Data Privacy and Digital Demand[*]

Long Chen, Yadong Huang, Shumiao Ouyang, Wei Xiong

September 2023

Abstract

We combine survey and behavioral data to analyze consumers' data-sharing choices in a pertinent context where they exchange personal data for digital services. Intriguingly, we find that respondents with stronger privacy concerns authorize more data sharing, underscoring the data privacy paradox. Different from conventional explanations of this paradox, such as inconsistent survey responses, privacy resignation, or inherent behavioral biases, we uncover a novel mechanism: the deepening of the data economy amplifies consumers' demand for digital services, even as their privacy concerns heighten. This suggests a nuanced market dynamic. While privacy concerns have been on the rise, the benefits from increasingly efficient digital services, fueled by consumer data, may offset or even dominate these concerns, encouraging continued data sharing.

[*] Long Chen and Yadong Huang are affiliated with the Luohan Academy, Shumiao Ouyang is affiliated with the University of Oxford, and Wei Xiong is affiliated with Princeton University. We appreciate comments and suggestions made by Alessandro Acquisti, Laura Brandimarte, Avi Goldfarb, Zhuang Liu, 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.

Consumers' sharing of personal data is the backbone of the thriving data economy, a potential cornerstone for the broader macroeconomy as identified by recent theoretical models of Jones and Tonetti (2020), Farboodi and Veldkamp (2020), and Cong, Xie, and Zhang (2020).[1] Yet, the rise of digital platforms and AI systems like Facebook, Amazon, and ChatGPT, which rely on extensive user data, has intensified long-held concerns about data privacy. This shift in consumer attitudes is strongly reflected by the recent enactments of the European Union's General Data Privacy Regulation (GDPR) in 2018, California's Consumer Privacy Act (CCPA) in 2020, and China's Personal Information Protection Law (PIPL) in 2021. At the core of these regulations is the belief that users can maintain privacy by withholding consent; if concerned about data privacy, they can simply opt not to share. However, comprehensive insights into consumers' data-sharing choices remain scant. With AI technologies advancing at a breakneck pace using personal data, grasping the evolving consumer privacy calculus becomes pivotal for shaping the future of data-driven innovation.

A salient observation is the "data privacy paradox." Numerous surveys and studies, including works by Spiekermann, Grossklags and Berendt (2001), Gross and Acquisti (2005), Norberg, Horne and Horne (2007), and Athey, Catalini and Tucker (2017), indicate a dichotomy: consumers express privacy concerns but often share personal data, sometimes for minor rewards. Acquisti, Brandimarte, and Loewenstein (2020) offer a recent overview. This paradox is frequently attributed to consumers' confusion or irrationality when consenting to data sharing. If true, even robust regulations like GDPR and CCPA might falter in safeguarding consumers. Hence, discerning how consumers decide on data-sharing in realistic contexts is paramount.

Against this backdrop, we use a unique dataset that combines both consumers' privacy attitudes and data sharing choices from a major digital platform to address several pivotal questions. Firstly, does the data privacy paradox manifest in realistic settings where consumers contemplate sharing data with digital service providers? This is not a trivial query. Solove (2021) contests the very existence of the paradox by critiquing its studies. These investigations often focus on specific contexts, contrasting starkly with the more generalized nature of self-reported privacy concerns. Secondly, if a gap exists between declared privacy concerns and actual data-sharing behaviors,

[1] See Chen et al. (2021) for an extensive report of data sharing in the booming data economy.

does this suggest consumers are ill-equipped to make informed decisions? Lastly, what underlying factors influence consumers' privacy concerns and data-sharing choices?

We delve into these issues by examining the data-sharing choices and digital service usage of a group of Alipay users. Alipay, a renowned payment and lifestyle platform in China, boasts over 900 million active users. Beyond its ubiquitous payment system, it houses over two million third-party mini-programs, essentially lightweight apps functioning within Alipay, providing a plethora of digital services. To access a mini-program, users must, upon initial entry, consent to sharing specific personal data. This data exchange, generally justified by the services offered, can range from benign details, like a user's Alipay nickname, to more sensitive information such as national ID numbers or credit scores. Such exchanges exemplify the digital platform landscape.

Our study involved surveying Alipay users about their data privacy concerns. We then paired their survey responses with comprehensive administrative data detailing their interactions with Alipay's mini-programs. This aimed to discern the relationship between their professed privacy concerns, their actual data-sharing decisions, and their engagement with these mini-programs. Given the diverse nature of Alipay's mini-programs in terms of service value and information sensitivity, this environment presents a prime opportunity to analyze how users balance privacy concerns and digital service needs.

We undertook a survey of Alipay users, encompassing 12 questions centered around their preferences and concerns about sharing data with Alipay's mini-programs. We collected responses from 14,250 Alipay users. Addressing their level of concern regarding data privacy when sharing personal data with mini-programs: 46% indicated significant concern, 39% expressed moderate concern, and a mere 15% felt no concern. In our main sample period from July 2019 to July 2020, data showed that "unconcerned" users, on average, consented to share data with 11.2 mini-programs. In contrast, "concerned" users shared with 11.5, and the "very concerned" segment shared with 11.3 mini-programs.

Surprisingly, even with the intuitive expectation that users with pronounced privacy concerns would be more conservative in their data sharing, both "concerned" and "very concerned" user groups, on average, allowed data sharing with nearly identical numbers of mini-programs as their "unconcerned" counterparts. This consistency in behavior persists even when accounting for user characteristics, such as digital familiarity, age, gender, and city of residence, as well as inherent characteristics of the mini-programs. This counterintuitive finding underscores the presence of a privacy paradox, especially in a context that is crucial to understanding the data economy.

Our methodology is impervious to Solove's (2021) criticisms since our survey specifically honed in on respondents' concerns regarding data exchanges with Alipay's mini-programs and was complemented with specific administrative data regarding these interactions. Further solidifying our survey's authenticity, we have also examined respondents' likelihood of undertaking privacy-centric actions on Alipay, such as revoking previous data-sharing permissions and altering default privacy configurations. The results substantiate that survey responses indeed mirror genuine user concerns regarding data-sharing on the platform.

