Evaluating the Privacy Valuation of Personal Data on Smartphones

LIHUA FAN*, Department of Computer Science and Technology, Tsinghua University, China
SHUNING ZHANG*, Institute for Network Sciences and Cyberspace, Tsinghua University, China
YAN KONG, Institute for Network Sciences and Cyberspace, Tsinghua University, China
XIN YI†, Institute for Network Sciences and Cyberspace, Tsinghua University and Zhongguancun Laboratory, China
YANG WANG, University of Illinois at Urbana-Champaign, United States
XUHAI "ORSON" XU, Information School, University of University of Washington, United States
CHUN YU, Department of Computer Science and Technology, Tsinghua University, China
HEWU LI, Tsinghua University and Zhongguancun Laboratory, China
YUANCHUN SHI, Department of Computer science and Technology, Tsinghua University, China

Smartphones hold a great variety of personal data during usage, which at the same time poses privacy risks. In this paper, we used the selling price to reflect users' privacy valuation of their personal data on smartphones. In a 7-day auction, they sold their data as commodities and earn money. We first designed a total of 49 commodities with 8 attributes, covering 14 common types of personal data on smartphones. Then, through a large-scale reverse second price auction (N=181), we examined students' valuation of 15 representative commodities. The average bid-price was 62.8 CNY (8.68 USD) and a regression model with 14 independent variables found the most influential factors for bid-price to be privacy risk, ethnic and gender. When validating our results on non-students (N=34), we found that despite they gave significantly higher prices (M=109.8 CNY, 15.17 USD), "privacy risk" was still one of the most influential factors among the 17 independent variables in the regression model. We recommended that stakeholders should provide 8 attributes of data when selling or managing it.

CCS Concepts: • **Security and privacy** → **Economics of security and privacy**; • **Human-centered computing** → **User studies**; **Smartphones**.

Additional Key Words and Phrases: smartphone personal data, selling, auction, price model, privacy risk, context, monetary benefits

ACM Reference Format:
Lihua Fan, Shuning Zhang, Yan Kong, Xin Yi, Yang Wang, Xuhai "Orson" Xu, Chun Yu, Hewu Li, and Yuanchun Shi. 2024. Evaluating the Privacy Valuation of Personal Data on Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 100 (September 2024), 33 pages. https://doi.org/10.1145/3678509

*These authors contributed equally to this work.
†Corresponding author.

Authors' addresses: Lihua Fan, air8218@163.com, Department of Computer Science and Technology, Tsinghua University, China; Shuning Zhang, zsn23@mails.tsinghua.edu.cn, Institute for Network Sciences and Cyberspace, Tsinghua University, China; Yan Kong, ky21@mails.tsinghua.edu.cn, Institute for Network Sciences and Cyberspace, Tsinghua University, China; Xin Yi, yixin@tsinghua.edu.cn, Institute for Network Sciences and Cyberspace, Tsinghua University and Zhongguancun Laboratory, China; Yang Wang, University of Illinois at Urbana-Champaign, United States; Xuhai "Orson" Xu, Information School, University of University of Washington, United States; Chun Yu, Department of Computer Science and Technology, Tsinghua University, China; Hewu Li, Tsinghua University and Zhongguancun Laboratory, China; Yuanchun Shi, Department of Computer science and Technology, Tsinghua University, China.

2474-9567/2024/9-ART100
https://doi.org/10.1145/3678509

1 INTRODUCTION

Smartphones, now ubiquitous, counted approximately 4.9 billion users globally by 2024, covering 60% of the population [7]. With the proliferation of mobile applications [67], these apps gather extensive personal data through sensor-based inputs (like GPS, cameras, and microphones) and non-sensor data (such as call logs and contact lists), utilized by researchers and businesses for profiling identities, psychological states, behaviors, habits, and environmental surroundings (e.g., compute users' identity [14], psychology [14, 63], behavior [63], habits [29] and surroundings [14]).

However, this data collection often breaches privacy. For example, app registration typically "forces" users to accept privacy policies, effectively coercing them into data sharing [62]. Moreover, some apps clandestinely access and sell data without user consent, via backdoor services, to third parties[1]. The consumer data market has evolved into a multi-billion dollar industry [50], with transactions mainly between data collectors and third parties, leaving the actual data producers—the users—unaware.

To enhance user engagement and awareness, researchers suggest allowing users to explicitly sell their personal data, creating a personal data market where users can trade their information for compensation [3, 47]. This concept views personal information as an economic commodity with quantifiable monetary value [47]. Users must weigh the trade-off between privacy and financial rewards when deciding to disclose their personal data.

Determining the correct "price" for personal information is crucial for realizing the proposed idea of a personal data market. Traditional social science methods, such as questionnaires, have been used to assess privacy and data values [54, 70]. Recently, finance researchers have explored additional financial mechanisms, like reverse second price auctions, to analyze privacy premiums and data sales. However, these approaches face two limitations: 1. previous studies primarily valued only specific types of data such as location, communication, media, and app usage [17, 18, 71], overlooking the broader spectrum available from the numerous sensors in modern smartphones. 2. most research focused on individual privacy valuations [27, 76] rather than treating smartphone data as commodities [71], and no study has yet conducted a fully realistic smartphone data transaction experiment. The social context mentioned in the location sharing [16] is important but limited. An increasing number of information types were adopted [20, 56] with different sharing norm and usage context [72]. The variety of data types required different investigation for their distinct functions [20, 56]. This gap [73], coupled with the "privacy paradox" [31], casts doubts on the reliability of valuation results obtained from participants.

In our study, we analyzed how users value their personal data on smartphones by enabling them to sell this data as commodities, employing a reverse second price auction to set the prices. To address previous research limitations, we initially designed 49 distinct commodities encompassing 14 common smartphone data types and designed 8 attributes for each commodity. We selected 15 representative commodities for a comprehensive data buying user study, involving 181 students, to assess their data valuation. The study uncovered key factors influencing participants' willingness to sell their data and the factors affecting their bid prices. Additionally, we developed regression models capable of predicting a commodity's price based on its attributes and the user's demographic characteristics.

To validate the generaliz-ability of our findings, we replicated the study with 34 non-student participants. Both student and non-student groups showed willingness to sell their data for compensation. Notably, the four commodities with the lowest prices were consistent across both groups. While non-students typically offered higher prices, "privacy risk" emerged as a significant factor influencing pricing decisions for both demographics.

The contributions of this paper were three-fold:

- We designed 14 types of smartphone data (opposed to at most 4 types in the past work) as commodities in social context based on *Data Context* [27, 30, 65] and *Privacy Risk* theory [23, 34]. We further proposed 8 attributes for each data to further facilitate users' understanding.

[1]https://www.purevpn.com/blog/third-party-applications-data-leak/, accessed by 5th April, 2024.

- Through a reverse second price auction realistic buying study ($N = 181$), we for the first time unveiled "privacy risk" as the most influential factors out of 14 independent variables in the regression model. We also proved the feasibility of valuating data through selling as all participants priced at least 1 commodity.
- Through bidding, we uncovered more influential factors regarding demographics (e.g., ethnics, students and non-students) compared with the past work [50, 58, 71]. With these findings, we recommended that stakeholders should provide attributes of data when selling or managing it.

We envision our work to be the first large-scale valuation study about smartphone personal data in an authentic and realistic setting. Our results are helpful for end-users to understand and benefit from their privacy more easily, and could also help data buyers to make use of the users' personal data in a more controlled and transparent way, finally, promoting the protection and legal use of user privacy in the era of privacy awakening[2].

2 RELATED WORKS

Estimating the valuation of personal data is an important problem in economics and HCI. Typically, researchers achieved this by measuring two kinds of metrics: *Willingness to Pay (WTP)* [2, 74] and *Willingness to Accept (WTA)* [17, 38]. WTP quantified the price an individual was willing to pay to protect his/her data or privacy from being used or sold to third-parties (e.g., Facebook data [9], school evaluation [15], smartphone permission [21], and location data [17, 18]). In comparison, WTA quantified the money or benefit a person believed can be gained by sharing his/her personal data or privacy (e.g., identity data [50], shared photos [12, 71], communication, location, media, and applications [71], camera, microphone, and GPS [58], battery [33] and wearable sensor data [40]).

While both metrics reflect users' perception of privacy or data value, they inherently differ in their assumptions about data ownership (user vs. service provider) [32]. Due to the *Privacy Paradox*, where consumers show high privacy concerns but often disregard them in practice [31], WTA typically exceeds WTP for similar data types [32]. Given the focus of this paper on evaluating user-generated smartphone data valuations, we exclusively used WTA in our study, aligning with existing personal data research [50, 71].

2.1 Valuation Methods of Privacy and Data

Researchers have developed various valuation methods for privacy and data, reflecting diverse pricing strategies and user study designs. From a technological perspective, researchers adopted differential privacy [83] or other designed algorithms [13] for determining price. In social science, common approaches involve privacy measurement scales such as the Westin Privacy Index [45], the Concerns for Information Privacy (CFIP) [70], and the Internet Users' Information Privacy Concerns (IUIPC) [54], which employ questionnaires to assess privacy concerns quantitatively. However, these stated concerns often do not match actual privacy-disclosing behaviors [36].

To mitigate bias, some methodologies involve expert guidance. The Take It or Leave It (TIOLI) method allows experts to set a price, then users decide whether to accept or reject it, simplifying implementation but potentially compromising accuracy [9]. The Becker-DeGroot-Marschak (BDM) method gives users more autonomy by letting them bid within an expert-defined range, although this can lead to biased bids due to the anchoring effect [9].

Auction-based methods, such as the Generalized Second-Price (GSP) [50] and reverse second price auctions [18, 71], do not require expert involvement. These methods ask users to bid the lowest price they would accept to sell their data, with winners chosen from the lowest bid upwards. In GSP, each winner is paid an amount matching the next highest bid, whereas in reverse second price auctions, all winners receive the lowest bid among non-selected bidders, enhancing incentives and potentially reflecting a more accurate user valuation. In our research, we adopted reverse second price auction as our valuation method.

[2] https://www.datagrail.io/resources/interactive/2022-consumer-privacy-survey/, accessed by 5th April, 2024.

2.2 Valuation of Smartphone Data

The extensive use of smartphones and their sensors facilitate detailed studies on user data valuation. Earlier studies focused primarily on location data , revealing factors like requester identity and data detail influencing disclosure decisions [16], yet they lacked in providing a monetary valuation. Other researchers estimated median bids for location data, linking them to user behaviours such as mobility and communication patterns [17, 18], but these studies were limited by their hypothetical bidding setups. They only collected bids through questionnaires and SMS messages, which was far from the real-life scenarios which used apps or conducted offline bidding.

Subsequently, Staiano et al. [71] assessed the value of smartphone data (communication, location, media and applications) over six weeks, finding unique daily events increase bid values and highlight a correlation with personality traits but not demographics factors. Although the data types were varied, this approach still involved non-realistic data collection methods, limiting its applicability to actual scenarios.

Prior research often portrayed smartphone data abstractly from a technical perspective, without addressing its practical commodification when valuing. Nguyen et al. [58] and Hosio et al. [33] investigated the valuation of specific sensors and smartphone batteries, considering factors like perceived value and personal sharing motives. However, their assessed sensors were limited and they still did not modeled data as commodities in social context to sell. The only work from the commodity perspective was from Kotut et al. [40], who used a third-party platform for trading hiker trail data, yet did not delve into its valuation. Designing personal smartphone data as commodities is closer to real-life scenarios and offers a conceptual representation that can enhance users' understanding, utilization, and management of this resource. *Thus, in contrast, our work introduces a novel application that evaluates smartphone data as real commodities within a practical, transaction-based framework on smartphone apps.*

2.3 *Data Context* and *Privacy Risk* Theory

We refer to data-privacy-related theories to guide our commodity design. When building values on top of user data, researchers suggested that conventional "privacy statement" was not comprehensible enough for users to protect their data [23]. To resolve this problem, Nissenbaum et al. [30, 65] proposed the theory of *Contextual Integrity*, which argued that the principle of "respect for context" in European legal regimes should be amended to state that companies should collect, use, and release personal data in a way that is consistent with the (social) context in which consumers provide it. Noah [61] present a survey method based on the Contextual Integrity privacy framework and apply this method to discover privacy norms in the smart home context. Based on this theory, the World Health Organization (WHO) and Microsoft [27] proposed the concept of *Data Context* to protect personal information, which used seven variables to describe data: data type, entity type, device type, collection method, data use, trust, and value exchange. However, these variables were designed in the context of technology system or platform [60] instead of social context. Therefore, they still cannot help normal users without technical background to understand data privacy. In addition to data context, both the US regulations [34], the EU Regulation [19] and researchers [23] also emphasized the assessment of *Privacy Risks* based on the specific scenarios of the data, which was defined as [34] the potential for emotional distress, physical, property, professional, or other harm to the user, either by itself or in comparison with other information. The researchers [48] investigated the study of privacy risks from the perspective of the participants - in the context of a large-scale sensor data collection having participants assess the privacy risks at the time of dataset collection. They expressed concerns about privacy risks, and privacy risks prompted them to reconsider their decisions regarding data collection and public release.

