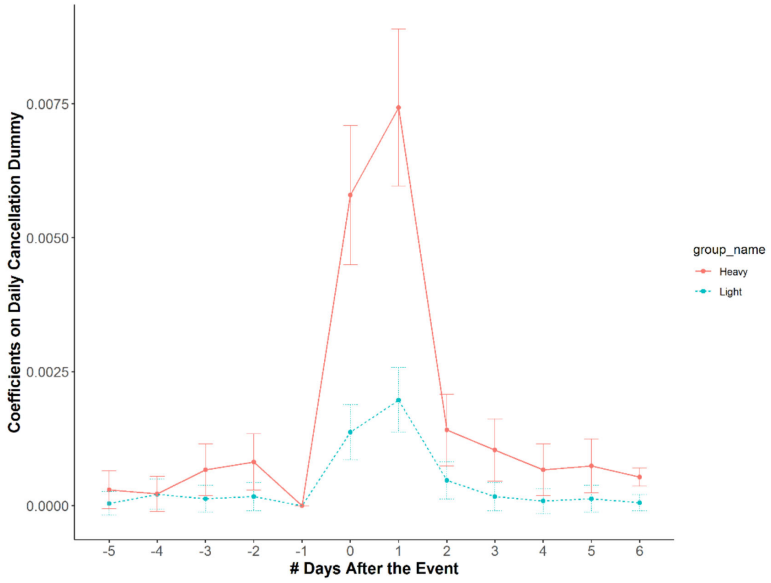


Figure 4: Activeness and Response to the 2017 Footprint Report Incident

The figures plot the $\beta_{H,\tau}$ and $\beta_{L,\tau}$ coefficients estimated by 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 canceled data sharing with at least one mini-program before November 30, 2017, in the random sample. The data are at individual and daily levels. 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

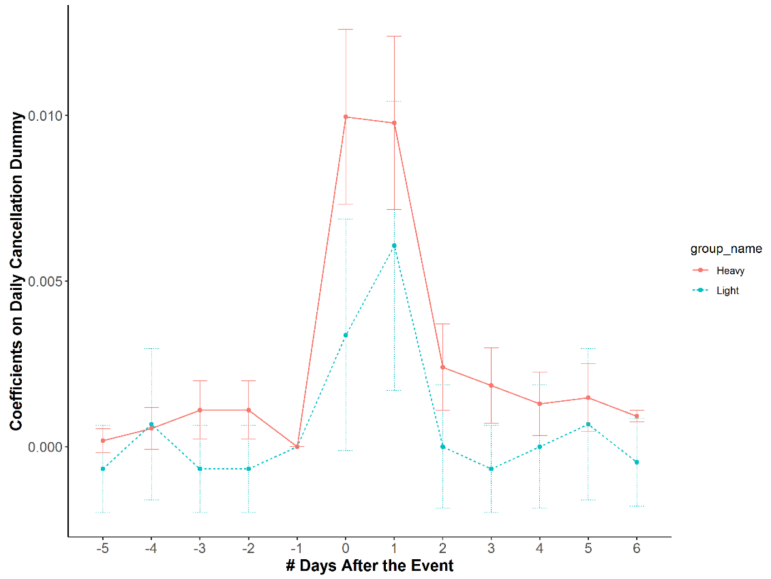


Table 1: Responses to Selected Survey Questions

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 choice)			
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. Panel 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 first 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 through July 2020 and the post-survey period of August 2020 through 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 through July 2020, including the number of active days, the number of uses, the number of launches, and the 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 higher, 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 (1)	Concerned (2)	Very Concerned (3)	Difference (2) – (1)	Difference (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: 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 through July 2020, while Panel B reports results for the post-survey period from August 2020 through 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 4: 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 through July 2020, while Panel B reports results for the post-survey period from August 2020 through 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 5: Validating Survey-Based Privacy Concerns

This table reports how the survey-based measure of privacy concerns is related to privacy-seeking actions, including canceling data-sharing authorizations with mini-programs and changing Alipay's default privacy settings. *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 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 through July 2020. In columns (3)–(4), the dependent variable is a dummy that indicates whether a user has changed Alipay's default privacy settings the period of May 2017 through April 2020. Panel B shows results for regressions at the user-mini–program level. In each pair of user-mini–program and existing data-sharing authorization, the dependent variable is a dummy that indicates whether the user canceled the authorization in July 2019 through July 2020. We cluster the standard errors at the user level. 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: Analysis at User-Mini–Program Level

	$Canceled\ Dummy_{ij}$	
	(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 *R*2	0.0001	0.107

Table 6: 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 through 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 $R2$	0.0002	0.119	0.0002	0.096	0.0001	0.086	0.0001	0.078

Table 7: Digital Demand 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 through 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 regressions at the user mini–program level, 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 $R2$	0.012	0.005	0.027	0.014

Panel B: Regression at User-Mini–Program Level

	$Canceled\ Dummy_{ij}$			
	(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 $R2$	0.127	0.127	0.172	0.170