Why might Alipay users with privacy concerns seemingly overlook these apprehensions when permitting data sharing? From our data set, we find that respondents declined data-sharing requests from mini-programs 26.5% of the time on average. This notable rate of rejection signifies that these users have not entirely given up protecting their data privacy. Existing literature on privacy identifies several psychological and behavioral explanations for this paradox. Some factors include users' lack of awareness regarding data-sharing repercussions (Kesan, Hayes, and Bashir, 2015), a present bias where users prioritize immediate digital conveniences while underestimating the long-term consequences of data sharing (Acquisti, 2004), and an "illusion of control" that leads users to believe they maintain control during data-sharing decisions (Brandimarte, Acquisti, and Loewenstein, 2013).

Distinct from these findings, our study introduces a unique perspective. We find a direct link between the intensity of privacy concerns and the utilization of digital services. Intriguingly, those with profound privacy concerns tend to engage with their permitted mini-programs more frequently and intensively. This might hint that while these users are privacy-aware, their demand for digital services could outweigh, or even eclipse, their reservations. Therefore, their evident data-sharing decisions might not contradict their altitudes but rather represent a balance struck between their digital needs and privacy concerns.

To determine the causal relationship between users' digital demand and their privacy concerns, we have employed an instrumental variable (IV) method to isolate exogenous shifts in digital demand. Specifically, we utilized the number of Alipay-bundled shared bicycles available in a user's city as an instrument for the number of mini-programs they engaged with. As Ouyang (2022) highlighted, the distribution of these shared bicycles across cities provides a plausibly exogenous variation in the demand for Alipay's digital offerings. An increased number of shared bicycles in a city encourages residents to use Alipay to access these bicycles, which could subsequently introduce them to use other digital services within Alipay. Leveraging this bicycle distribution data, we identified a significant causal relationship between users' engagement with digital services and their privacy concerns.

Considering the nascent stage of our digital data ecosystem, a comprehensive analysis of the interplay between users' digital demand and privacy concerns seems premature. However, broadening our dataset by 17 more months (from August 2020 to December 2021) has unveiled some compelling trends. In this span, "unconcerned" users explored 27.8 mini-programs and consented to data sharing with 22.5, while the "concerned" explored 32.8 and shared data with 24.6, and the "very concerned" explored 33.4 and shared data with 23.8. Across these segments, both initial mini-program visits and data-sharing consent rates saw significant upticks compared to the main sample period (July 2019 to July 2020). Remarkably, the growth rate for the "concerned" and "very concerned" segments outpaced that of the "unconcerned," even after adjusting for individual characteristics. This trend underscores that, despite their heightened privacy concerns, the former groups' rising thirst for digital services likely propelled them to share data more freely.

To mitigate the potential bias stemming from the likelihood of more active Alipay users responding to our survey, we further analyzed a representative sample of 100,000 users. This sample was randomly selected from the entire pool of active Alipay users. Employing a behavior-based measure of privacy concerns—namely, whether a user has modified Alipay's default privacy settings—we corroborated the key findings from our survey sample. Specifically, those with stronger privacy concerns tend to approve more data sharing with mini-programs and interact with the authorized mini-programs more frequently and more intensely. Additionally, we leveraged this representative sample to investigate the evolution of privacy concerns among diverse consumers, especially after a notable incident on January 3, 2018. This event, instigated by Alipay, significantly heightened awareness of data privacy among its users. Intriguingly, this incident rendered the avid users of mini-programs more predisposed to privacy concerns.

Overall, our study not only validates the data privacy paradox within a context highly pertinent to the data economy, but also leverages this paradox to examine a critical dynamic inherent to the data economy: the simultaneous growth of consumers' privacy concerns and digital demands as the data economy evolves. Our results depict a complex scenario. Although privacy concerns are escalating, the advantages derived from progressively sophisticated digital services, powered by consumer data, might counterbalance or even surpass these apprehensions, prompting continued data sharing.

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. This is particularly evident in the case of fintech firms that harness vast amounts of consumer data to provide comprehensive financial services. Berg et al. (2020) undertook predictive studies that showcased the promise of digital footprints in facilitating novel financial offerings. Ouyang (2022) furnishes empirical evidence indicating that as Alipay's user base expanded, it could offer credit to those consumers who were overlooked by traditional banks. This not only improved consumer welfare but also boosted profits for lenders.

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.[2]

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.

[2] 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).

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 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.[3] 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.[4] This data-sharing

[3] 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 bicycles. 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.

[4] 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

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.

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 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

opposed to nonpersonalized services. In our setting, an Alipay user cannot use any service from a mini-program unless she authorizes data sharing.

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.[5] These 10,875 respondents are the main sample for our analysis.

Due to the natural tendency that more-active users are more likely to pay attention to the message box in the Alipay application and thus to open the survey link, this sample of survey respondents is representative of more-active Alipay users rather than the whole population of Alipay users. For robustness and comparison, we will also examine in Section VI a representative sample of 100,000 Alipay users 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.

[5] 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 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%.

We also asked the respondents this specific question: "*What privacy issues are you concerned about when using mini-programs in Alipay?*" This question allowed each respondent to select more than one option from a list of four, including: 1) data leakage and security, 2) price discrimination by merchants, 3) seductive advertising and temptation consumption, and 4) others. The first choice represents potential concerns about insufficient protections provided by mini-programs to secure user data and prevent hacking and other data leakage, as modeled by Fainmesser, Galeotti and Momot (2019). The second choice represents a concern that extensive data sharing by consumers may allow merchants to infer consumers' reservation prices and thus employ price discrimination. There is a large body of economics literature analyzing this concern in the digital economy, as reviewed by Acquisti, Taylor and Wagman (2016), Bergemann and Morris (2019), and Goldfarb and Tucker (2019). The third choice represents a new concern that in the booming digital economy, extensive data sharing by consumers may expose consumers' personal weaknesses, such as a lack of self-control, to online advertisers and sellers, as recently emphasized by Liu, Sockin and Xiong (2020). Interestingly, 86% of the respondents selected data leakage and security, 49% selected seductive advertising and temptation consumption, and 21% selected price discrimination by merchants. 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. [6] 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 offers a comparison among three user categories: "unconcerned," "concerned," and "very concerned," based on their reactions to the survey question, "*Are you concerned about negative impacts caused by information shared with mini-programs in Alipay?*" Notably, while age doesn't present any significant variance across groups, both "concerned" and "very concerned" users tend to have a more extended digital history, exhibit a higher probability of being female, and are more likely to possess 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 present some empirical results, which are consistent with the data privacy paradox. We then validate the survey-based measure of privacy concerns.