The *Data Context* and *Privacy Risk* theory reinforced the traditional informed consent and helped the users to better understand their data, and served as mechanisms for information protection [23]. **Therefore, we designed**

our data commodities according to these theories, which could help to reveal the most accurate results on users' perceptions of data privacy.

3 DESIGNING SMARTPHONE PERSONAL DATA AS COMMODITIES

In this paper, we used reverse second price auction [18, 71] to value users' personal data as commodities on smartphones, which ensured that the bid price truly reflected the users' valuation. In addition, we employed WTA in our study as it was more commonly adopted to assess the valuation of the participants compared to WTP [77]. In this section, we introduced the two-step procedure of designing personal data as commodities, which is a key component of the auction: 1) choosing the personal data, and 2) designing the commodity attributes.

3.1 Step 1: Choosing the Personal Data on Smartphones

In order to cover common data types on smartphones, we collected a total of 14 data types. They not only included the types of data that have been valued in existing works (e.g., location [17, 18, 21, 71], camera [21, 71], call logs [21, 71], and app usage [71]), but also included more common sensors in daily lives (e.g., WiFi and touchscreen [25]). The detail considerations of the data was shown in Section A.2 in the appendix. Past works with contextual sensing also focused on these sensors (e.g., Ferreira et al. [25] focused on 12 types of sensors including accelerometer, barometer, gyroscope, etc., which we all covered), thus the valuation of these potentially collected information was necessary and important.

The 14 types of data correspond to 14 different sensors, capturing information about the users' environment (e.g., barometer, temperature, compass and light) and their personal behavior (e.g., accelerometer, gyroscope and touchscreen). Therefore, these data were highly private, but also valuable for third-party usage (e.g., advertising [49, 64], authentication [14], and behavior detection [14, 29, 63]).

3.2 Step 2: Designing the Commodities

While most smartphone users are somewhat familiar with sensor data, selling it as a commodity is uncommon in real life. Designing these data as commodities can significantly facilitate the selling process and help reveal the authentic valuation of users' private data. Prior research [27] indicates that how data is described significantly affects users' perceptions, thereby influencing the validity of valuation results. In designing our data attributes, we drew on the *Data Context* [27, 30, 65] and *Privacy Risk* [23, 34] theories to help users correctly understand data usage and associated privacy risks. Also, we referred to the description of physical commodities on online markets (e.g., Amazon and Taobao) to further modify the attributes. This can help the commodities become more compatible with the auction mechanism and help the users to value their personal data accurately.

3.2.1 Designing the Commodity Attributes. The *Data Context* theory [27] defined seven variables to describe the user's perception of a particular personal data usage scenario. However, as these variables were designed in the context of technology systems or platforms, it was not suitable to use them directly in our user study. Given our focus on smartphone data, we adapted these into six new attributes:

- *Data Type*. This attribute was derived from "data type", which corresponds to the type of the personal data.
- *Description*. This attribute was derived from both "data type" and "type of equipment". For each data type, this attribute contains a more detailed description about the collected personal data on the user's smartphone (see Table 1).
- *Permission*. This attribute was derived from "collection method", which specifies whether obtaining this data requires permission in the Android system. The value was "yes" or "no".
- *Buyer*. This attribute was derived from "entity type", which described the type of data buyers. In our study, as the data was bought from the users to conduct scientific researches, the value of this attribute was fixed to be "a laboratory".

- *Purpose*. This attribute was derived from "data usage". According to existing works, there were four major kinds of purposes: 1) physical and psychological research [14, 29, 57, 63]; 2) personalized advertisement [49, 64]; 3) fitness recommendations [59]; and 4) improving user experience [24].
- *Period*. This attribute was also derived from "data usage", which quantified the length of time during which the data was collected (measured in number of days, similar as in related works [17, 18]). In our study, the value of this attribute was fixed to be "7 days".

Besides the above six attributes, we also added another two attributes that was necessary for our study. The first is "Commodity Name". According to the *Data Context* theory [27, 30, 65], the range of the above variables should be determined in specific social domains. In real life, the same personal data on smartphones could be used to infer different behavioral, psychological, or environmental aspects of an individual in different social domains, which lead to intrinsically different commodities. Therefore, we defined commodity name to describe the inferred information to facilitate the users' comprehension. To cover as many social domains as possible, we referred to existing works of the corresponding data types [14, 29, 57, 63], and envisioned additional usage of the data as the complement. In total, we designed 49 commodities with different names (see Table 5 and 6 in the appendix). The second is "Privacy Risk". *Privacy Risk* theory pointed out that displaying privacy risk was very helpful for users to understand and assess the privacy of their data [34, 74]. Thus, we also involved this attribute in our design. Privacy risk reflects the level of risk that the users perceived regarding their private personal data. We used the commonly used five-point likert scale to quantify this attribute, where 1 and 5 represented the lowest and highest privacy risks, respectively.

In total, we designed eight attributes: commodity name, data type, description, permission, buyer, purpose, period, and privacy risk.

3.2.2 Effect of Commodity Attributes on Privacy Risk. In order to determine the privacy risk level of the commodities, we recruited 10 master/Ph.D. students specializing in HCI (7 male, 3 female) with an average age of 24.5 (SD = 2.1). Their research focused on sensors, interaction, and privacy, with an average publication count of 2.3 HCI-related conference papers (e.g., CHI, Ubicomp, CSCW; SD = 0.8). During the study, the participants were firstly asked to read and comprehend the definition of "privacy risk" in the U.S. Draft Consumer Privacy Bill of Rights [34] and the EU General Data Protection Regulation [19]. They then rated the privacy risk of the 49 commodities according to the value of the attributes, as described above. Across all the commodities, the Cronbach's α coefficient [46] was 0.9, confirming the internal consistency of the ratings. Finally, we set the privacy risk value of each commodity as the mean value of the 10 participants and rounded them (see Table 5 and Table 6 in the appendix).

As the privacy risk of each commodity was rated according to the other seven attributes, we performed an analysis of the effect of these attributes on privacy risk. As the value of "buyer" and "period" was identical for all commodities, we excluded them from the analysis. Furthermore, as the distribution of "data type" was very unbalanced (e.g., only 1 for magnetic field and barometer, and only 2 for camera, gyroscope and touchscreen), we did not test "data type" and "description". Finally, we tested "purpose" and "permission", as showed in Figure 1.

The Mann-Whitney U test indicated a significant impact of "permission" on privacy risk ($U = 76.5$, $p < .001$), with commodities requiring permission rated as more sensitive (average risk rating of 3.7) compared to those that did not require permission (average rating of 2.3), aligning with findings from previous research [21].

We perform a Friedman test on "purpose", which also showed a significant effect on privacy risk ($\chi^2(3) = 0.925$, $p<.05$). However, *post hoc* analysis with Bonferroni correction only found a significant difference between "physical and psychological research" and "improving user experience" ($p < .05$). "Improving user experience" received the lowest privacy risk value (1.8), followed by "fitness recommendations" (2.8). Notably, "personalized advertisement" and "physical and psychological research" both scored the highest at 3.4.

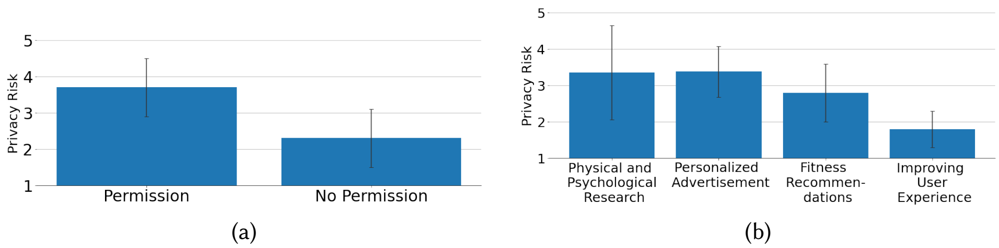


Fig. 1. Mean privacy risk of the commodities with different (a) permission and (b) purposes. Error bar showed one standard deviation.

3.2.3 Justification of Privacy Risk. Although we assigned the aspect "privacy risk", it is worthy to note that privacy risk did not mean the sum of the rest attributes. In fact, privacy risk evaluated the potential privacy loss our product would impose on users. The other perspectives of the product would also affect the price valuation of the product.

Besides, phone manufacturers would also give privacy risk valuation regarding different application permissions. For example, Android provided the detailed information about application permissions and classified the privacy risk into two classes: restricted and dangerous. This was similar to our privacy risk valuation. Thus, the presentation of privacy risk facilitated a more realistic price valuation and bidding.

4 USER STUDY: A REVERSE SECOND PRICE AUCTION

In this section, we executed a large-scale user study to investigate how users value their personal data on smartphones. Following the approach of prior studies [71], we set up a scenario where a laboratory aimed to purchase participants' smartphone data for scientific research. To authentically reveal the participants' data valuation, we utilized the commodities designed in the previous section and conducted a reverse second price auction where participants placed bids on their data. Unlike simulated auctions used in other studies, we actually compensated the winning bidders to purchase their data, ensuring a more realistic auction experience and addressing potential discrepancies between participants' stated behaviors and actual decisions [36].

4.1 Participants

We recruited 181 university participants (69 males, 112 females), all of whom used Android-based smartphones daily. The participants consisted of freshmen with an average age of 18.6 (SD = 0.9), including 113 Han Chinese and 68 ethnic minorities. Their average monthly living expenses were approximately 1,241 CNY (about 171.51 USD), with a standard deviation of 552 CNY (76.29 USD).

4.2 Experiment Design

Ideally, this study would have used all 49 commodities designed in the previous section. However, our pilot study revealed that evaluating 49 commodities could take participants over 1.5 hours, potentially compromising the validity of the results. Consequently, we opted to use a subset of these commodities.

To maintain balanced sampling, we ensured equal representation from each privacy risk level, which integrates all other attributes. We selected three commodities from each of the five privacy risk levels, totaling 15 commodities. These were chosen to cover all 14 types of data listed in Table 1. The selected 15 commodities are detailed in Table 1.

Table 1. The 15 commodities that were used in the user study. We did not list the buyer and the period because the values of these two attributes were the same for all the commodities.

Commodity Name	Data Type	Permission	Privacy Risk	Purpose	Description
Your working hour	GPS	Yes	3	Physical and psychological research	Longitude and latitude measured by GPS of phone
The number of Bluetooth devices around you	Bluetooth	Yes	3	Personalized advertisement	The number and time of Bluetooth connected devices
The outdoor temperature at your location	Temperature	No	2	Fitness recommendations	Outdoor temperature measured by phone
Which contact you receive the most calls	Call records	Yes	5	Physical and psychological research	Number and duration of calls
The horizontal direction of your phone	Compass	No	1	Physical and psychological research	The direction measured by the compass in the phone
The altitude at your location	Barometer	No	1	Physical and psychological research	The altitude of the mobile phone location measured by the barometer
The length of time you spend on online shopping	App usage	Yes	5	Personalized advertisement	App usage time recorded by app log
Your facial expression	Front camera	Yes	4	Physical and psychological research	Facial photos taken by the front camera
Your surroundings	Rear camera	Yes	4	Physical and psychological research	Environmental video taken by the rear camera
The length of time you drive	Accelerometer	No	3	Fitness recommendations	The phone's movement degree
Your sleeping environment's light	Light	No	2	Physical and psychological research	The brightness of the light on the front of the phone
Your emotion	Microphone	Yes	5	Physical and psychological research	Sound around the phone
Your location	WiFi	No	4	Fitness recommendations	Number and time of mobile WiFi connections
The spatial orientation of your phone	Gyro	No	2	Improve user experience	Rotation and tilt angle of the phone
Your unintentional touches when the phone was placed in the pocket	Touch screen	No	1	Improve user experience	Actions on the touch screen

In our study, we requested that participants place their bids *before* data collection, diverging from prior practices such as those in [71]. Earlier research [18] demonstrated a positive correlation between the bid amount and the extent of user behavior (i.e., information) captured in the data. Bidding after data collection, like after a week's worth of data which might include unexpected events such as attending a social event, could introduce significant bias. By having bids placed beforehand, the prices reflect expectations of future behaviors, resulting in a more stable and less biased valuation.

To emulate a more realistic data transaction scenario, winners were paid the actual monetary incentive after bidding. Additionally, participants were allowed to withdraw from the study at any time, whether before data collection or before final payment, mirroring real-world conditions where users can choose not to sell their personal data. We also monitored the number of participants remaining at various stages of the study to gain insights into their perceptions of data privacy.

We gathered participants' demographics as detailed in Section 4.4 and assessed their psychological attributes using the Chinese Big Five Personality Inventory (CBF–PI) [78], as all participants were located in China. The Big Five model is widely used to describe personality traits [4, 37] and is known for its cross-cultural validity [81]. The BFI-2 questionnaire results outlined participant traits across five dimensions—neuroticism, conscientiousness, agreeableness, openness, and extraversion, each scored on varying scales with higher scores indicating stronger tendencies in the respective dimension.

4.3 Experiment Platform

To facilitate real-data transactions, we developed an Android application using Android Studio with three main functions:

User Guidance: The app features guidance pages that explain the concept of personal data transactions, the bidding process, the study's objectives, the 14 data types involved, an overview of the commodities, the mechanics of the reverse second price auction, and the privacy policy governing the data (see Figure 9 in the appendix).

Bidding: During the bidding phase, users enter the registration page (see Figure 2 (a)). The app then displays the 15 commodities in a list format similar to online shopping platforms (see Figure 2 (b)). Participants can select any commodity to view its attributes and place their bid (see Figure 2 (c)). They can enter any positive price, but once entered, the bid cannot be changed.

Data Collection and Transaction: After the bidding winners are determined (see Figure 2 (f)), they can view the results in the app (see Figure 2 (d)) and decide whether to sell the data at the bid price. If they choose to sell, the app collects the corresponding data from their smartphone over the next 7 days, as outlined in the "period" attribute. Users generally are not aware of the sensor tracking granularity, so this was not considered in our experiment. Winners then select "Receive Benefit" to receive their compensation through WeChat Payment, and they can view the current status of the commodity transaction (see Figure 2 (e)). As the auction concluded before any data collection, no participant data was displayed in the app.

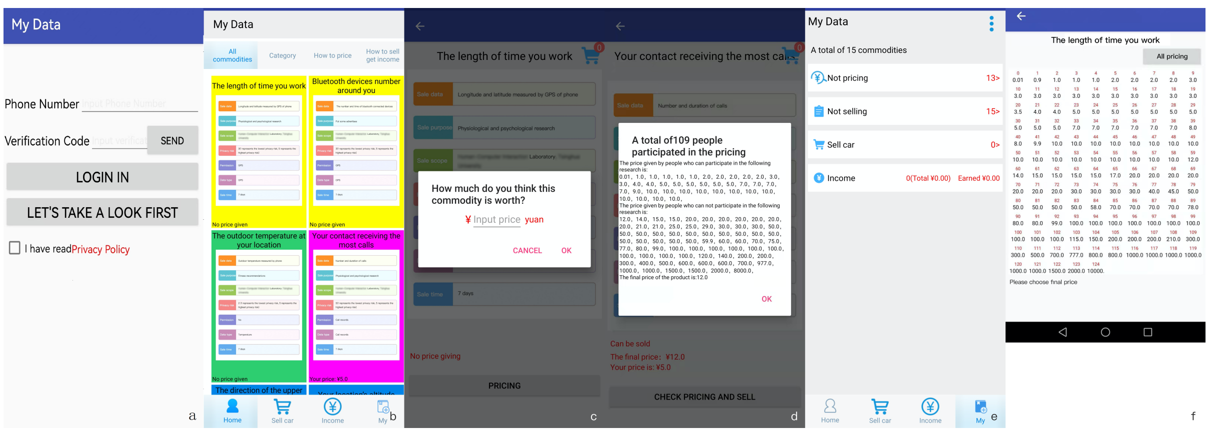


Fig. 2. The user interface of the experiment platform. The original page is in Chinese and we translated that to English. (a) The Login page. (b) The bidding page that showed the list of commodities. (c) The bidding page for entering the bid price. (d) The checking page of the result of winner and the selling page of commodities. (e) The status of the commodity. (f) The page determining the winning of the bidding.

Besides, as we used a bmob[3] cloud server as the backend, after collecting the data, the data would be delivered using encrypted HTTPS protocol via POST method to the server. After the experiment, the experimenters do collect and download the data. However, experimenters did not use them or share the data to any other servers. The experimenters would promptly delete all the data immediately after the experiment.

4.4 Procedure

The experiment procedure consisted of 6 stages (See Figure 3):

Stage 1: Initial Explanation. We orally presented the goals and procedures of the user study during an online meeting and provided additional text descriptions via WeChat groups. The app's privacy statement, which also served as the user consent form, detailed data processing and collection while disclosing potential risks associated

[3]https://www.bmobapp.com/

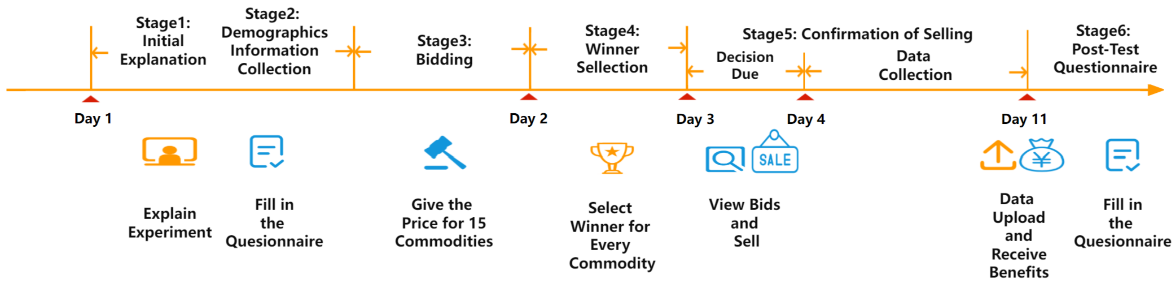


Fig. 3. The experiment procedure.

with the app (discussed in Section 9). Participants were required to sign this consent before participating in the experiment.

Stage 2: Demographics Information Collection. Participants installed the app, registered an account, and completed an online questionnaire to collect demographic information, including daily smartphone usage (see Table 9 in the appendix), familiarity with the 14 data types, Big5 personality traits [78], and their perceptions of smartphone privacy [71] (see Table 10 in the appendix).

Stage 3: Bidding. Following the questionnaire, participants were informed about the start of the bidding process. They had 24 hours to submit their bids for all 15 commodities on the app, instructed to fully understand each commodity based on its attributes and bid independently without discussing with others. We sent three reminders during the bidding phase via the WeChat group to ensure participants remembered to place their bids. If participants chose not to sell a specific commodity, they could opt not to place a bid for it. As confirmed in Sections 5.2 and 6.3.2, all participants bid on at least one commodity, demonstrating the experiment's validity.

Stage 4: Winner Selection. Following the reverse second price auction method [18], we determined the winners for each commodity and set the prices within 24 hours post-bidding. Unlike previous studies where the number of winners varied [18, 71], we limited the number of winners to one-third of all bidders due to budget constraints. The final price for each commodity was set at the lowest bid among the non-selected bidders.

Stage 5: Confirmation of Selling. Participants were able to view the results, including selected bidders, all bid prices, and final commodity prices. Within 24 hours, winning bidders could confirm or cancel their intention to sell their data. Upon confirmation, the app collected the relevant data from their smartphones over the next 7 days and transferred it to the laboratory server, and the winners received their payment. If participants did not successfully sell any commodities, they received a small gift; however, if they sold at least one item, they did not receive the gift. Notably, we do collect participants' data to better simulate the real transaction scenario. In fact, some participants in the study do examine the permissions that we ask them to turn on in order to collect the data. If the data was not collected, the participants' mental model about whether the data is collected would change, thus affecting the final post-test questionnaire and payment receiving process. Participants also reported that if the data is not collected, they may have less privacy perception compared with a real case. This echoed past literature that people's valuation can be quite different in hypothetical settings than in real situations with data collection and money transaction [51, 52]. **As described in Section 4.3, we tried our best to design the experiment in order to protect users' data. We used TLS 1.3 encrypted HTTPS transmission, and deleted the data immediately after the collection phase.** Thus, our setting is more practical and ensures less negative effect compared with a deceptive design.

Stage 6: Post-Test Questionnaire. Following the confirmation process, all participants were asked to complete a post-test questionnaire that explored the factors influencing their decision to sell their data.

5 RESULTS AND ANALYSIS

In this section, we reported the decisions that the participants made during the user study, and analyzed the factors that affected these decisions. After that, we reported the bidding prices of the commodities, which reflected the participants' valuation of the corresponding personal data. We also built a regression model between different factors and the bid prices, which helped explain the results. Finally, we analyzed the results from the post-test questionnaire, which uncovered additional findings besides the data. As our data violated the assumption of normality, we used non-parametric testes for the test of significance during analysis. All *post hoc* analysis was performed with Bonferroni correction. As the experiment was conducted in the summer of 2023, the exchange rate stood at 7.12 CNY to 1 USD at that time. Consequently, all monetary values denominated in CNY could be converted to USD at this prevailing exchange rate.

5.1 Statistics of the Bidding

We analyzed the number of bids that each participants gave, as well as the bid prices. We observed that some participants would give unreasonable high bids when they refuse to sell the data. Therefore, in order to remove these outliers, we considered all bid with price > 1,000 CNY (138.20 USD) as "unwilling to sell", and did not count them as valid bids. In total, we removed 70/1,918 = 3.6% from participants (ranged from 1,500 CNY , 207.3 USD to 1E+28 CNY ≈ 1E+27 USD), covering 15 commodities.

5.1.1 Ratio of bids. Figure 4 showed the ratio of bids for each commodity, which was calculated as the number of bids received from the participants, divided by the total number of participants. On average, the big ratio was 68.1% (SD=12.2%).

The commodity with the highest bid ratio was "The outdoor temperature at your location" at 88%. The top three commodities participants bid on included this temperature data, "The altitude at your location" and "The number of Bluetooth devices around you". Conversely, the four commodities with the lowest bid ratios were "The length of time you spend on online shopping", "Your facial expression", "Your surroundings" and "Your emotion", indicating these were perceived as more private.

Further analysis of bidding behavior showed participants bid on an average of 10.3 commodities (SD=4.8), with 58 participants (39.3%) bidding on all 15 commodities, demonstrating a strong willingness to sell their smartphone data. The decision to bid on a commodity was not significantly correlated with participants' familiarity with the data ($R^2 = 0.14$). Interestingly, all participants, without requirement, bid on at least one commodity, suggesting that the concept of selling personal smartphone data was generally acceptable to them.

5.1.2 Winners and commodity prices. For each commodity, we designated 1/3 of the bidders as winners. According to the principles of reverse second price auctions, the price for each commodity was set to the lowest bid among the non-selected bidders, as detailed in Table 2. Interestingly, we observed no significant correlation between the commodity price and the bid ratio ($R^2 = 0.65$), indicating that commodities receiving more bids did not necessarily result in lower prices. Additionally, there was no significant correlation between commodity prices and participants' familiarity with them ($R^2 = 0.27$). As shown in Table 2, "Your facial expression" and "The length of time you spend on online shopping" were among the commodities with the highest prices. A more detailed analysis of the bid prices will be provided in Section 5.2.

5.1.3 Compensation. Before actually collecting the data, we allowed the winners to cancel the transaction for each commodity. During the study, 9/108 participants canceled the selling of at least one commodity at this stage, possibly due to last-minute concerns about privacy. 73 participants did not win any commodity, thus

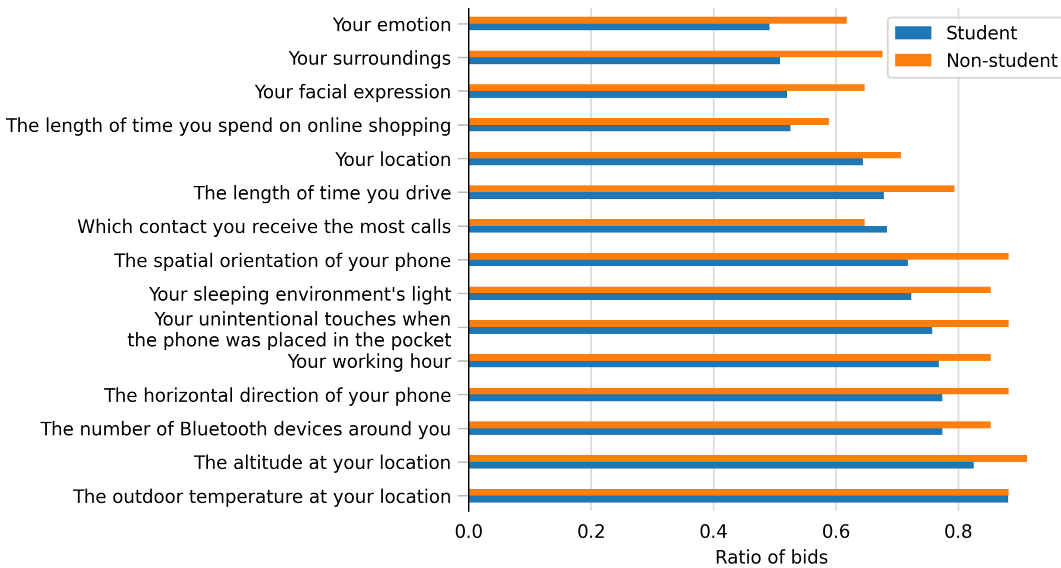


Fig. 4. The average ratio of bids for the 15 commodities.

Table 2. The final bidding price (in CNY) of the 15 commodities.

Commodity	Student	Non-student	Non-student/Student×100%
The length of time you spend on online shopping	15	30	200%
Your facial expression	15	40	267%
Which contact you receive the most calls	10	20	200%
Your surroundings	10	21	210%
Your location	10	30	300%
Your emotion	10	30	300%
The number of Bluetooth devices around you	8	20	250%
The length of time you drive	7	30	428%
Your sleeping environment's light	7	20	286%
Your working hour	7	20	286%
The spatial orientation of your phone	6	20	333%
The outdoor temperature at your location	5	10	200%
The horizontal direction of your phone	5	20	400%
The altitude at your location	5	10	200%
Your unintentional touches when the phone was placed in the pocket	5	10	200%

only received some snacks as compensation, valued at 1 CNY (approximately $0.15), significantly less than any winning bid. Based on the set prices of the commodities and the final number of sellers, we distributed a total of 4,405 CNY to participants. The average payment received by each participant was 44.5 CNY, with individual amounts ranging from 5 to 125 CNY.