A. Privacy Concerns and Data Sharing

To determine whether to share their personal data with a specific mini-program, Alipay users weigh the benefits they gain from the mini-program against the potential privacy risks. Both these benefits and

[6] 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.

costs might vary depending on the user and the particular mini-program in question. For clarity, we propose a linear decomposition of the cost for user i sharing data with mini-program j, denoted as c_{ij}:

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

where c_i encapsulates the user's inherent privacy concerns, c_j reflects the mini-program's potential risk, especially if it requires more sensitive data or has a questionable privacy protection reputation, and ϵ_{ij} is a stochastic component independent of both the user and mini-program. Similarly, we linearly decompose the user's benefit from the mini-program, b_{ij}, as

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

where b_i pertains to the user's inclination or receptiveness to digital services, b_j represents the value or usefulness of the mini-program's services, and ε_{ij} is another noise component, also independent of the user-mini-program pairing.

The user authorizes 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 adjusting for the mini-program-specific factors, the decision to share hinges predominantly on the user's intrinsic factors, which can be captured by $b_i - c_i$. For a start, let's consider a scenario where b_i and c_i are unrelated—meaning a user's valuation of digital services is not influenced by their privacy concerns. This perspective often emerges in policy debates around data privacy, typically treating privacy apprehensions without acknowledging consumers' demand for digital services. Consequently, we can hypothesize:

Hypothesis 1: Given other factors being constant, users with heightened privacy concerns will be more hesitant to approve data sharing with mini-programs.

This hypothesis aligns with the conventional understanding encapsulated in the data privacy paradox discourse. Our initial empirical endeavors will be aimed at testing this hypothesis. As an alternative perspective, there could exist a positive correlation between b_i and c_i across users. In this scenario, a user's demand for digital services could counterbalance their privacy concerns, rendering their data-sharing decisions largely impervious to those concerns. We will also examine this possibility in our later analysis.

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. This pattern contradicts 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:

$$Y_i = a_1\, Concerned_i + a_2\, Very\ Concerned_i + a_3\, Age_i$$

$$+a_4\, Digital\ Experience_i + \delta_i + \epsilon_i\ , \qquad (1)$$

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.

Panel A in Table 3 reports the regression results, using the sample from July 2019 to July 2020. 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.

A user’s data-sharing authorization with a mini-program may also depend on the services offered and the data requested by the mini-program. To control for these 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.

Panel B of Table 3 reports the analysis at the user-mini–program level. Even after controlling for mini-program fixed effects, the results are very similar to that from the user-level analysis. 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.

We have also explored how data sharing authorizations may vary across users with different characteristics. In Panel C of Table 3, 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 period from July 2019 to July

2020.[7] Interestingly, the first column shows that higher-educated users who are “concerned” or “very concerned” tend to authorize data sharing more than their “unconcerned” counterparts. This implies that the data privacy paradox is notably intensified among the well-educated. Conversely, the second column shows no significant variance between “concerned” and “very concerned” users compared to “unconcerned” ones in terms of the self-control metric. Thus, the data privacy paradox isn’t exclusive to users with limited educational backgrounds or weaker self-control.

Overall, the results from Table 3 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.

B. Validating Survey-Based Privacy Concerns

A potential criticism surrounding the data privacy paradox is that it might merely be a manifestation of survey respondents not genuinely or consistently revealing their actual privacy preferences. Such skepticism regarding survey results is not new, as noted by scholars like Bertrand and Mullainathan (2001). Solove (2021) similarly challenged the credibility of self-reported privacy concerns in paradox studies, questioning their alignment with observed behaviors.

To address this, we leveraged our expansive administrative dataset. Our goal was to determine if there's a positive correlation between survey-based measures of privacy concerns and tangible actions users take to safeguard their data privacy. Specifically, we looked at two actions: the cancellation of previously approved data sharing with mini-programs and modifications to Alipay's default privacy settings. Conceptually, one would anticipate that users expressing greater privacy concerns would be more inclined to undertake these actions.

Our analysis is again bifurcated into user-level and user-mini-program-level regressions. For the former, we used the regression model from Equation (1), substituting the dependent variable with an indicator: if a user ever rescinded any data-sharing authorization from January 2013 to July 2020 or if they modified Alipay's default privacy settings between May 2017 and April 2020.[8]

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

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

It's worth noting that executing either action not only necessitates privacy concerns but also awareness about how to revoke data-sharing permissions or adjust Alipay's privacy settings. As Table 1 illustrates, a mere 60% of our survey participants knew how to tweak Alipay's default settings. Our regression factors in extensive controls, such as user's digital experience, age, gender, and city fixed effects, to account for their digital literacy.

Table 4's Panel A shows the results from these user-level regressions. Columns (1) and (2) present the dependent variable as the '*Has Canceled*' indicator. Controlling for other variables, those who expressed "concerned" or "very concerned" sentiments about data sharing with mini-programs were notably more prone to having withdrawn data sharing permissions with at least one mini-program compared to their "unconcerned" counterparts. Interestingly, the propensity to cancel was even more pronounced in the "very concerned" cohort than the "concerned" group.

Columns (3) and (4) pivot to the '*Privacy Setting Changed*' indicator. Without factoring in controls, those with higher privacy concerns were more likely to have adjusted Alipay's default settings compared to the "unconcerned" group. However, upon including extensive controls in column (4), only the "very concerned" group maintained a significantly higher likelihood to adjust settings, whereas the "concerned" group's probability diminished.

Furthermore, there's a clear link between taking these protective measures and both digital experience and age, suggesting younger and more digitally experienced users are better equipped to make privacy-related decisions.

Panel B of Table 4 extends this analysis to the user-mini-program level, focusing on data-sharing cancellation. This more granular approach lets us account for mini-program specific effects, offering insights into varying propensities to cancel data-sharing agreements with identical mini-programs but differing user privacy anxieties. We employed the regression model from Equation (2), analyzing all existing data-sharing consents between any user and mini-program pairing from July 2019 to July 2020, a sample size of 481,143. We observed that "very concerned" users displayed a substantially higher inclination to cancel data-sharing consents.