5.2 Analysis on Bid Prices

5.2.1 Distribution of bid prices. We analyzed the distribution of bid prices to understand their privacy valuation of the data. Table 11 and Table 12 in the appendix showed the detailed statistics of the bid prices for all the commodities. Figure 5 showed the average bid price of the 15 commodities. The mean bid prices across all commodities ranged from 25.8 to 127.8 CNY.

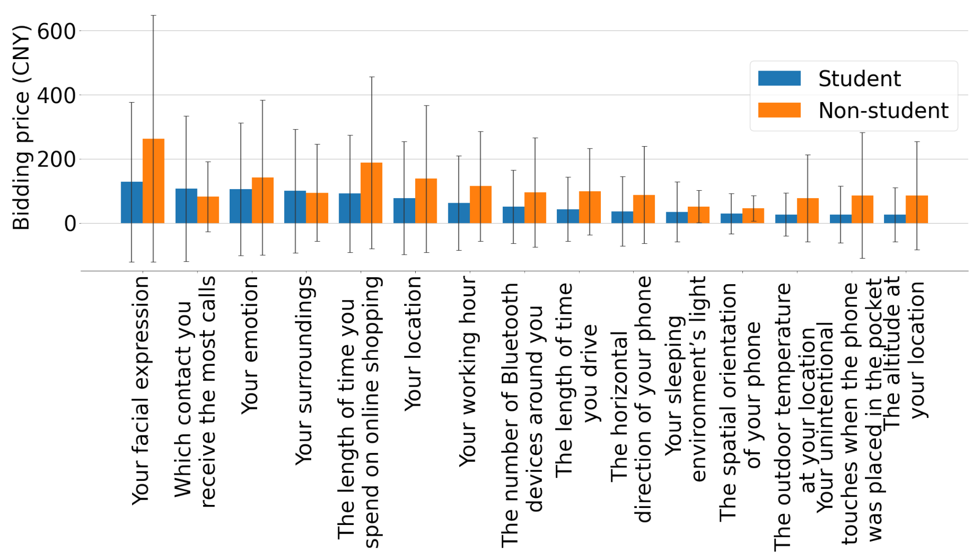


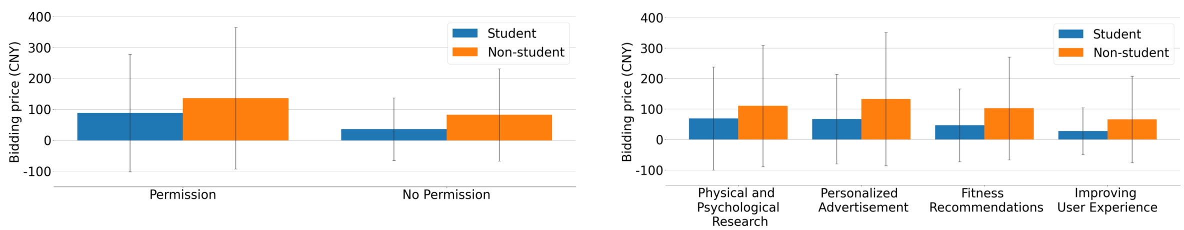
Fig. 5. The average bidding price of the 15 commodities. Error bar showed one standard deviation.

The mean bid price of the commodities correlated significantly with the final prize ($R^2 = 0.78$). The top three commodities with the highest mean bid prices were "Your facial expression" (127.8 CNY), "Which contact you receive the most calls" (107.1 CNY), and "Your emotion" (104.5 CNY). Conversely, the commodities with the lowest mean bid prices were "The altitude at your location" (25.8 CNY), "Your unintentional touches when the phone was placed in the pocket" (26.0 CNY), and "The outdoor temperature at your location" (26.6 CNY). Notably, six student participants (0.3% of the sample) placed a total of 29 very low bids (ranging from 0.01 to 0.99 CNY). Interviews with these participants indicated that these low bids reflected a willingness to essentially give away the data for free.

The commodity with the highest bids from students was "Your facial expression"(1E+28 CNY) and from non-student was "Your Emotions"(480000 CNY). We asked the highest bidders, both students and non-students, for their reasons via telephone interviews. We interviewed two types of users: one was ridiculously high (1E+28 CNY), and they said they wouldn't sell their data for any amount of money, so it was too high to win the bid. One kind of bid is a few hundred to a thousand CNY or so, they said they can sell it, but it's worth that much, because the data is very private and important to them. The lowest bidder indicated that he had checked the price of the data online and also indicated that it would be free if used for research. Another argument was that their low bids on some commodities were due to the fact that the data of those commodities had been leaked over in their lives.

5.2.2 Effect of commodity attributes. To further explore the effect of the commodity attributes on the participants' bid price, we performed a series of analysis on the collected data. As the value of "buyer" and "period" were the same across all the commodities, we focused on analyzing the remaining attributes: "data type", "permission", "purpose" and "privacy risk". According to Kruskal-Wallis H test, data type yielded a significant effect on the bid price for participants ($\chi^2(14) = 112, p<.001$), reflecting participants' concerns about the data type.

Figure 6a showed the bid price for commodities that require permission or not. For participants, the bid price for commodities that require permission was significantly higher than those who did not ($U = 314835$, $p < .001$), with a mean value of 88.1 v.s. 36.2 CNY. This suggested that the users considered data that required permission to be "more risky" and more private. Existing works [21] have also found that the users would give higher bid values for data that required more permissions.



(a) (b)

Fig. 6. The average bidding price for commodities with different (a) permission and (b) purpose. Error bar showed one standard deviation.

Figure 6b showed the bid price for commodities with different purpose. For participants, the purpose with the highest and lowest bid price was "Physiological and psychological research" and "improving user experience", respectively. A significant effect of purpose on the bid price was found ($\chi^2(3) = 25.2, p<.001$). *Post hoc* analysis revealed significant differences between "Improving user experience" and both "Physiological and psychological research" ($p < .001$) and "Personalized advertisement" ($p < .05$).

Figure 7 showed the bid price for commodities with different levels of privacy risks. For participants, the bid price increased monotonically with privacy risk, confirming that users would give higher bids for commodities that were more private. According to Kruskal-Wallis H test, a significant effect of privacy risk was found on the bid price ($\chi^2(4) = 105, p<.001$). *Post hoc* analysis found significant differences between the commodities with privacy risk ≤ 3 and those with privacy risk ≥ 4 ($p<.001$).

5.2.3 Regression model of commodity prices. Using the collected data, we employed stepwise regression[4] to develop a regression model for predicting commodity prices, a method commonly used in previous research (e.g., [22, 28]). This model not only predicts prices but also clarifies the impact of various factors, excluding variables that did not significantly influence the outcome from the final model. We designed 14 variables for student participants as showed in Table 3.

In the regression process, we input raw values for continuous variables, such as age. For ordinal variables like hours per day using a smartphone, we assigned values from 1 to n corresponding to different levels, where

[4]https://en.wikipedia.org/wiki/Stepwise_regression

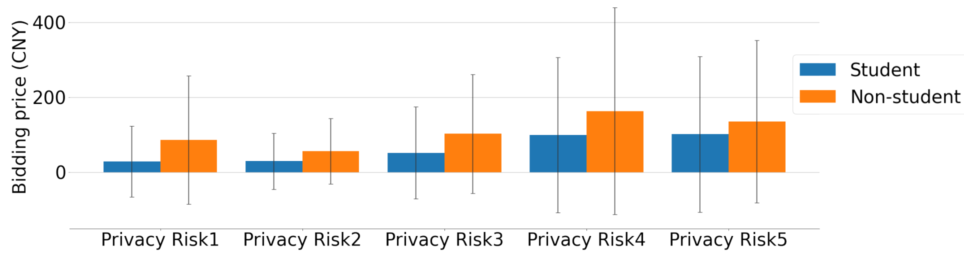


Fig. 7. The average bidding price for commodities with different privacy risks. Error bar showed one standard deviation.

Table 3. The variables used in the price regression model, as well as the value ranges. Red and blue indicated the variables that were specific to student and non-student participants, respectively.

	Continuous variables	Ordinal variables	Categorical variables
Variables associated with commodities		Privacy risk (1-5)	Permission (yes, no), Purpose (physiological and psychological research, personalized advertisement, fitness recommendations, improving user experience)
Variables associated with individuals	Age, Monthly living expenses, Monthly income	Hours per day using a smartphone, Neuroticism (13-43), Conscientiousness (13-43), Agreeableness (17-32), Openness (8-48), Extroversion (18-38), Familiarity to the data types (1-5)), Education (elementary school 1, middle school 2, high school 3, bachelor's degree 4, master's degree 5, doctoral degree 6	Ethnicity (Han Chinese 1, minority 0), Gender (Male 1, Female 0), Marriage (married 1, unmarried 0), Children (yes 1, no 0), Working in IT (yes 1, no 0)

n is the number of levels. Categorical variables, such as purpose, were converted into $n-1$ dummy variables, each represented by boolean values to reflect the categories of the original variable. The model's output offers a predicted valuation (i.e., price) of each commodity for a given participant.

Table 4 showed the regression model for participants. The R^2 value of the model was 0.068. These values were acceptable in the field of social sciences [43], as the effect of the large amount of variables were usually complex. In comparison, existing regression models on the valuation of identity information [50] and privacy behavior [26] yielded R^2 between 0.045 and 0.269, which was similar to our results.

As observed in Section 5.3.2, the models showed that "privacy risk" significantly affected participants' valuation of the commodities, which aligns with the design intention of "privacy risk" as a composite of other attributes. Additionally, commodities purposed for "physiological and psychological research" received significantly higher valuations.

Demographic factors such as ethnicity and gender affect participants' valuations. Additionally, participants with higher conscientiousness scores value their privacy more. Previous studies [4, 37] have also found that

Table 4. The regression model for students and non-students. Significance level: $^{*}\alpha = .05$, $^{**}\alpha = .01$, $^{***}\alpha = .001$

	Regression model	R^2	Adjusted R^2	F	p
Student	$Price = -42.5 + 21.4 \cdot \text{Privacy risk}^{***} + 28.9 \cdot \text{Han Chinese}^{***} + 29.3 \cdot \text{Male}^{***} + 19.8 \cdot \text{“Physiological and psychological research”}^{***} + 1.9 \cdot \text{Conscientiousness}^{*}$	0.068	0.065	25.8	$< .001$
Non-student	$Price = 346.6 - 3.6 \cdot \text{Age}^{***} + 21.9 \cdot \text{Privacy risk}^{**} - 4.5 \cdot \text{Conscientiousness}^{*} - 42.8 \cdot \text{Male}^{**}$	0.099	0.089	10.0	$< .001$

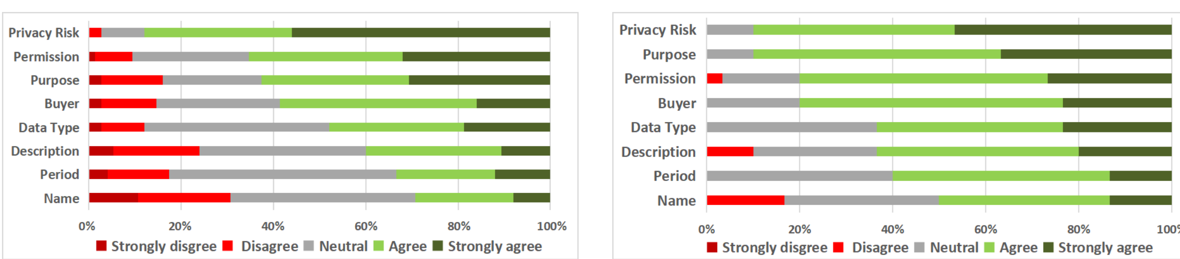
conscientiousness influences privacy concerns. Similar to the effects of age and race noted in past literature [50], this study identified significant effects of ethnicity and in students, which probably because ethnicity influences privacy attitude and consumption habits before graduation [35].

5.3 Effect of Minor Factors through Questionnaire Results

5.3.1 Effect of buyer. In the user study, we only tested one kind of buyer (a research lab), as we designed to perform realistic data transaction. However, different buyers could also affect the participants' willingness of selling the data. To explore this, we asked the participants to rank the priority of selling their smartphone personal data to different buyers. We designed a total of five kinds of buyers. Three of them were from existing works [71] (banks, telecommunication companies, and insurance companies). We also added research institutions and big internet companies.

We calculated the average rank for each kind of buyers. The buyer that they were the most willing to sell data to was banks, followed by research institutions, telecommunication companies, big internet companies and insurance companies. This result was in consistent with existing works [12, 71].

5.3.2 Effect of the commodity attributes on whether to sell. In Section 5.1.1, we observed that some participants did not bid for all the commodities, as they refuse to sell specific commodities. Therefore, after the user study, we used a questionnaire to examine which of the eight commodity attributes affected their decision on whether to sell it. The answers resulted in a rating from 1 (strongly disagree) to 5 (strongly agree). Figure 8 showed the distribution of the answers from participants. The order of the factors that most affected the participants' decision was privacy risk, permission and purpose.