In summary, Table 4 corroborates the idea that survey-derived measures of privacy concerns are intrinsically tied to tangible actions taken by Alipay users to bolster their data privacy. Our findings indicate that Solove's (2021) skepticism doesn't pertain to our analysis.

IV. Digital Demands

How can we account for the absence of a negative correlation between privacy concerns and the frequency of data-sharing authorizations? As deliberated in Section III.A, this apparent paradox might make sense if there's a positive correlation between a user's apprehensions about sharing personal data with a mini-program and the perceived advantages of using it. In this section, we delve deeper into the interplay between privacy concerns and digital needs.

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. We focus on the following hypothesis:

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

We examine this hypothesis by using the following regression specification:

$$\mathrm{Y_{ijt}} = \mathrm{a}_1\, Concerned_i + \mathrm{a}_2\, Very\ Concerned_i + \mathrm{a}_3\, Age_{it} + \mathrm{a}_4\, Digital\ Experience_{it} + \delta_\mathrm{i} + \mu_\mathrm{j} + \theta_t + \varepsilon_{\mathrm{ijt}}, \qquad (3)$$

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

Table 5 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 supporting Hypothesis 2.

This surprising finding, where privacy-concerned individuals also exhibit higher digital demands, suggests a nuanced understanding. The higher data-sharing authorizations by these individuals don't necessarily contradict their expressed privacy concerns, a phenomenon often linked with the data privacy paradox. Instead, it highlights a potential balance between their privacy concerns and digital needs. This balance could render their data-sharing decisions unaffected by, or even positively aligned with, their privacy concerns.[9]

B. Causal Effect of Digital Demand on Privacy Concerns

Our initial correlation analysis showed an association between greater digital demand and higher privacy concerns. However, this correlation can neither establish causality nor rule out reverse causality. For example, underlying factors might influence both variables and cause the observed correlation.

To circumvent this challenge, Table 6 uses an instrumental variable (IV) approach to isolate exogenous variation in digital demand. Finding a valid instrument for digital demand is intricate. We use the number of Alipay-bundled shared bicycles placed in a user's city as an instrument for the number of mini-programs engaged by the user. As discussed in Ouyang (2022), the placement of shared bicycles across cities provides plausibly exogenous variation in users’ demand for Alipay's digital services. An increased bicycle count in a city naturally boosts residents' reliance on Alipay for bicycle access, spurring them to utilize Alipay's other digital offerings.[10]

[9] 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.

[10] In the three years leading up to our July 2020 survey, bike-sharing services experienced a notable surge in China. As detailed by Ouyang (2022), bike-sharing enterprises vied for dominance, strategically deploying shared bicycles across various Chinese cities in a staggered manner. This was done to entice residents from these cities to adopt their

Ouyang (2022) validates that bicycle availability is a potent predictor for Alipay's digital service use, satisfying the relevance condition for a valid instrument. This is confirmed in Panel B of Table 6. The F-statistic on the instrument ranges from 16.9 to 52.7 across specifications. This indicates a strong first-stage relationship between bicycle availability and mini-program visits.

The exclusion restriction is also satisfied—bicycle placement affects privacy concerns only through increased digital service demand, rather than through other channels. It is unlikely bicycle placement has a broad and direct effect on privacy concerns outside of intensified platform engagement. While some may speculate that bicycle distributions could mirror unseen city traits that alter privacy perspectives, Ouyang (2022) dispels this by finding no link between bicycle distribution and local economic conditions.

Employing number of Alipay-bundled shared bicycles placed in each city as the digital demand instrument, the Two-Stage Least Squares (2SLS) regression results in Panel A reveal a user's engagement with mini-programs notably boosts their privacy concerns. This causal effect persists across varied controls and fixed effects. Economically, it's significant: each added mini-program engagement in the year leading up to the survey augments users' data privacy worries by 1.3% to 2.1%, as indicated by the estimated coefficients.

Conclusively, the IV analysis fortifies our core findings, suggesting that surging digital demands genuinely lead to amplified privacy concerns. This bolsters our conviction that escalating digital service engagement exacerbates these concerns.

C. Digital Demand and Cancellation

To solidify the connection between increasing digital demands and heightened privacy concerns, we delve into the relationship between users' digital demands and their tendencies to revoke existing data-sharing authorizations, which serves as a tangible measure of their privacy concerns. We aim to evaluate the following hypothesis.

services. Though each of these companies had proprietary apps, several of the major contenders collaborated with Alipay. This partnership allowed Alipay users to access and unlock the shared bicycles using mini-programs, eliminating the need to download another application. Such an integration led to a cross-pollination of services within Alipay. As the number of Alipay-integrated shared bicycles grew in a city, its residents found it more convenient to access and utilize these bikes. Following their initial use, users gained familiarity with Alipay and became more inclined to explore its array of services. Ouyang (2022) emphasizes that these spill-over effects were both substantial and enduring.

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, further confirming the key notion that users with greater digital demands tend to be more concerned about data privacy.

V. Data Sharing Evolution

In this section, we probe a pivotal question: How do consumers' privacy concerns and data-sharing habits evolve? Our earlier findings align with economic literature suggesting that privacy isn't a fixed preference. Instead, it's intertwined with the economic implications of revealing personal data to others (Stigler, 1980; Posner, 1981). The cost of privacy, particularly if personal data is compromised (Fainmesser et al., 2019), rises with the volume of shared data. Extensive data sharing also empowers digital providers to fine-tune price discrimination (Taylor, 2004; Acquisti & Varian, 2005) and zero in on users' vulnerabilities (Liu et al., 2020). In both scenarios, the more data consumers provide, the more their privacy anxieties amplify.

Despite the nascent state of the data economy, Figure 2 underscores the uptick in privacy worries. It depicts privacy concerns across respondents, segmented by their digital experience ranging from one to 12 years. We gauge privacy anxieties for each group based on the fraction that voice “concerned” or “very concerned” sentiments about data-sharing. Notably, as digital experience grows, so do privacy concerns.

However, does this escalation in privacy concerns curb data sharing? It's essential to realize that more consumer data allows service providers to refine user experiences through tailored services, capitalizing on the exponential benefits of data sharing (Jones & Tonetti, 2020; Farboodi & Veldkamp, 2020; Cong et al., 2020). Hence, even as privacy-related apprehensions grow, consumers might persist in data sharing due to its escalating benefits.