(a) (b)

Fig. 8. The participants' response to the commodity attributes that influenced their decision on whether to sell the commodity. (a) Students. (b) Non-students.

6 GENERALIZATION AND VERIFICATION ON PARTICIPANTS WITH VARIOUS DEMOGRAPHICS

In this section, we further conducted a study on non-students to verify whether the results derived in the previous section is effective for other demographics.

6.1 Participants and Apparatus

We recruited 34 non-student (14 male, 20 female) participants. All of them used Android smartphones on a daily basis. The non-student participants were recruited from a capital city of a province using the snowball sampling method [1]. Their ethnic composition consisted of 31 Han Chinese and 3 ethnic minorities. The average age was 39.9 (SD=10.8). Their average monthly income was 6,066 CNY (SD=3,360 CNY). Among them, 8 had a bachelor's degree or above, and 14 worked in the field of information technology (see Table 8 in the appendix for details).

6.2 Study Design and Procedure

We used the same study design and procedure in the main study, comprising six stages. In fact, the experiment was conducted in the same time along with the main study although they were conducted in a separately manner.

6.3 Results

6.3.1 Statistics of the Bidding. Consistently, bids exceeding 1,000 CNY were classified as "unwilling to sell" and excluded from the analysis. We discarded 9 out of 407 bids (2.2%) from non-student participants (ranged from 2,000 to 480,000 CNY) participants, covering 10 commodities.

Ratio of Bids The average bid ratio for non-student participants was 77.8% (SD = 11.7%). Compared to student participants, non-student participants showed significant higher tendency of selling their smartphone personal data ($U = 65.0$, $p < .05$).

The commodity with the highest bid ratio was "The altitude at your location" (91%) for non-students. The top-3 commodities of student and non-student participants did not overlap. However, although with different orders, the top-4 commodities with the lowest bid ratio was the same for student and non-student participants (See Section 5.1.1), indicating that they both perceived these data to be more private.

Focusing on the number of commodities bid on, non-student participants were generally more willing to sell their data than students, with an average of 12.0 commodities bid on (SD = 4.6), with 20 (58.8%) non-student participants bid on all 15 commodities. The decision to bid on a commodity was not significantly correlated with their familiarity with it ($R^2 = 0.26$). The fact that all participants sold at least one commodity confirms the wide acceptance of privacy valuation through bidding.

Winners and Commodity Prices Consistent with previous findings, no significant correlation was observed between the commodity price and the ratio of bids ($R^2 = 0.54$), indicating that higher numbers of bids did not necessarily result in lower prices. Additionally, non-student participants consistently offered higher prices than student participants ($U = 12$, $p < .001$). For all 15 commodities, prices from non-students were substantially higher than those from students, with a mean ratio of 271% and ranging from 200% to 428%. This was probably correlated with the anchoring effect of non-students [9].

Compensation During the study, 3/26 non-student winners canceled the selling of at least one commodity. Eight non-students did not win for any commodity, thus only received a small gift as compensation. In total, 2,707 CNY was distributed among non-student participants, with the average payment 117.7 CNY per participant (ranged from 10 to 331 CNY). The higher average payment could be attributed to the anchoring effect [75].

6.3.2 Analysis on Bid Prices. **Distribution of Bid Prices**

Across all the commodities, the mean bid price ranged from 45.6 to 263.1 CNY for non-student participants. Notably, non-students offered higher mean bid prices than students for 13 out of 15 commodities.

For non-student participants, no significant correlation was found between the mean bid price of the commodities and the final prize ($R^2 = 0.59$). The top-3 commodities with the highest mean bid price was "Your facial expression" (263.1 CNY), "The length of time you spend on online shopping" (187.9 CNY) and "Your emotion" (141.8 CNY), aligning with final commodity prices. Meanwhile, the top-3 commodities with the lowest median bid price was "The spatial orientation of your phone" (45.6 CNY), "Your sleeping environment's light" (50.9 CNY) and "The outdoor temperature at your location" (77.1 CNY). 1 (3%) of the non-student participants gave a low bid (0.01 CNY) for 1 commodity.

Effect of Commodity Attributes According to Kruskal-Wallis H test, data type yielded no significant effect on the bid price for non-student participants ($\chi^2(14) = 13.2, p = .50$). This implied that non-student participants were not so sensitive about the data types. During interviews, non-students expressed greater concern about the social context of information rather than the data types.

We further analyzed the bid price for commodities that require permission or not. For non-student participants, the bid price for commodities that require permission was significantly higher than those who did not ($U = 16789$, $p < .05$), with a mean value of 136.0 vs. 82.2 CNY.

Regression Model of Commodity Prices Table 4 showed the regression model for non-student participants. The R^2 value of the model was 0.099. Consistent with Section 5.3.2, the model indicated that "privacy risk" significantly impacted their valuation, with the effect strength appearing universally across different participant groups. No other attributes remained significant in the model for either group.

Given the different valuations between student and non-student participants, We developed separate regression models. Specifically, we designed 17 variables for non-student participants as showed in Table 3. Eleven of the variables were shared between both groups, including three variables associated with the commodity, and ten variables associated with the individual. Two and four variables were specific to student and non-student participants respectively.

In terms of demographics information, gender and conscientiousness influenced non-student valuations, though the effects were opposite to those observed in students, possibly reflecting changes in priorities post-graduation. Besides, younger non-students tended to value their data higher. It should be noted that since the number of non-student ethnic minorities and Han Chinese is unbalanced (3 vs 31), we did not include ethnicity as a dependent variable.

6.3.3 Effect of Minor Factors through Questionnaire Results. Similar to Section 5.3, we calculated the average rank for each kind of buyers. Surprisingly, the preferences for student and non-student participants were identical. Both groups were most willing to sell their data to banks, followed by research institutions, telecommunication companies, big internet companies and insurance companies. This result was in consistent with existing works [12, 71]. The average rankings was 1.9, 2.9, 3.1, 3.6, 3.8 for students, and 1.8, 2.8, 3.2, 3.5, 3.9 for non-students, respectively.

Figure 8 showed the distribution of the answers from non-student participants, respectively. Generally, the order of the factors that most affected the participants' decision was consistent between both groups. The only exception was that student participants considered "permission" more than "purpose", while non-student participants considered "purpose" more. For all participants, "privacy risk" was the top-1 attribute impacting their decision.

7 DISCUSSION

7.1 Feasibility of Selling Personal Data

Privacy trading was proposed as an alternative approach to protect personal information [3, 47]. While earlier research discussed the possibility [5, 11, 50], conducted privacy trading with other context [69] or investigating the data selling price with algorithmic methods [13, 83], our work was the first to make privacy trading feasible through bidding on apps. We designed smartphone data as commodities under the guidance of Data Context theory, which was commonly considered in investigating privacy norm [10, 44] but less investigated in valuation context [66].

In Section 5.1.1, our analysis reveals that all participants bid for at least one commodity, and approximately half of the participants (39.3% for students and 58.8% for non-students) even bid for all the commodities. This indicates the acceptance of data selling among the participants, which echoed previous literature [33, 71]. Furthermore, despite being influenced by the commodity attributes, participants exhibited a high willingness to sell specific commodities (above 40% in Figure 4). Compared with valuing location data in the social context [71], our valuation of different data types on various demographics considered various privacy risk. The bidding price of different data also have marked difference (see Figure 5). This comprehensive result suggests selling smartphone data is both acceptable and feasible in practical use.

This was significant given the growing global attention towards privacy protection in various regions (e.g., China[5], the U.S.A. [6], and European[7]). Our findings provided the possibility of protecting participants' data through valuing as commodities.

7.2 Bidding Price and Influential Factors

In this study, we adhered to the Data Context [27, 30, 65] and Privacy Risk [23, 34] theory to design eight attributes including "privacy risk". The resulting bidding prices were not consistent in contrast to previous research [71]. The most valuable data types, locations and applications, declined to the fifth and sixth rank in our valuation, possibly due to their widespread usage, leading users to become accustomed to their collection. Conversely, facial expressions extracted from media ascended from the fourth to the top position, possibly owing to the higher resolution and using frequency of facial data[8]. Furthermore, increasing advocacy efforts [41] have raised people's awareness regarding the importance of protecting facial data. The awareness of participants was also verified through questionnaire in our study, which privacy awareness rating (see Table 10 in the appendix) was consistently higher than the past work [71]. This could probably be attributed to increased literacy on privacy protection [68].

Personal data valuation, especially the bid prices was among the most significant results. We evaluated 15 types of smartphone personal data commonly mentioned in the commercial products and literature [53, 71]. 5 were previously assessed in other studies: "Your location" (WiFi) [17, 18, 71], "The length of time you spend on online shopping" (app usage) [71], "Which contact you receive the most calls" (call records) [71], "Your facial expressions" (front camera) [71] and "Your surroundings" (rear camera) [71]. The rank of different commodity is similar to the past literature focusing on information collection and sharing [42, 72]. Facial expression are valued [72] while location and surroundings are generally de-valued [42]. The mean bid prices in our study were higher, suggesting our commodity design effectively raised participants' privacy awareness and the increasing use of smartphones brought additional data concerns. However, students' bids for "Location" were lower than Staiano's [71] participants, suggesting a devaluation possibly influenced by familiarity and ubiquitous access. Beyond the five common commodities, our study included ten others covering typical smartphone personal data. Analyses revealed that "privacy risk", "purpose", and "permission" significantly influenced bid prices, with demographic features also playing a role, consistent with existing findings [50].

Contrary to previous studies [71] that did not find a demographic correlation with bid prices for location data, our research identified the link between bids and demographic characteristics, possibly due to 1) the participants in our study may have a more unified cultural background than the past [53], and our categorization

[5]https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2022/art_e0f06662e37140808d43d7735e9d9fd3.html, accessed by 1st May, 2024
[6]https://mp.weixin.qq.com/s/qPrR0mCsI-wvGa-oRYttzg, accessed by 1st May, 2024
[7]http://news.sohu.com/a/636656485_120076174, accessed by 1st May, 2024
[8]https://blogs.iadb.org/administracion-publica/en/face-id-is-our-biometric-facial-data-being-safeguarded/, accessed by April 20, 2024

into student and non-student groups, allowing for more detailed results. Within these categories, males bid higher than females, while the opposite was true for non-student participants. This is due to the fact that women enter society with more responsibilities in the family and thus are more conscious of their privacy than men. Younger non-students bid higher than older ones, which was consistent with Carrascal's [12] findings. While conscientiousness negatively correlated with bids among students, it positively correlated among non-students. Our study did not explore the relationship between participants' bids and behavior [18, 71]. However, we treated data as commodities and explored the relationships between pricing and 3 additional product attributes (e.g., privacy risk, sales purpose and permissions). These provide design implications for stakeholders to better regulate data usage.

7.3 Design Implications for Different Parties

We valued 15 commodities and identified different influential factors (see Section 5.2.2, 5.2.3 and 6.3.2). These findings provided the guidelines for stakeholders and governments.

We recommend app developers explicitly disclose privacy risks and data usage purposes for sensor data (e.g., accelerometers and cameras) within the system settings (Android[9] or IOS[10]). As the results indicated these attributes strongly influenced personal data bids and user willingness to share data (see Table 4), we recommend they should be stated in apps' privacy policies and highlighted for sensitive data.

Additionally, data should be categorized based on privacy risk and bidding price (see Table 2) during the app design process. For data with less valuation price (e.g., logs), app developers could present simpler notification mechanisms when collecting (e.g., op-up notifications), while for data with higher valuation price (e.g., camera, facial expression), more detailed notifications and repeated confirmations are recommended to enhance user awareness and consent.

For governments, users' bid prices on smartphone data could inform the structuring of penalties for illegal data trading [39, 80]. Governments could impose higher penalties for misuse of data types that attract higher bids and lower fines for those with lesser bids in our study, thereby enforcing data protection more effectively.

8 LIMITATION AND FUTURE WORK

We presents three limitations that pave the way for future exploration. Firstly, due to practical constraints in real data transaction settings (e.g., data buyer and period), we could not formally test all attributes. Instead, we analyzed their effects through questionnaires and interviews (see Figure 8). Nonetheless, our results demonstrated the impact of several attributes (e.g., privacy risk) on participants valuation of smartphone data commodity, contributing significantly compared to existing works [50, 71] which did not evaluate the effect of commodities' attribute.

Secondly, our study only involved Chinese users, warranting investigation into cultural variations to generalize findings. We made efforts to diversify participants, but recruiting non-students proved challenging, resulting in a smaller sample size. Analyzing data from a more balanced participant group may yield further insights.

Lastly, certain patterns observed in our main study (see Section 5) was not mirrored by non-students during verification (see Section 6.3). Non-students priced commodities with privacy risk 4 than with privacy risk 5. Interviews revealed non-students priced commodities based on various attributes, with privacy risk only as a reference. Our research highlighted the importance of understanding the common influential factors and advocating for tailored privacy approaches for different users (also see Section 7.2).

[9]https://developer.android.google.cn/
[10]https://developer.apple.com/

9 ETHICAL CONSIDERATIONS

We acknowledge potential ethical concerns regarding privacy in our experiments and took measures to mitigate them while respecting participants. Our experiment received approval from the Institutional Review Board (IRB) and adhered to ethical guidelines outlined in the Menlo Report [6] and Belmont Report [8].