Shifting focus to Alipay users, Figure 3 depicts the monthly average of data-sharing authorizations across three user groups with varied privacy concerns. Though monthly patterns oscillate, an overall upward trend is evident across all groups. Interestingly, post-July 2020, the 'concerned' and 'very concerned' clusters exhibited heightened data-sharing behaviors.

Furthermore, Panel B of Figure 1 summarizes data-sharing and initial visits in the post-survey period from August 2020 to December 2021. "Unconcerned" users engaged with 27.8 mini-programs and shared data with 22.5, "concerned" users with 32.8 and 24.6 respectively, and "very concerned" users with 33.4 and 23.8. A pronounced trend emerges: all groups were more active and shared more data post-survey than pre-survey, with 'concerned' and 'very concerned' users outpacing their 'unconcerned' counterparts in data sharing.

Table 8 formally shows that the data privacy paradox is not confined to the pre-survey period but in fact intensifies subsequently. This table presents regression analysis of the data privacy paradox results comparing the post-survey period with the pre-survey period. All regressions are at the user-mini-program level. They show that the "concerned" and "very concerned" users are significantly more likely to authorize mini-programs than the "unconcerned" users in the post-survey period, even after controlling for user characteristics and mini-program fixed effects.

Taken together, we find an encouraging pattern: despite growing privacy concerns over time, Alipay users display a propensity to authorize more data sharing. This finding corroborates a pivotal premise: as data economy continues to prosper, it may harness the potential to provide more powerful digital services, effectively attracting users to perpetuate data sharing despite their growing privacy concerns.

VI. A Representative Sample

In our study, we initially faced the challenge regarding the sample bias favoring active users, which raised concerns about the generalizability of our findings. To overcome this challenge, we constructed a representative sample of 100,000 users and found results consistent with our original survey sample, thereby affirming the robustness and applicability of our main conclusions. Additionally, we delved into the differential impact of a privacy awareness event on heavy versus light users of digital services. This analysis provides a new result to show that privacy concerns intensify with greater digital service usage.

A. 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 A2, the random sample is indeed less active in using mini-programs than the survey sample.[11] 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.[12]

In Table 9, 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

[11] 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.

[12] 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.

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 5. 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 4. We again observe that the cancellation rate is significantly and positively correlated with user activeness. Panel D shows the results of testing whether users who visit more mini-programs are more likely to change their privacy settings, indicating a higher level of privacy concern. It uses specifications similar to Table 6 and establishes the causal relationship between digital demand and revealed privacy concern. Taken together, we confirm that the key results of our analysis are robust in the representative sample of Alipay users.

B. Heterogeneous Responses in a Privacy Related Incident

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:

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

where t corresponds to the number of days after the incident on January 3, 2018, $\mathit{Daily\ Cancellation\ Dummy}_{i,t}$ is a dummy variable indicating whether user i has canceled at least one mini-program during the day t, $\mathit{Heavy\ User}_i$ 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, $\mathit{Light\ User}_i$ is a dummy that equals $1 - \mathit{Heavy\ User}_i$, 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 A2 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 5, 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.

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*.

Berg, T., Burg, V., Gombovic´, A., and Puri, M. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7):2845–2897.

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. (2022). 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.

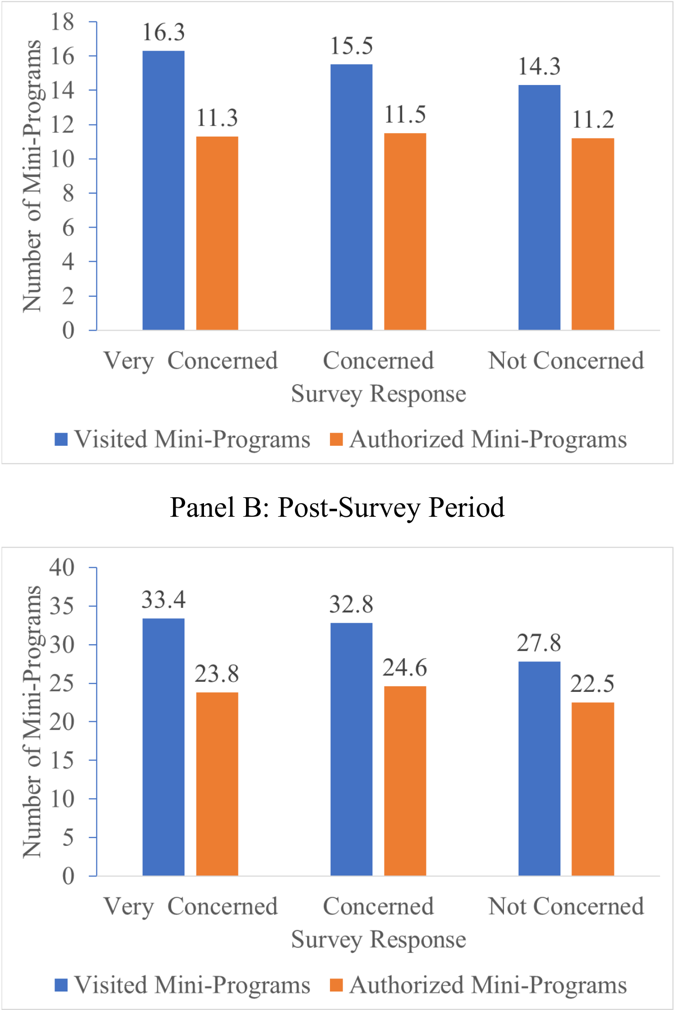
Stigler, G. J. (1980). An Introduction to Privacy in Economics and Politics. *Journal of Legal Studies*, 9(4), 623-644.

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: 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.