Prior to the study, we carefully crafted questions to avoid revealing participants' real-world usage or personal information. Participants were briefed on the experiment's purpose and potential risks (See Section A.1), with consent integrated into the process (See Table 13 and Table 14). They retained the right to withdraw at any time and delete their data post-experiment. Income diversity was considered to ensure fairness, with compensation provided according to average income value. Data transmission and processing were conducted securely via HTTPS protocol, with encryption during transmission and offline analysis by experimenters only. Although we collected participants' smartphone data to ensure the realistic effect of the transaction, we transmitted them with secure protocols and retained them on a local server. We deleted all the data immediately after the experiment. Participants retained the option to contact the lab regarding any concerns arising from the experiment.

10 ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of China under Grant No. 62132010.

11 CONCLUSION

In this paper, we proposed a design that enables the buying and selling of personal data on smartphones as commodities. Based on the *Data Context* and *Privacy Risk* theory, we proposed 49 commodities accompanied by 8 commodity attributes that made the smartphone data more transparent and simple for the user to understand, which helped the users to weigh the pros and cons of selling more easily. We conducted a large-scale user study (N=215) and employed the reverse second price auction mechanism to allow participants to bid on 15 commodities. We regressed price models for students and non-students respectively, which revealed the commodity and demographic factors that affected their bidding price. Our results also found the commodity factors that influenced students and non-student' decision-making, including privacy risks, permissions, and purpose of use. Our work provided novel empirical findings for evaluating personal data on smartphones, and could serve as a practical reference for both stakeholders and end-users.

REFERENCES

[1] GOODMAN L A. 1961. Snowball sampling. *Annals of Mathematical Statistics* 32, 1 (1961), 148–170.

[2] Alessandro Acquisti and Jens Grossklags. 2005. Privacy and Rationality in Individual Decision Making. 3, 1 (jan 2005), 26–33. https://doi.org/10.1109/MSP.2005.22

[3] E. Adar and B. A. Huberman. 2001. A market for secrets. *First Monday* 6, 8 (2001), 200–1.

[4] Vinitzky G Amichai-Hamburger Y. 2010. Social network use and personality. *Computers in Human Behavior* 26(6) (2010), 1289–1295.

[5] Ian Ayres and Matthew Funk. 2003. Marketing privacy. *Yale J. on Reg.* 20 (2003), 77.

[6] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. 2012. The menlo report. *IEEE Security & Privacy* 10, 2 (2012), 71–75.

[7] BankMyCell 2022. *HOW MANY SMARTPHONES ARE IN THE WORLD?* BankMyCell. https://www.bankmycell.com/blog/how-many-phones-are-in-the-world.

[8] Tom L Beauchamp et al. 2008. The belmont report. *The Oxford textbook of clinical research ethics* (2008), 149–155.

[9] Volker Benndorf and Hans-Theo Normann. 2018. The Willingness to Sell Personal Data. *The Scandinavian Journal of Economics* 120, 4 (2018), 1260–1278. https://doi.org/10.1111/sjoe.12247 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjoe.12247

[10] Jaspreet Bhatia and Travis D Breaux. 2018. Empirical measurement of perceived privacy risk. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 6 (2018), 1–47.

[11] Ryan Calo. 2015. Privacy and markets: a love story. *Notre Dame L. Rev.* 91 (2015), 649.

[12] J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. D. Oliveira. 2013. Your browsing behavior for a Big Mac: Economics of Personal Information Online. In *the 22nd international conference.*

[13] Raul Castro Fernandez. 2023. Data-sharing markets: model, protocol, and algorithms to incentivize the formation of data-sharing consortia. *Proceedings of the ACM on Management of Data* 1, 2 (2023), 1–25.
[14] Delphine Christin, Andreas Reinhardt, Salil S. Kanhere, and Matthias Hollick. 2011. A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software* 84, 11 (2011), 1928–1946. https://doi.org/10.1016/j.jss.2011.06.073
[15] Janis Cloos, Björn Frank, Lukas Kampenhuber, Stephany Karam, Nhat Luong, Daniel Möller, Maria Monge-Larrain, Nguyen Tan Dat, Marco Nilgen, and Christoph Rössler. 2019. Is Your Privacy for Sale? An Experiment on the Willingness to Reveal Sensitive Information. *Games* 10, 3 (2019). https://doi.org/10.3390/g10030028
[16] Sunny Consolvo, Ian E. Smith, Tara Matthews, Anthony Lamarca, and Pauline Powledge. 2005. Location disclosure to social relations: Why, when, & what people want to share. In *Conference on Human Factors in Computing Systems*.
[17] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis. 2006. A Study on the Price of Location Privacy. *Proceedings of Acm Workshop on Privacy in the Electronic Society* (2006), 109–118.
[18] George Danezis, Stephen Lewis, and Ross Anderson. 2005. How much is location privacy worth. *Proceedings of the Workshop on the Economics of Information Security* (2005).
[19] EU Directive. 1995. 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the EC* 23, 6 (1995).
[20] Justin Drake, Karl Schulz, Radek Bukowski, and Kelly Gaither. 2021. Collecting and analyzing smartphone sensor data for health. In *Practice and Experience in Advanced Research Computing*. 1–4.
[21] Serge Egelman, Adrienne Porter Felt, and David Wagner. 2013. *Choice Architecture and Smartphone Privacy: There's a Price for That*. Springer Berlin Heidelberg, Berlin, Heidelberg, 211–236. https://doi.org/10.1007/978-3-642-39498-0_10
[22] Hany Fahmy. 2014. Modelling nonlinearities in commodity prices using smooth transition regression models with exogenous transition variables. *Statistical Methods & Applications* 23, 4 (2014), 577–600.
[23] Wei Fan. 2017. New Approaches of Personal Data Protection in A World of Big Data——Concept of Contextual and Risk-based Approaches inU S and EU's data protection regulatory reform. *Journal of Cyber and Information Law* (2017), 248–286,393–39.
[24] Ben Fehnert and Alessia Kosagowsky. 2008. Measuring user experience: complementing qualitative and quantitative assessment. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. 383–386.
[25] Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. 2015. AWARE: mobile context instrumentation framework. *Frontiers in ICT* 2 (2015), 6.
[26] Piers Fleming, Andrew P. Bayliss, S. Gareth Edwards, and Charles R. Seger. 2021. The role of personal data value, culture and self-construal in online privacy behaviour. *PLoS ONE* 16, 7 (2021), 1 – 20. https://search.ebscohost.com/login.aspx?direct=true&db=asn&AN=151434525&lang=zh-cn&site=ehost-live
[27] World Economic Forum. 2014. Rethinking Personal Data: Trust and Context in User-Centred Data Ecosystems. (2014).
[28] Euna Han and Lisa M. Powell. 2013. FAST FOOD PRICES AND ADULT BODY WEIGHT OUTCOMES: EVIDENCE BASED ON LONGITUDINAL QUANTILE REGRESSION MODELS. *Contemporary Economic Policy* (2013).
[29] Gabriella M. Harari, Nicholas D. Lane, Rui Wang, Benjamin S. Crosier, Andrew T. Campbell, and Samuel D. Gosling. 2016. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854. https://doi.org/10.1177/1745691616650285 arXiv:https://doi.org/10.1177/1745691616650285 PMID: 27899727.
[30] Nissenbaum Helen. 2004. Privacy As Contextual Integrity. *Washington Law Review* 79 (05 2004).
[31] S. Hermes, anela.halilovic@tum.de, M. Schreieck, and H. Krcmar. 2020. Who Quits Privacy-Invasive Online Platform Operators? A Segmentation Study with Implications for the Privacy Paradox. In *54th Hawaii International Conference on Systems Sciences*.
[32] J. K. Horowitz and K. E. Mcconnell. 2000. A Review of WTA/WTP Studies. 3 (2000).
[33] Simo Hosio, Denzil Ferreira, Jorge Goncalves, Niels Van Berkel, and Vassilis Kostakos. 2016. Monetary Assessment of Battery Life on Smartphones. *ACM* (2016).
[34] White House. 2012. Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. *White House, Washington, DC* (2012), 1–62.
[35] Ahmad Jamal. 2003. Marketing in a multicultural world: The interplay of marketing, ethnicity and consumption. *European journal of marketing* 37, 11/12 (2003), 1599–1620.
[36] C. Jensen, C. Potts, and C. Jensen. 2005. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human - Computer Studies* 63, 1-2 (2005), 203–227.
[37] Iris A Junglas, Norman A Johnson, and Christiane Spitzmüller. 2008. Personality traits and concern for privacy: an empirical study in the context of location-based services. *European Journal of Information Systems* 17(4) (2008), 387–402.
[38] L. H. Kai, H. H. Teo, and Syt Lee. 2007. The Value of Privacy Assurance: An Exploratory Field Experiment. *MIS Quarterly* 31, 1 (2007), 19–33.
[39] T Tony Ke and K Sudhir. 2023. Privacy rights and data security: GDPR and personal data markets. *Management Science* 69, 8 (2023), 4389–4412.

[40] Lindah Kotut, Timothy L. Stelter, Michael Horning, and D. Scott McCrickard. 2020. Willing Buyer, Willing Seller: Personal Data Trade as a Service. In *Companion of the 2020 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '20)*. Association for Computing Machinery, New York, NY, USA, 59–68. https://doi.org/10.1145/3323994.3369899
[41] Murat Koyuncu, Tolga Pusatli, et al. 2019. Security awareness level of smartphone users: An exploratory case study. *Mobile Information Systems* 2019 (2019).
[42] Frauke Kreuter, Georg-Christoph Haas, Florian Keusch, Sebastian Bähr, and Mark Trappmann. 2020. Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review* 38, 5 (2020), 533–549.
[43] Maria Kuecken. 2013. Book review: a tale of two cultures: qualitative and quantitative research in the social sciences. *lse review of books* (2013).
[44] Abhishek Kumar, Tristan Braud, Young D Kwon, and Pan Hui. 2020. Aquilis: Using contextual integrity for privacy protection on mobile devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–28.
[45] P. Kumaraguru and L. F. Cranor. 2005. Privacy Indexes: A Survey of Westin's Studies. (2005).
[46] Liron Lamash, Tsameret Ricon, and Sara Rosenblum. 2022. Time Organization Patterns of Adolescents: Agreement between Self-Report and Parent Report. *Physical & Occupational Therapy In Pediatrics* 42, 3 (2022), 319–332.
[47] Kenneth Laudon. 1995. *Markets and Privacy.* Academic Press, Inc., USA, 697–726.
[48] Hyunsoo Lee, Soowon Kang, and Uichin Lee. 2022. Understanding Privacy Risks and Perceived Benefits in Open Dataset Collection for Mobile Affective Computing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 61 (jul 2022), 26 pages. https://doi.org/10.1145/3534623
[49] Pedro Giovanni Leon, Blase Ur, Wang Yang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. 2013. What Matters to Users? Factors That Affect Users' Willingness to Share Information with Online Advertisers. In *Proceedings of the Ninth Symposium on Usable Privacy and Security* (Newcastle, United Kingdom) *(SOUPS '13)*. Association for Computing Machinery, New York, NY, USA, Article 7, 12 pages. https://doi.org/10.1145/2501604.2501611
[50] Xiao-Bai Li, Xiaoping Liu, and Luvai Motiwalla. 2021. Valuing Personal Data with Privacy Consideration. *Decision Sciences* 52, 2 (2021), 393–426. https://doi.org/10.1111/deci.12442 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/deci.12442
[51] John A List and Craig A Gallet. 2001. What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and resource economics* 20 (2001), 241–254.
[52] Jayson L Lusk, Corinne Alexander, and Matthew C Rousu. 2007. Designing experimental auctions for marketing research: The effect of values, distributions, and mechanisms on incentives for truthful bidding. *Review of Marketing Science* 5, 1 (2007), 0000102202154656161059.
[53] Aurel Ruben Mäder, Lakmal Meegahapola, and Daniel Gatica-Perez. 2024. Learning About Social Context from Smartphone Data: Generalization Across Countries and Daily Life Moments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
[54] N. K. Malhotra, S. S. Kim, and J. Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC). *Information Systems Research* (2004).
[55] Juha Matero and Ashley Colley. 2012. Identifying Unintentional Touches on Handheld Touch Screen Devices. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 506–509. https://doi.org/10.1145/2317956.2318031
[56] Lakmal Meegahapola, Florian Labhart, Thanh-Trung Phan, and Daniel Gatica-Perez. 2021. Examining the social context of alcohol drinking in young adults with smartphone sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–26.
[57] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I. Hong. 2014. Toss 'n' Turn: Smartphone as Sleep and Sleep Quality Detector. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 477–486. https://doi.org/10.1145/2556288.2557220
[58] Ngoc Thi Nguyen, Maria Zubair, Agustin Zuniga, Sasu Tarkoma, Pan Hui, Hyowon Lee, Simon Tangi Perrault, Mostafa H Ammar, Huber Flores, and Petteri Nurmi. 2023. Is the Price Right? The Economic Value of Sharing Sensors. *IEEE Transactions on Computational Social Systems* (2023).
[59] Jianmo Ni, Larry Muhlstein, and Julian McAuley. 2019. Modeling Heart Rate and Activity Data for Personalized Fitness Recommendation. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1343–1353. https://doi.org/10.1145/3308558.3313643
[60] H Nissenbaum. 2018. Respecting Context to Protect Privacy: Why Meaning Matters. *Sci Eng Ethics* 24 (2018), 831–852.
[61] Apthorpe Noah, Shvartzshnaider Yan, Mathur Arunesh, Reisman Dillon, and Feamster Nick. 2018. Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity.
[62] OCDE. 2002. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data / OCDE. (2002).
[63] Fabian Okeke, Michael Sobolev, and Deborah Estrin. 2018. Towards A Framework for Mobile Behavior Change Research *(TechMindSociety '18)*. Association for Computing Machinery, New York, NY, USA, Article 27, 6 pages. https://doi.org/10.1145/3183654.3183706

[64] Omer Rashid, Paul Coulton, and Reuben Edwards. 2008. Providing Location Based Information/Advertising for Existing Mobile Phone Users. *Personal Ubiquitous Comput.* 12, 1 (Jan. 2008), 3–10. https://doi.org/10.1007/s00779-006-0121-4

[65] Marc Rotenberg, Jeramie Scott, and Julia Horwitz. 2015. *Privacy in the modern age: The search for solutions*. New Press.