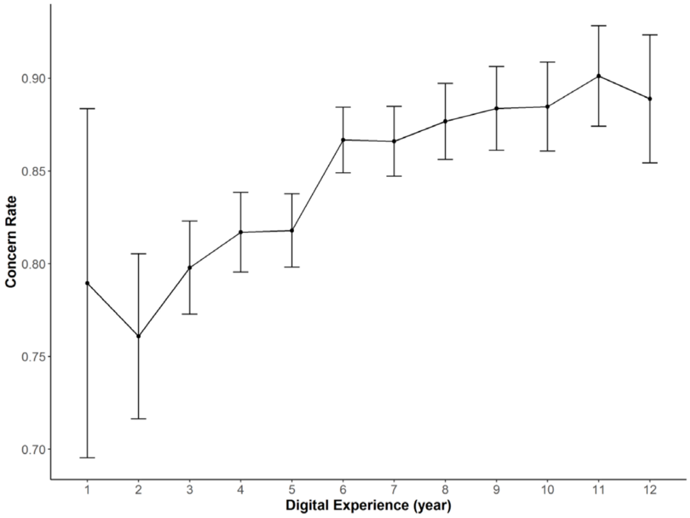


Figure 3: Time Trend in Data-Sharing Authorizations

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.

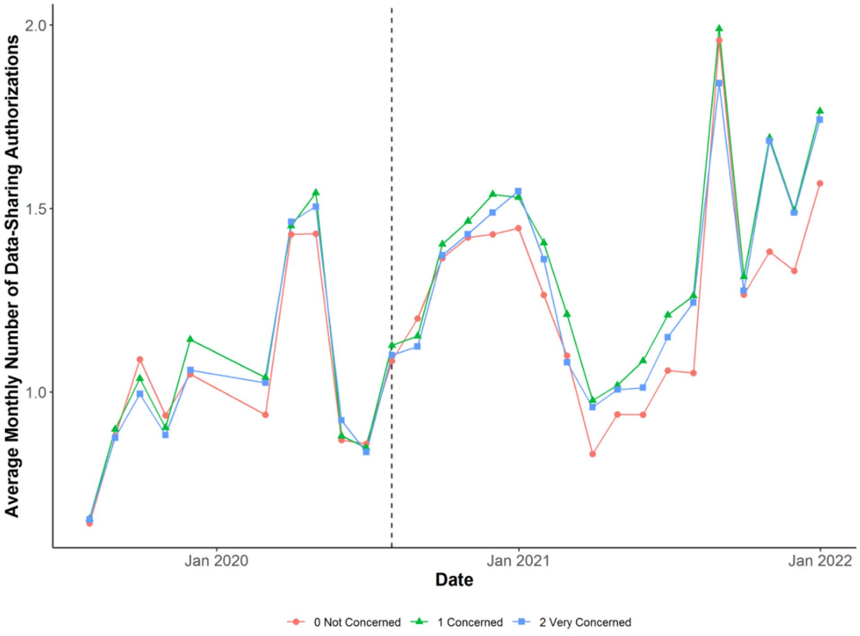
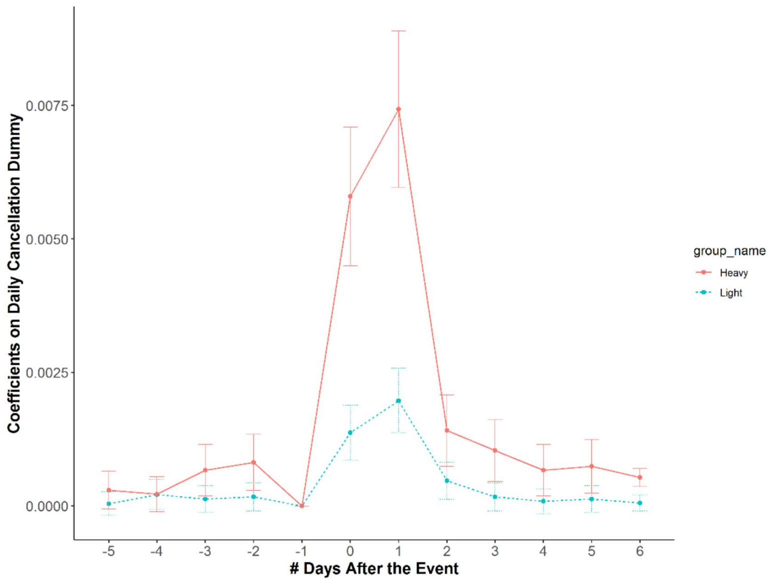


Figure 4: Activeness and Response to the 2017 Footprint Report Incident

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

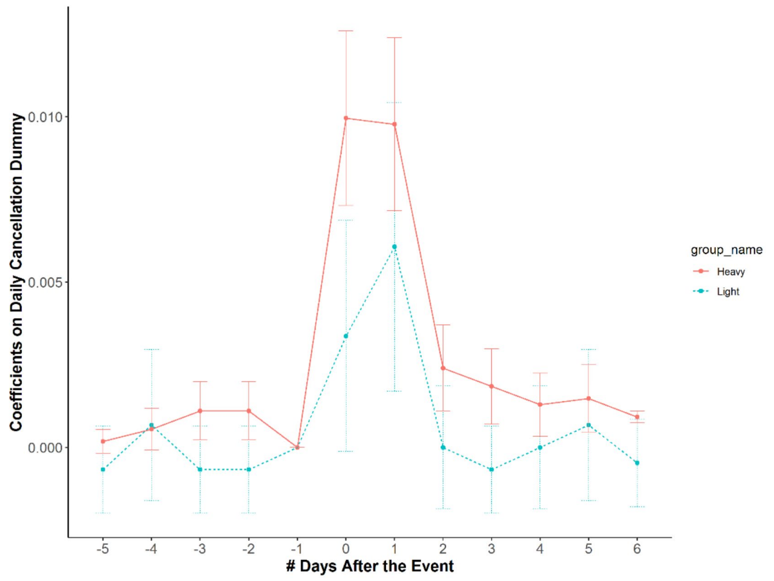


Table 1: Responses to Selected Survey Questions

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$_i$	10,875	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Very Concerned Dummy$_i$	10,875	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Privacy Setting Changed$_i$	10,875	0.49	0.5	0.00	0.00	0.00	1.00	1.00
Digital Experience$_i$ (month)	10,871	74.97	35.07	4.00	48.00	70.00	97.00	190.00
Age$_i$ (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$_i$	10,875	34.22	22.78	0.00	19.00	30.00	43.00	422.00
# Entered Mini-Programs$_i$	10,875	46.57	55.45	1.00	26.00	38.00	53.00	1609.00
Has Canceled$_i$	10,875	0.48	0.50	0.00	0.00	0.00	1.00	1.00
# Cancellations$_i$	10,857	2.66	5.54	0.00	0.00	0.00	3.00	80.00
Cancellation Rate$_i$	10,857	0.05	0.10	0.00	0.00	0.00	0.06	1.00
Monthly mini-program use								
# Active Days$_{it}$	1,521,645	0.57	2.92	0.00	0.00	0.00	0.00	31.00
# Uses$_{it}$	1,521,645	0.81	5.01	0.00	0.00	0.00	0.00	75.00
# Launches$_{it}$	1,521,645	2.29	15.07	0.00	0.00	0.00	0.00	230.00
# Visited Pages$_{it}$	1,521,645	5.20	33.67	0.00	0.00	0.00	0.00	503.00

Panel B: Privacy Concern and Personal Characteristics

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

Table 3: The Data Privacy Paradox

This table presents regression analysis of the data privacy paradox results for the pre-survey period from July 2019 through July 2020. *Concerned Dummy* and *Very Concerned Dummy* in 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." *Education* is a dummy indicating whether the user has a college degree or higher. *Self Control* is a dummy indicating whether the user's opt-in rate of seemingly addictive mini-programs is higher than the opt-in rate of other mini-programs in the pre-survey period. Panel A displays results for user-level regressions. Columns (1)–(2) present results for the number of authorized mini-programs and columns (3)–(4) for the number of initially visited mini-programs. Panel B provides results for regressions at the user-mini–program level. In each pair of user-mini–program, columns (1)–(2) indicate results for the dummy whether the user allowed the authorization and columns (3)–(4) for the dummy whether the user visited at least once. Panel C reports heterogeneity analysis of the data privacy paradox, where we interact privacy concern measures with a user's personal characteristics. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: User-Level Analysis

	# Authorized Mini-Programs$_i$		# Visited Mini-Programs$_i$	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$	0.334	0.207	1.262***	1.243***
	(0.213)	(0.214)	(0.322)	(0.320)
Very Concerned Dummy$_i$	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)	
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
Observations	10,875	10,858	10,875	10,858
Adjusted R2	0.0001	0.021	0.003	0.045

Panel B: Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$ (0/1)		Visited Dummy$_{ij}$ (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$ (× E-4)	0.862	0.386	2.897***	2.552***
	(0.745)	(0.735)	(0.848)	(0.836)
Very Concerned Dummy$_i$ (× 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)	
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
Observations	25,414,875	25,364,288	25,414,875	25,364,288
Adjusted R2	0.000	0.105	0.000	0.129

Panel C: Heterogeneity Analysis at User-Mini–Program Level

	Authorized Dummy$_{ij}$ (0/1)		Visited Dummy$_{ij}$ (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$ (× E-4)	-0.649	0.600	1.592*	2.826***
	(0.811)	(0.771)	(0.931)	(0.875)
Very Concerned Dummy$_i$ (× E-4)	-1.096	-0.327	2.696***	3.679***
	(0.807)	(0.765)	(0.939)	(0.881)
Concerned Dummy$_i$ × Characteristics Measure$_i$ (× E-4)	5.120***	-2.644	4.354**	-3.283
	(1.833)	(1.855)	(2.001)	(2.096)
Very Concerned Dummy$_i$ × Characteristics Measure$_i$ (× E-4)	3.329*	-2.501	2.982	-3.619*
	(1.800)	(1.827)	(1.981)	(2.083)
Characteristics Measure$_i$	-0.000	0.001***	0.000	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Characteristics Measure	Education	Self-Control	Education	Self-Control
Mini-program FE	Y	Y	Y	Y
City FE	Y	Y	Y	Y
Gender FE	Y	Y	Y	Y
Control Age	Y	Y	Y	Y
Control Digital Experience	Y	Y	Y	Y
Observations	25,364,288	25,364,288	25,364,288	25,364,288
Adjusted *R*2	0.105	0.105	0.129	0.129

Table 4: 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$_i$ (0/1)		Privacy Setting Changed$_i$ (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$	0.060***	0.033***	0.028*	0.012
	(0.014)	(0.014)	(0.015)	(0.015)
Very Concerned Dummy$_i$	0.082***	0.051***	0.060***	0.041***
	(0.014)	(0.014)	(0.014)	(0.015)
Digital Experience$_i$		0.004***		0.001***
		(0.0001)		(0.0001)
Age$_i$		-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 the User-Mini–Program Level

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

Table 5: 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$_{it}$		# App Uses$_{it}$		# App Launches$_{it}$		# Visited Pages$_{it}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Concerned Dummy$_i$	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$_i$	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$_i$		-0.0001		-0.0003		-0.001		-0.001
		(0.000)		(0.001)		(0.001)		(0.003)
Age$_i$		0.020***		0.033***		0.080***		0.128***
		(0.001)		(0.002)		(0.005)		(0.011)
Constant	0.468***		0.651***		1.864***		4.339***	
	(0.023)		(0.039)		(0.112)		(0.226)	
Mini-program FE	N	Y	N	Y	N	Y	N	Y
Year-Month FE	N	Y	N	Y	N	Y	N	Y
City FE	N	Y	N	Y	N	Y	N	Y
Gender FE	N	Y	N	Y	N	Y	N	Y
Observations	1,521,645	1,519,020	1,521,645	1,519,020	1,521,645	1,519,020	1,521,645	1,519,020
Adjusted *R*2	0.0002	0.119	0.0002	0.096	0.0001	0.086	0.0001	0.078

Table 6: Causal Effect of Digital Demand on Privacy Concerns

This table examines the causal relationship between digital demand and revealed privacy concern in the survey. *Concerned Dummy$_i$* is a dummy variable that equals 1 if the user *i*'s 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." *# Visited Mini-Programs$_i$* indicates the number of initially visited mini-programs by the user *i* in the pre-survey period from July 2019 through July 2020. *log(Average Monthly # Active Bikes)$_c$* is the log transformed average of the monthly number of Alipay-bundled shared bicycles placed in the user *i*'s city *c* before July 2020. Panel A reports 2SLS estimates, instrumenting for number of visited mini-programs with average bicycle placement; Panel B reports the first stage. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Concerned Dummy$_i$		
	(1)	(2)	(3)
Panel A. Two-Stage Least Squares			
# Visited Mini-Programs$_i$	0.021***	0.013***	0.013*
	(0.005)	(0.005)	(0.008)
Panel B. First Stage for # Visited Mini-Programs$_i$			
log(Average Monthly # Active Bikes)$_c$	0.373***	0.427***	0.311***
	(0.051)	(0.061)	(0.076)
F-Statistic	52.7	48.5	16.9
Adjusted *R2*	0.005	0.042	0.070
Birth City FE	N	N	Y
Gender, Education, Occupation FE	N	Y	Y
Control Age and Digital Experience	N	Y	Y
Observations	9,849	6,140	6,140