[66] Vera Schmitt, Zhenni Li, Maija Poikela, Robert P Spang, and Sebastian Möller. 2023. What is your location privacy worth? Monetary valuation of different location types and privacy influencing factors. In *Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. 19–29.

[67] Seven Wheat Data 2022. *White Paper on Mobile Internet Industry in 2021 - Seven Wheat Data*. Seven Wheat Data. https://www.sohu.com/a/518965750_120855974.

[68] Cornelia Sindermann, Helena Sophia Schmitt, Frank Kargl, Cornelia Herbert, and Christian Montag. 2021. Online privacy literacy and online privacy behavior–the role of crystallized intelligence and personality. *International Journal of Human–Computer Interaction* 37, 15 (2021), 1455–1466.

[69] Anya Skatova, Rebecca McDonald, Sinong Ma, and Carsten Maple. 2023. Unpacking privacy: Valuation of personal data protection. *Plos one* 18, 5 (2023), e0284581.

[70] Smith, H. Jeff, Milberg, J. Sandra, and Burke. 1996. Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly* (1996).

[71] Jacopo Staiano, Nuria Oliver, Bruno Lepri, Rodrigo De Oliveira, Michele Caraviello, and Nicu Sebe. 2014. Money walks: a human-centric study on the economics of personal mobile data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 583–594.

[72] Bella Struminskaya, Vera Toepoel, Peter Lugtig, Marieke Haan, Annemieke Luiten, and Barry Schouten. 2020. Understanding willingness to share smartphone-sensor data. *Public Opinion Quarterly* 84, 3 (2020), 725–759.

[73] Joshua Tan, Mahmood Sharif, Sruti Bhagavatula, Matthias Beckerle, Michelle L Mazurek, and Lujo Bauer. 2018. Comparing hypothetical and realistic privacy valuations. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. 168–182.

[74] Tsai, S. Egelman, L. F. Cranor, and A. Acquisti. 2011. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research* (2011), 254–268. https://ssrn.com/abstract=3305347

[75] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science* 185, 4157 (1974), 1124–1131.

[76] Michele Vescovi, Christos Perentis, Chiara Leonardi, Bruno Lepri, and Corrado Moiso. 2014. My Data Store: Toward User Awareness and Control on Personal Data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 179–182. https://doi.org/10.1145/2638728.2638745

[77] A. Wagner, N. Wessels, P. Buxmann, and H. Krasnova. 2018. Putting a Price Tag on Personal Information - A Literature Review. *publications of darmstadt technical university institute for business studies* (2018).

[78] Mengcheng Wang, Xiaoyang Dai, and Shuqiao Yao. 2011. Preliminary development of the Chinese Big Five Personality Inventory III: Development of the short form version and reliability testing. *Chinese Journal of Clinical Psychology* 19, 4 (2011), 4.

[79] Zhixian Yan, Jun Yang, and Emmanuel Munguia Tapia. 2013. Smartphone Bluetooth Based Social Sensing. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (Zurich, Switzerland) *(UbiComp '13 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 95–98. https://doi.org/10.1145/2494091.2494118

[80] Razieh Nokhbeh Zaeem and K Suzanne Barber. 2020. The effect of the GDPR on privacy policies: Recent progress and future promise. *ACM Transactions on Management Information Systems (TMIS)* 12, 1 (2020), 1–20.

[81] Bo Zhang, Yi Ming Li, Jian Li, Jing Luo, and Oliver P John. 2021. The Big Five Inventory-2 in China: A Comprehensive Psychometric Evaluation in Four Diverse Samples. *Assessment* 4 (2021).

[82] Sicheng Zhao, Hongxun Yao, Xiaoshuai Sun, Pengfei Xu, Xianming Liu, and Rongrong Ji. 2011. Video Indexing and Recommendation Based on Affective Analysis of Viewers. In *Proceedings of the 19th ACM International Conference on Multimedia* (Scottsdale, Arizona, USA) *(MM '11)*. Association for Computing Machinery, New York, NY, USA, 1473–1476. https://doi.org/10.1145/2072298.2072043

[83] Xiao Zheng. 2020. Data trading with differential privacy in data market. In *Proceedings of 2020 6th International Conference on Computing and Data Engineering*. 112–115.

[84] Pengfei Zhou, Mo Li, and Guobin Shen. 2014. Use It Free: Instantly Knowing Your Phone Attitude. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking* (Maui, Hawaii, USA) *(MobiCom '14)*. Association for Computing Machinery, New York, NY, USA, 605–616. https://doi.org/10.1145/2639108.2639110

A APPENDICES

A.1 Disclose of Potential Risks in User Study

We disclosed the potential risk as follows in the user consent. The original language was in Chinese.

The data we collect includes not only the physiological and psychological data, behavioral data, and environmental data written in the product names within our application "My Data." It also includes other data collected once permissions related to these products are activated:

(1) The data collected when the "Call Logs" permission is activated reveals not only with whom you communicate most frequently but also exposes your social relationships.
(2) The data collected with the "Usage Access" permission exposes your shopping preferences and times, as well as your usage of other applications.
(3) The data from the "Front Camera" permission reveals your facial features and can be used for identity recognition.
(4) The data from the "Rear Camera" permission exposes not only your environment but also your social and home environments.
(5) The data collected when the "Microphone" permission is activated exposes your mental health as well as the content of your conversations.

We will state in the "Privacy Policy" on the application's introduction and registration pages that all collected data will be anonymized, ensuring your identity information is not disclosed. It will be collected, used, and stored by the laboratory and destroyed after one month, without being transferred to any third party. Although this data collection may feel like a privacy invasion, potentially causing psychological impacts, we assure you it will not cause physical, psychological, social adaptation, or any other kind of harm.

In fact, we do not use the collected data for any purpose. Only the project leader has access to download and store data from the server. The data will be encrypted, and only the project leader will have access to it in later stages.

A.2 Detail Considerations of the Data Chosen

We collected 9 types of sensor data, including accelerometer, front (rear) camera, microphone, touch screen, gyroscope, barometer, compass, temperature, light. Sensors on smartphones can sense the behavior and psychology of individuals. The data captured by the accelerometer was used [63]to record the user's travel patterns (e.g., walking, running, driving). The front camera was used to capture the user's facial features, which could also be used as authentication [14] or facial expression recognition [82]. The photo and video data from the rear camera could reflect the user's hobbies, diet, environment [14] and socialization [29]. The microphone recorded sound could be used not only for noise measurement of the surroundings [14], but also for recognizing the user's emotions and conversations [29], and for inferring social interactions [14, 29]. Touch screens [55] collected users' click data to study user mis-touch on capacitive touch screens. Gyroscopes could be used for posture estimation [84].

In addition to sensing the behavior and psychology of individuals, sensors on smartphones could also sense environmental information [29]. Data from barometer, compass, and temperature sensors could be used to calculate the height, direction, and temperature of the user's environment, respectively. Data from light sensors[29] could be used to monitor the brightness of the environment and to infer the user's daily activities. These environmental data could be used alone or in combination with other sensors to infer the user's behavior and activities [29].

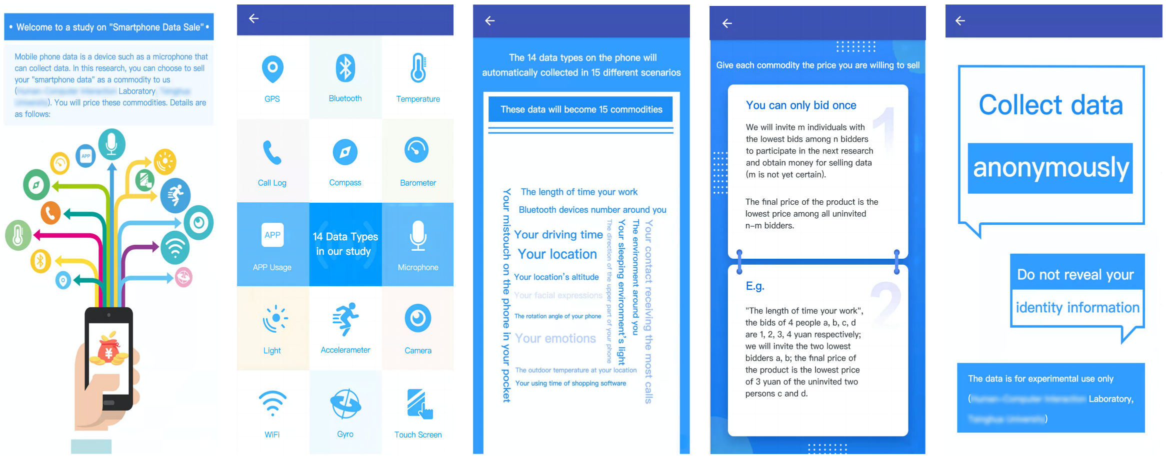
We collected a total of 5 types of non-sensor data, including bluetooth, location, log data [29]-SMS logs, call logs, application usage logs, and identity data. Bluetooth could be used for social and context-awareness [79]. The two main methods of location were GPS and WiFi. GPS and WiFi [14], [63] could collect user's trip, home address, workplace, hospital, etc. through location information. SMS logs, call logs [63] could be used to infer the user's daily activities (work, home, classroom). application usage logs could reflect [29] users' social interactions.

Table 5. The attribute values of all the 49 commodities. (1-25)

No.	Commodity Name	Data Type	Permission	Purpose	Privacy Risk	Description
1	Your movement path	GPS	Yes	Physical and Psychological research	4	Longitude and latitude measured by GPS of phone
2	Your working hour	GPS	Yes	Physical and Psychological research	3	Longitude and latitude measured by GPS of phone
3	The length of time you stay at home	GPS	Yes	Physical and Psychological research	3	Longitude and latitude measured by GPS of phone
4	The length of time you spend in the park	GPS	Yes	Physical and Psychological research	3	Longitude and latitude measured by GPS of phone
5	Your location	GPS	Yes	Physical and Psychological research	4	Longitude and latitude measured by GPS of phone
6	The length of time you stay in the hospital	GPS	Yes	Physical and Psychological research	4	Longitude and latitude measured by GPS of phone
7	Your wireless headset usage	Bluetooth	Yes	Personalized advertisement	3	The number and time of Bluetooth connected devices
8	Your connection to other phones	Bluetooth	Yes	Personalized advertisement	4	The number and time of Bluetooth connected devices
9	Your use of shared bikes	Bluetooth	Yes	Personalized advertisement	3	The number and time of Bluetooth connected devices
10	The number of Bluetooth devices around you	Bluetooth	Yes	Physical and Psychological research	3	The number and time of Bluetooth connected devices
11	Your connection to the wireless speaker	Bluetooth	Yes	Personalized advertisement	3	The number and time of Bluetooth connected devices
12	The outdoor temperature of your location	Temperature	No	Fitness recommendations	2	Outdoor temperature measured by phone
13	Your body temperature	Temperature	No	Fitness recommendations	3	Your temperature measured by phone
14	The indoor temperature of your location	Temperature	No	Fitness recommendations	2	Indoor temperature measured by phone
15	How many contacts did you call	Call records	Yes	Physical and Psychological research	4	Number and duration of calls
16	How many contacts did you receive calls from	Call records	Yes	Physical and Psychological research	4	Number and duration of calls
17	Which contact you call the most	Call records	Yes	Physical and Psychological research	5	Number and duration of calls
18	Which contact you receive the most calls	Call records	Yes	Physical and Psychological research	5	Number and duration of calls
19	The horizontal direction of your phone	Compass	No	Physical and Psychological research	1	The direction measured by the compass in the phone
20	The altitude at your location	Barometer	No	Physical and Psychological research	1	The altitude of the mobile phone location measured by the barometer
21	Your facial expression	Camera	Yes	Physical and Psychological research	4	Facial photos taken by the front camera
22	Your surroundings	Camera	Yes	Physical and Psychological research	4	Environmental video taken by the rear camera
23	The length of time you use music software	app Usage	Yes	Personalized advertisement	3	App usage time recorded by app log
24	The length of time you use WeChat	app Usage	Yes	Personalized advertisement	3	App usage time recorded by app log
25	The length of time you play games	app Usage	Yes	Personalized advertisement	3	App usage time recorded by app log