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

Panel B: Regression at User-Mini–Program Level

	$Canceled\ Dummy_{ij}$			
	(1)	(2)	(3)	(4)
Active-Month Ratio$_{ij}$	0.047***		0.081***	
	(0.007)		(0.011)	
log(1+ # Avg. Monthly Active Sessions)$_{ij}$		0.003**		0.007***
		(0.001)		(0.003)
Digital Experience$_i$ (× E-4)	1.557***	1.464***	-2.358***	-2.534***
	(0.218)	(0.217)	(0.410)	(0.409)
Age$_i$ (× 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

Table 8: The Data Privacy Paradox in the Post-Survey Period

This table presents regression analysis of the data privacy paradox results comparing the post-survey period from August 2020 through December 2021 with the pre-survey period from July 2019 to July 2020. *Concerned Dummy* and *Very Concerned Dummy* in 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." The results for regressions are at the user-mini–program level. In each pair of user-mini–program, columns (1)–(2) indicate results for the dummy whether the user allowed the authorization and columns (3)–(4) for the dummy whether the user visited at least once. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

	Authorized Dummy$_{ij}$ (0/1)		Visited Dummy$_{ij}$ (0/1)	
	(1)	(2)	(3)	(4)
Concerned Dummy$_i$ (× E-4) × Post-Survey Dummy$_{ij}$	274.587***	256.997***	3.314***	-9.945
	(53.601)	(52.630)	(0.530)	(9.511)
Very Concerned Dummy$_i$ (× E-4) × Post-Survey Dummy$_{ij}$	535.489***	502.886***	2.696***	-23.389**
	(53.176)	(52.183)	(0.524)	(9.492)
Concerned Dummy$_i$ (× E-4)	-272.477***	-255.647***	-0.000	12.561
	(53.697)	(52.690)	(0.000)	(9.630)
Very Concerned Dummy$_i$ (× E-4)	-534.387***	-502.433***	-0.000	25.502***
	(53.268)	(52.240)	(0.000)	(9.620)
Post-Survey Dummy$_{ij}$	-0.863***	-0.799***	-0.997***	-0.923***
	(0.004)	(0.004)	(0.000)	(0.001)
Constant	0.866***		1.000***	
	(0.004)		(0.000)	
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
Observations	72,525,375	72,401,144	72,525,375	72,401,144
Adjusted R2	0.298	0.367	0.368	0.440

Table 9: Results from the Representative Sample

This table reports four sets of robustness tests from using the representative random sample of 100,000 Alipay users. Panel A presents the robustness test for the digital privacy paradox, where the regressions are at the user level. *Privacy Setting Changed* is a behavior-based measure for privacy concerns, defined as a dummy variable that equals 1 if a user changed the default privacy settings at least once between May 2017 and April 2020, and 0 otherwise. Columns (1) and (2) show results for the number of authorized mini-programs, and columns (3) and (4) show results for the number of initially visited mini-programs. In columns (2) and (4), we control for digital experience and age, along with gender and city fixed effects. Panel B tests the positive relationship between privacy concerns and demand for digital services, where the regressions are at the user-mini–program-month level, and the standard errors are clustered at the user level. We use four variables from July 2019 to July 2020 to capture demand for digital services, namely, number of active days, number of uses, number of launches, and number of visited pages. Columns (1), (3), (5), and (7) show regression results without any controls, while columns (2), (4), (6), and (8) control for digital experience and age, as well as user gender, user city, mini-program, and year-month fixed effects. Panel C examines the positive relationship between user activeness and cancellation of mini-programs, where the regressions are at the user-mini-program level, and the standard errors are clustered at the user level. The sample covers user-mini–program pairs that had been active between July 2019 and July 2020. We use two measures of user activeness. The first one is an active-month ratio that refers to the total number of months the user is active as a percentage of the total number of months from the beginning to the end of authorizations in all mini-programs. The second one is the logarithm of the average monthly active uses. We use the whole sample in columns (1) and (2) and a subsample of users who canceled at least one mini-program before July 2019 in columns (3) and (4). In all the regressions, we control for digital experience and age, as well as gender and city fixed effects. Panel D examines the causal relationship between digital demand and revealed privacy concern in terms of privacy setting changing behavior. *# Visited Mini-Programs$_i$* indicates the number of initially visited mini-programs by the user *i* in the pre-survey period from July 2019 through July 2020. *log(Average Monthly # Active Bikes)$_c$* is the log transformed average of the monthly number of Alipay-bundled shared bicycles placed in the user *i*'s city *c* before July 2020. Part 1 reports 2SLS estimates, instrumenting for number of visited mini-programs with average bicycle placement; Part 2 reports the first stage. We denote ***, **, and * as the 1%, 5%, and 10% confidence levels, respectively. We report standard errors in parentheses.

Panel A: Analysis at User Level of the Data Privacy Paradox

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

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

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

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

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

Panel D: Analysis at User Level of the Effect of Digital Demand Shock on Privacy Concern

	Privacy Setting Changed$_{i}$		
	(1)	(2)	(3)
Part 1. Two-Stage Least Squares			
# Visited Mini-Programs$_{i}$	0.017***	0.014***	0.019***
	(0.002)	(0.003)	(0.006)
Part 2. First Stage for # Visited Mini-Programs$_{i}$			
log(Average Monthly # Active Bikes)$_{c}$	0.190***	0.181***	0.114***
	(0.006)	(0.009)	(0.011)
F-Statistic	957.1	422.5	109.0
Adjusted *R2*	0.010	0.031	0.051
Birth City FE	N	N	Y
Gender, Education, Occupation FE	N	Y	Y
Control Age and Digital Experience	N	Y	Y
Observations	90,645	45,666	45,666