Table 6. The attribute values of all the 49 commodities. (26-49)

No.	Commodity Name	Data Type	Permission	Purpose	Privacy Risk	Description
26	The length of time you spend on online shopping	App usage	Yes	Personalized advertisement	5	App usage time recorded by app log
27	The length of time you use your cell phone	App usage	Yes	Physical and Psychological research	2	App usage time recorded by app log
28	Your attention span	App usage	Yes	Physical and Psychological research	3	App usage time recorded by app log
29	The length of time you use your phone	Accelerometer	No	Fitness recommendations	2	The phone's movement degree
30	The number of steps you walk	Accelerometer	No	Fitness recommendations	2	The phone's movement degree
31	The Length of time you run	Accelerometer	No	Fitness recommendations	2	The phone's movement degree
32	The length of time you drive	Accelerometer	No	Fitness recommendations	3	The phone's movement degree
33	Your sleeping environment's light	Light	No	Physical and Psychological research	2	The brightness of the light in your environment
34	The light of your phone use indoors	Light	No	Physical and Psychological research	2	The brightness of the light on the front of the phone
35	The light of your phone use outdoors	Light	No	Physical and Psychological research	2	The brightness of the light on the front of the phone
36	The noise of your environment	Microphone	Yes	Physical and Psychological research	4	Sound around the phone
37	How many people you talked to	Microphone	Yes	Physical and Psychological research	5	Sound around the phone
38	The length of time you talk	Microphone	Yes	Physical and Psychological research	4	Your sound
39	Your emotion	Microphone	Yes	Physical and Psychological research	5	Your sound
40	The length of time you watched TV	Microphone	Yes	Physical and Psychological research	3	Sound around the phone
41	Your movement path	WiFi	No	Fitness recommendations	4	Number and time of mobile WiFi connections
42	Your working hour	WiFi	No	Fitness recommendations	3	Number and time of mobile WiFi connections
43	The length of time you stay at home	WiFi	No	Fitness recommendations	3	Number and time of mobile WiFi connections
44	The length of time you spend in the park	WiFi	No	Fitness recommendations	3	Number and time of mobile WiFi connections
45	Your location	WiFi	No	Fitness recommendations	4	Number and time of mobile WiFi connections
46	The spatial orientation of your phone	Gyro	No	Improve user Experience	2	Rotation and tilt angle of the phone
47	The tilt of your phone	Gyro	No	Improve user Experience	2	Rotation and tilt angle of the phone
48	Your unintentional touches when the phone was placed in the pocket	Touch screen	No	Improve user Experience	1	Actions on the touch screen
49	Your single-touch habits	Touch screen	No	Improve user Experience	2	Actions on the touch screen



(a) (b) (c) (d) (e)

Fig. 9. The user interface of the experiment platform. The original page is in Chinese and we translated that to English. (a) The goal of the study. (b) Selected 14 data types. (c) The overview of the commodities. (d) The procedure of reverse second price auction. (e) The privacy policy of the data.

Table 7. Demographics of student participants

Demographic Information	Count	Percentage
Gender		
Female	112	61.9%
Male	69	38.1%
Ethnic Group		
Han	113	62.5%
Hui	29	16.0%
Tibetan	24	13.3%
Tu	10	5.5%
Salar	3	1.7%
Mongolian	2	1.0%
Living expenses per month (in CNY)		
0-500	22	12.2%
500-1000	65	35.9%
1000-1500	65	35.9%
1500-2000	22	12.2%
2000-2500	2	1.0%
2500-3000	5	2.8%

Table 8. Demographics of non-student participants

Demographic Information	Count	Percentage
Gender		
Female	20	58.8%
Male	14	41.2%
Ethnic Group		
Ethnic minorities	3	8.8%
Han Chinese	31	91.2%
Marital Status		
Married	28	82.3%
Never married	6	17.7%
Divorced	0	0%
Have Children		
Yes	28	82.3%
No	6	17.7%
A career related to IT (Information Technology)		
Yes	14	41.2%
No	20	58.8%
Occupation		
E-commerce	1	2.9%
Consultants/consulting	1	2.9%
Technical or R&D	2	5.9%
Teachers	10	29.4%
Drivers	1	2.9%
Professionals	2	5.9%
Freelancers	7	20.6%
Others	10	29.4%
Education		
Junior high school	1	2.9%
High school	5	14.7%
Undergraduate	20	59.0%
Master's degree	8	23.4%
Income per month (in CNY)		
0-1000	1	2.9%
1000-2000	1	2.9%
2000-3000	4	11.8%
3000-4000	5	14.7%
4000-5000	8	23.4%
5000-6000	2	6.0%
6000-7000	3	8.9%
7000-8000	4	11.8%
8000-9000	1	2.9%
9000-10000	4	11.8%
>10000	1	2.9%

Table 9. Distribution of the length of time (in hours) that the participants spent using smartphones.

	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	>12	Sum
Student	0	1.7%	11.2%	14.6%	15.7%	9.6%	9.0%	6.2%	6.7%	10.7%	4.5%	2.2%	7.9%	100%
Non-student	3.1%	6.3%	25.0%	21.9%	25.0%	15.6%	0	0	0	3.1%	0	0	0	100%

Table 10. The ratings of the participants for the different questions on their perceptions of privacy on smartphones. (7: totally agree; 1: totally disagree).

	Student		Non-student	
Question	Mean	SD	Mean	SD
1. I am concerned about the protection of the data collected on my smartphone.	5.3	1.2	5.8	1.3
2. I trust the apps I install and run on my smartphone and how they use my data.	4.2	1.4	4.4	1.6
3. I trust how telecom providers (Mobile, Unicom, Telecom) use my data.	3.8	1.5	3.6	1.6
4. I always read the privacy terms and conditions of the applications I use.	3.3	1.8	3.8	1.7
5. I am aware of the legislation on mobile data protection.	3.1	1.6	4.2	1.9

Table 11. Statistics of prices of 15 commodities given by students

Commodity	Min.	1st Q	Med.	Mean	3rd Q	Max.
Your working hour	0.1	5.0	10.0	62.3	35.0	700.0
The number of Bluetooth devices around you	0.01	5.0	10.0	50.2	30.0	700.0
The outdoor temperature at your location	0.01	3.25	10	26.6	25.0	700.0
Which contact you receive the most calls	0.1	7.0	20	107.1	100.0	1000.0
The horizontal direction of your phone	0.1	4.0	10.0	36.4	25.0	1000.0
The altitude at your location	0.1	3.0	10.0	25.8	25.0	1000.0
The length of time you spend on online shopping	0.01	9.0	25.0	91.6	80.0	1000.0
Your facial expression	0.1	10.0	20.0	127.8	100.0	1000.0
Your surroundings	0.1	10.0	20.0	99.5	80.0	700.0
The length of time you drive	0.1	5.0	10.0	43.1	37.5	1000.0
Your sleeping environment's light	0.1	5.0	10.0	34.6	34	1000.0
Your emotion	0.5	7.0	20.0	104.8	66.0	1000.0
Your location	0.5	8.0	15.0	77.5	50.0	1000.0
The spatial orientation of your phone	0.1	5.0	10.0	28.9	20.0	509.0
Your unintentional touches when the phone was placed in the pocket	0.09	5.0	9.0	26.0	20.0	1000.0

Table 12. Statistics of prices of 15 commodities given by non-students

Commodity	Min.	1st Q	Med.	Mean	3rd Q	Max.
Your working hour	3.0	20.0	50.0	114.7	100.0	700.0
The number of Bluetooth devices around you	2.0	10.0	35.0	94.8	70.0	700.0
The outdoor temperature at your location	1.0	10.0	40.0	77.1	100.0	600.0
Which contact you receive the most calls	2.0	10.0	55.0	82.2	100.0	500.0
The horizontal direction of your phone	1.0	10.0	30.0	87.8	100.0	700.0
The altitude at your location	1.0	10.0	40.0	85.5	80.0	700.0
The length of time you spend on online shopping	4.0	20.0	55.0	187.9	350.0	1000.0
Your facial expression	2.0	30.0	56.5	263.1	500.0	1000.0
Your surroundings	2.0	17.0	50.0	94.1	90.0	600.0
The length of time you drive	2.0	14.0	55.0	98.3	100.0	500.0
Your sleeping environment's light	2.0	18.0	30.0	50.9	80.0	200.0
Your emotion	4.0	20.0	50.0	141.8	100.0	1000.0
Your location	2.0	24.0	55.0	137.7	100.0	1000.0
The spatial orientation of your phone	1.0	10.0	30.0	45.6	80.0	140.0
Your unintentional touches when the phone was placed in the pocket	0.001	10.0	25.0	85.9	88.0	1000.0

Table 13. Subject's informed consent 1

Research Topic	Evaluating the Valuation of Personal Data on Smartphones
Researcher	*Sensitive Concealment*
Project Contacts and their Addresses	*Sensitive Concealment*
Purpose of the Study	

(1) The significance of studying the monetary value of data on smartphones, whether in the fields of economics, decision science, management, or human-computer interaction, is that it can inform future data markets.
(2) Our research can help users better understand personal data and its privacy on smartphones.
(3) Our research can also improve the transparency of information and help users evaluate their data accurately and truthfully.

Experimental Procedures

We will create a WeChat group at the beginning of the experiment for all subjects participating in the experiment to join. In this WeChat group, we will notify all subjects about the experiment. We will post the following text to that WeChat group.
Cell phone sensors are devices such as microphones that collect data. In this study, you have the option of selling your ‘cell phone sensor data” to us as a commodity. You will first price these goods. The steps are as follows:

(1) Fill out the basic information questionnaire.
(2) Please download and install the application ‘My Data”.
(3) Please register the application with your cell phone number and agree to the privacy policy on the registration page before registration.
(4) Our experiment includes two parts: the first part is the bidding of 15 commodities, which requires you to price 15 commodities; the second part is the sale of 15 commodities, if your bidding is successful (it will be shown in app), you can sell the commodities.
(5) The bidding process is as follows: please check the descriptions of the 15 items carefully, and if you are willing to sell, please give an estimate of the price of the items you are willing to sell; if you are not willing to sell, you may not give an estimate. You may sell as many commodities as you wish. You can only estimate the price of each commodity once, and you can't change the estimate afterward. To ensure that the price is correct, you may write it out on paper first. This process is completed within 24 hours.
(6) After the bidding is completed, we will notify you to log in the app in the WeChat group, and the final price of each commodity will be displayed in the app interface. If the interface where you estimate the commodity shows ‘can be sold”, it means that your bidding is successful, and you will be able to continue to sell the data in the app and get the revenue; if the commodity you estimate shows ‘can not be sold”, it means that your bidding is unsuccessful, and you will not be able to sell it.
(7) After your commodity has been sold for 7 days, please click ‘Receive Revenue” in the Revenue” of the app interface. After receiving the proceeds, we will give you cash through the transfer function of WeChat.
(8) After receiving the proceeds, please fill out 1 questionnaire. The entire experiment will not be videotaped and you can exit at any stage of the experiment.

Table 14. Subject's informed consent 2

Potential Risks and Side Effects
The data we collect includes physical and psychological data , behavioral data and data about the environment in which you are located on your smartphone. This data can make you feel that your privacy is being exposed and can cause potential psychological effects. However, it will not harm you physically, psychologically, socially adapted or otherwise. We state in the "Privacy Policy" on the app's navigation and registration pages that all data we collect is anonymous, does not reveal your identity, is collected, used and stored by the laboratory and destroyed after one month, and is not transferred to third parties.
Benefits
Although your participation in this study does not provide any other direct benefit to you personally, there are possible benefits that will result from the research in this program. Our apps can help you understand the information about personal data on your smartphone, including the type of data, the specific data collected, the psychological, behavioral and environmental information that can be calculated about an individual, and can also help you understand the privacy risks of personal data on your smartphone. The information we provide about the personal data on your smartphone can also help you to better value the commodities and decide whether to sell your personal data. If you are successful in selling your data, you may also receive monetary compensation. This study can also contribute to research on the buying and selling of personal data and inform the future data market, making it more transparent for buyers to purchase personal data. As long as you participate in the experiment, even if you don't price any commodities or don't sell a single commodity, we'll send you a small gift at the end of the experiment.
Privacy
The results of this study may be published in academic journals/books or used for teaching purposes. However, your name or other information that identifies you will not appear in any published or instructional materials unless you give your permission.
Termination of the experiment
Your participation is entirely on a voluntary basis; you may request to withdraw at any point during the experiment, and you will not be penalized or lose money for withdrawing from the experiment.
Principal Subject Statement
I have explained the purpose of the study, the procedures of the study, the potential dangers and discomforts, and the rights and interests of the subjects, and have answered questions related to the study to the best of my ability. Signature: Date:
Subject
I declare that I have been informed of the purpose, procedure, possible dangers, and potential benefits and costs of this study. All my questions have been answered to my satisfaction. I have read this subject consent form in detail. My signature below indicates my willingness to participate in this study. Signature: Date: