NBER WORKING PAPER SERIES

THE DIGITAL WELFARE OF NATIONS:
NEW MEASURES OF WELFARE GAINS AND INEQUALITY

Erik Brynjolfsson
Avinash Collis
Asad Liaqat
Daley Kutzman
Haritz Garro
Daniel Deisenroth
Nils Wernerfelt
Jae Joon Lee

Working Paper 31670
http://www.nber.org/papers/w31670

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2023

A.L., D.K., H.G., and D.D. are employees of Meta Platforms and hold a financial interest in Meta. N.W. was an employee of Meta while this research was conducted though he no longer holds a financial interest in the company. All other authors declare no current competing interests, although E.B. and A.C. have in the past been awarded unrestricted gifts from Meta and receive final support from the Stanford Digital Economy Lab which is in turn partially funded by a variety of entities. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w31670

The Digital Welfare of Nations: New Measures of Welfare Gains and Inequality
Erik Brynjolfsson, Avinash Collis, Asad Liaqat, Daley Kutzman, Haritz Garro, Daniel Deisenroth, Nils Wernerfelt, and Jae Joon Lee
NBER Working Paper No. 31670
September 2023
JEL No. O30,O4

ABSTRACT

Digital goods can generate large benefits for consumers, but these benefits are largely unmeasured in the national accounts, including GDP and productivity. In this paper, we measure welfare gains from 10 popular digital goods across 13 countries by conducting large-scale incentivized online choice experiments on representative samples of nearly 40,000 people. We estimate that these goods—many of which are free to users—generate over $2.5 trillion in aggregate consumer welfare across these countries per year, which is roughly equivalent to 6% of their combined GDP. We find that lower-income individuals and lower-income countries obtain relatively larger welfare gains from these goods compared to higher-income individuals and countries. This suggests that digital goods may reduce inequality in welfare within and across countries by disproportionately benefiting lower-income groups.

Erik Brynjolfsson
Stanford Digital Economy Laboratory
353 Jane Stanford Way, Office 136
Stanford, CA 94305
and NBER
erik.brynjolfsson@gmail.com

Avinash Collis
Carnegie Mellon University
Pittsburgh, PA
avinash.collis@gmail.com

Asad Liaqat
Meta Platforms
1 Hacker Way
Menlo Park, CA 94063
asadliaqat@meta.com

Daley Kutzman
Meta
1 Hacker Way
Menlo Park, CA 94025
dkutzman@meta.com

Haritz Garro
Meta
1 Hacker Way
Menlo Park, CA 94025
haritz.garro@gmail.com

Daniel Deisenroth
Meta Platforms
1 Hacker Way
Menlo Park, CA 94025
dbdeisenroth@meta.com

Nils Wernerfelt
Kellogg School of Management
2211 Campus Drive
Evanston, IL 60208
nils.wernerfelt@gmail.com

Jae Joon Lee
Stanford Digital Economy Lab
353 Jane Stanford Way
Stanford, CA 94305
jaejoonl@stanford.edu

The Digital Welfare of Nations: New Measures of Welfare Gains and Inequality

Erik Brynjolfsson✉
Stanford University
and NBER

Avinash Collis✉
Carnegie Mellon
University

Asad Liaqat
Meta

Daley Kutzman
Meta

Haritz Garro
Meta

Daniel Deisenroth
Meta

Nils Wernerfelt
Northwestern
University

Jae Joon Lee*
Stanford University

September 2023

Abstract

Digital goods can generate large benefits for consumers, but these benefits are largely unmeasured in the national accounts, including GDP and productivity. In this paper, we measure welfare gains from 10 popular digital goods across 13 countries by conducting large-scale incentivized online choice experiments on representative samples of nearly 40,000 people. We estimate that these goods—many of which are free to users—generate over $2.5 trillion in aggregate consumer welfare across these countries per year, which is roughly equivalent to 6% of their combined GDP. We find that lower-income individuals and lower-income countries obtain relatively larger welfare gains from these goods compared to higher-income individuals and countries. This suggests that digital goods may reduce inequality in welfare within and across countries by disproportionately benefiting lower-income groups.

Introduction

Digital technologies create challenges for economic measurement. On one hand, with the spread of the Internet, time spent on digital goods has increased dramatically and these goods are affecting more and more aspects of daily life. For instance, the average person in both the US and UK now spends almost 24 hours a week online (Coyle and Nakamura, 2022) while the

*Authors listed in order of contribution with E.B. and A.C. contributing equally. A.L. and D.K. led the project internally at Meta. A.L., D.K., and H.G designed and implemented the survey, and analyzed and validated the data. A.L. and A.C. wrote the paper, with contributions from E.B., D.K., H.G., D.D., N.W.; E. B. and A. C. were the joint external principal investigators; D. D. and N. W. were the joint internal principal investigators.
Competing interests: A.L., D.K., H.G., and D.D. are employees of Meta Platforms and hold a financial interest in Meta. N.W. was an employee of Meta while this research was conducted though he no longer holds a financial interest in the company. All other authors declare no current competing interests, although E.B. and A.C. have in the past been awarded unrestricted gifts from Meta.

✉Corresponding authors. Email: erikb@stanford.edu; avinashcollis@cmu.edu.

number of photos shared increased from 350 billion to 2.5 trillion between 2010 and 2017 (Brynjolfsson et al., 2019b). On the other hand, the officially-measured size of the information sector as a share of total GDP has remained almost unchanged at around 4-5% for the past four decades. The discrepancy is a reflection of the fact that, regardless of the value they create for consumers, most digital goods are available at zero price (Brynjolfsson and Collis, 2019).[1]

To better understand the welfare effects of digital goods, new metrics are needed to complement existing production-based metrics such as GDP and productivity (Masood, 2022).[2] Recent research has proposed new ways of directly measuring consumer surplus from digital goods using massive online choice experiments (Brynjolfsson et al., 2019a; Brynjolfsson et al., 2019b). However, existing research in this area has only looked at a select few regions and goods using convenience samples that were not representative of the population of digital users.

In this paper, we conduct large-scale incentivized choice experiments involving nearly 40,000 representative users of the Facebook digital service in 13 countries to estimate the welfare gains generated by 10 popular digital goods. While previous research looks at specific digital goods in a single country using smaller non-representative samples (Allcott et al., 2020; Brynjolfsson et al., 2019a), the scale and scope of our sample enable us to estimate valuations for ten leading digital goods with sufficient precision to compare welfare gains from digitization across countries. Similar contingent valuation methods have been used in the past to value non-market goods, including environmental goods (Bishop et al., 2017) and accepted as evidence in legal cases (Lowensohn, 2012).

We find that digital goods generate substantial welfare for consumers across these countries. Specifically, our analysis implies that digital goods create $2.52 trillion of value across all 13 countries, which corresponds to 5.95% of their aggregate GDP. We also find that lower-income individuals within countries and lower-GDP countries obtain disproportionately more welfare from these digital goods compared to higher-income individuals and higher-GDP countries—not only relative to income, but in some instances even in absolute terms. Because the free digital goods we examine (e.g. search engines or instant messaging platforms) are available for free to both higher- and lower-income individuals, they serve to reduce welfare inequality both within and across countries in our sample.

Since none of the digital goods we examine existed a few decades ago, our findings suggest that economic growth may have been underestimated by conventionally-measured GDP.[3]

---

[1] Most are instead supported by advertising revenues, while others are supported by volunteers.
[2] Production based metrics may also not properly include welfare gains from non-digital goods. For e.g., Nordhaus (1996) calculates that lighting generated $275 billion in consumer surplus between 1800 and 1992, and this is not measured in existing metrics. Jones and Klenow (2016) propose a more comprehensive measure of welfare accounting for consumption, leisure, mortality, and inequality. In this work we focus on digital goods because the changes in welfare over the past few decades may be the highest for digital goods compared to other types of goods.
[3] As discussed in Brynjolfsson et al (2019b), the fact that some of these goods are paid for by advertising or other means does not fundamentally change this fact. That said, it is also possible, even likely, that

Furthermore, since labor productivity is typically defined as GDP per hour worked, it also fails to reflect the full contribution of digital goods. As the digital economy becomes relatively more important, it will be increasingly important to explicitly measure the trillions of dollars of value created by these sorts of goods. The reduction in inequality is consistent with the idea that greater access to free public goods can be effective in reducing inequality (Fischer, 2017) and that relative price changes have an important effect on inequality (Slottje, 1987). In particular, as argued by Goolsbee and Klenow (2006), low wage workers should be expected to consume relatively higher quantities of free digital goods, not only because they have less money to spend on other goods, but also because their opportunity cost of time may be lower.

Results

Measuring welfare gains from digital goods

We surveyed 39,717 Facebook users in 13 countries on the Facebook internal survey platform (see Materials and Methods section in the Supplementary Materials for more details). The two main components of the survey were (i) a best-worst scaling task (N=23,752) where users select their most and least preferred options from a list (Louviere et al., 2015), and (ii) an incentivized willingness-to-accept estimate (N=39,717) using single-binary discrete choice experiments (Brynjolfsson et al., 2019a). The former gives us relative valuations of different goods while the latter gives us valuation for our benchmark good using a revealed preference approach.[4] In this paper, we use Facebook's valuation as the benchmark to calibrate valuations of the other goods.[5]

The survey sample for each component was weighted to be representative of the population of "monthly active" Facebook users in each country (i.e., users who have been active on Facebook within the last 30 days). Facebook has nearly 3 billion monthly active users, and using a sample that is representative of Facebook users in each country is a stark improvement over existing

economic growth was misestimated in earlier eras as well, as other unmeasured or poorly measured goods and services entered (and exited) the economy. Our paper is only one in a long line of efforts to improve on past measures.

[4] These are well-established methods for valuing non-market goods. Note that there are other valuation methods available, including, for example, Becker-Degroot-Marschak's willingness-to-pay approach ("BDM" - Becker et al., 1964). Prior to running our main survey we tested several such methods and found that our preferred approach worked best.

[5] These valuations capture the consumer surplus generated by these goods since they measure consumers' valuation of a good beyond the costs associated with consuming the good. In most instances, the costs associated with consuming digital goods are very small or nonexistent as most users have a phone with an internet plan. However, some users might purchase a phone with an internet plan with the main purpose of being able to use Facebook or a similar app. In these cases, the purchase of the phone and internet plan will be reflected in the GDP numbers, and our measure of consumer surplus will be an underestimate.

estimates that were based on laboratory experiments or off-platform surveys of the general public.[6]

(a) Measuring relative valuations of digital goods using best-worst scaling

We used a best-worst scaling methodology to measure the relative willingness to accept for stopping the use of 10 selected digital goods,[7] as well as not "meeting friends in person", for one month. We included the category "meeting friends in person" to compare the relative valuation of in-person communication versus online social networks. Users are provided with a list of items and asked to select their most valuable and least valuable option from that list (i.e. relative comparison of these items, see Appendix Figure 2 for an example task). The digital goods include Facebook, Twitter, Instagram, WhatsApp, Snapchat, TikTok, Google Search, Google Maps, YouTube, and Amazon Shopping. Using a balanced-incomplete block design to ensure that all pairs of goods are evaluated together sufficiently (Hanani, 1961), we generated 70 questions with the 11 items mentioned above plus 10 monetary amounts. We dropped 7 questions that included *only* monetary amounts, resulting in 63 questions from which we presented three random questions to each respondent.

Figure 1 shows the ranking of items from most preferred to least preferred based on the disutility of giving up access to that item (the omitted category is Snapchat).[8] Google Search is the most preferred good and is even preferred to meeting friends in person. YouTube is also highly ranked, along with Google Maps, WhatsApp, Amazon Shopping, and Facebook. Other apps such as TikTok, Instagram, Twitter, and Snapchat—which is the least preferred digital service and thus the base category—receive the least broad-based appeal among the suite of digital goods.[9]

---

[6] The monthly active user (MAU) population in Facebook in these 13 countries constitutes a large share of the total population. The average Facebook penetration rate (defined as MAU divided by population) among these countries is 59.4%. These users are likely to be fairly representative of the users of digital technologies in each of these countries, though not necessarily of people who don't use any digital goods.

[7] These included three digital goods provided by Meta Platforms, Inc. (Facebook, Whatsapp, Instagram) as well as several other popular apps covering a range of product types.

[8] This approach also allows us to analyze heterogeneity in valuations based on various user characteristics. Appendix Figure 15 analyses relative utility of these goods based on gender.

[9] Some of the digital goods in the list, such as Snapchat, are not broadly used in many countries, which contributes to them being least preferred.

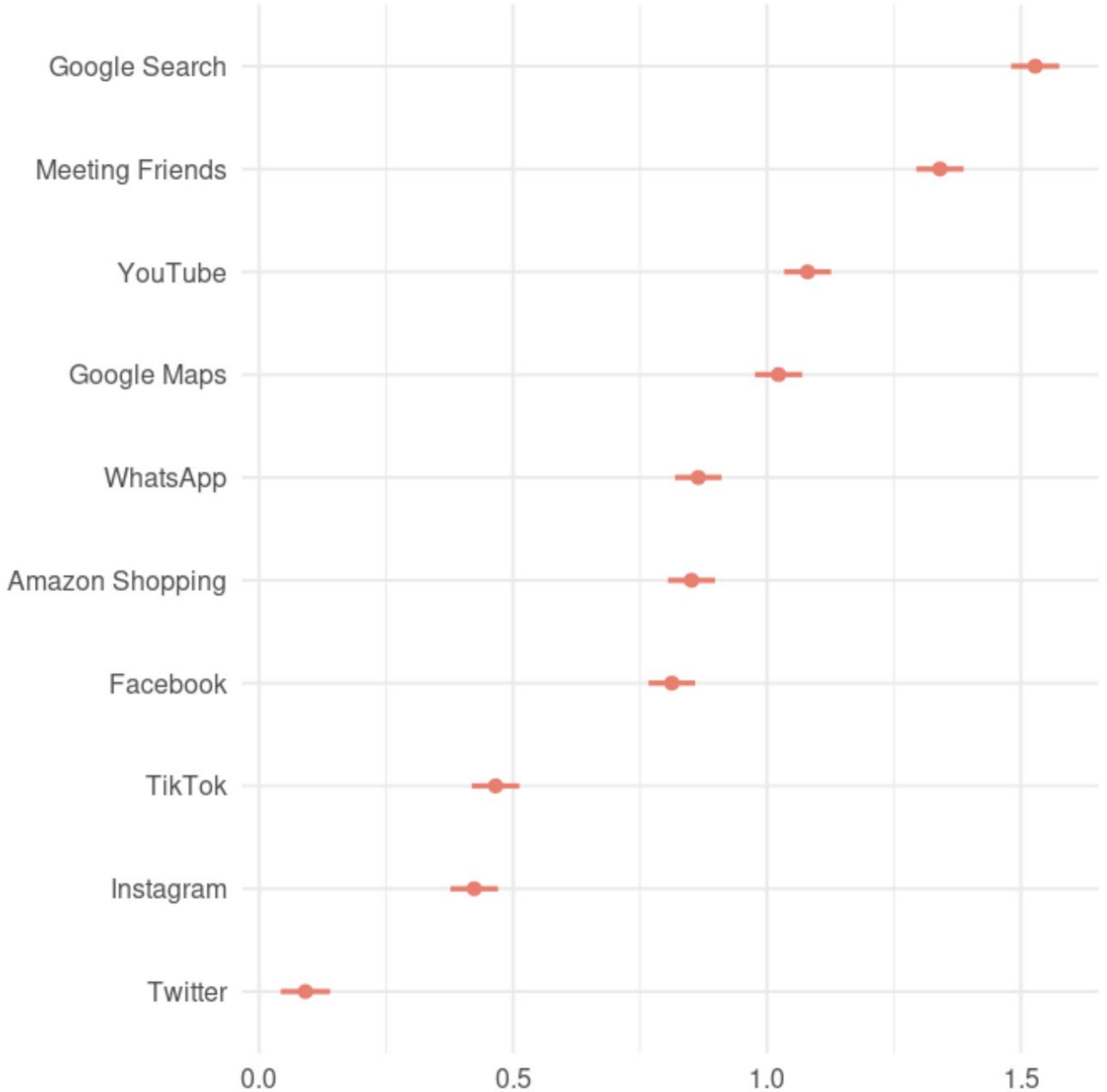


Figure 1: Relative disutility from stopping use, estimated using a conditional logit model. Snapchat is the omitted category.

Figure 1 Notes: The figure depicts the disutility from stopping use of the digital service relative to stopping use of Snapchat, which is the least preferred digital service. For instance, stopping use of Google Search for a month causes the largest disutility—relative to stopping use of Snapchat—among users of the 13 countries. Note that these figures reflect the valuation of the average Facebook user and not necessarily the valuation of the average user of these other platforms or non-users. Note also that value to users is different from value to advertisers for these services, the estimation of which would require a different measurement strategy

(b) **Calibrating welfare gains using Facebook**

In the same survey of Facebook users from 13 countries, we used an incentivized single-binary discrete choice (SBDC) experiment to elicit respondents' willingness to accept to stop using Facebook for one month. We chose Facebook for this study because we can ensure compliance of choices for selected users (we are able to observe their true Facebook usage). Respondents were asked: "Would you be willing to stop using Facebook for one month in exchange for X?", where X was chosen randomly from a set of nine monetary values ranging from $5 to $100. We clarified to the respondents that they could be randomly selected for their choices to be fulfilled, and if so, they would actually be eligible to receive the offer amount if they deactivated their Facebook account for a month. Respondents were given offers in their own currency. For instance, if a respondent in France was chosen to receive an offer equivalent to $50, they were given an offer of 45 Euros, which was equivalent to US$50 at the time of the experiment.

Figure 2a below shows the proportion of respondents that rejected each offer. 81% of users rejected the lowest offer (equivalent to US$5) and 24% of respondents rejected the highest offer (equivalent to US$100). As predicted by the law of demand, the proportion of users rejecting the offer decreases monotonically with increasing offer amounts.

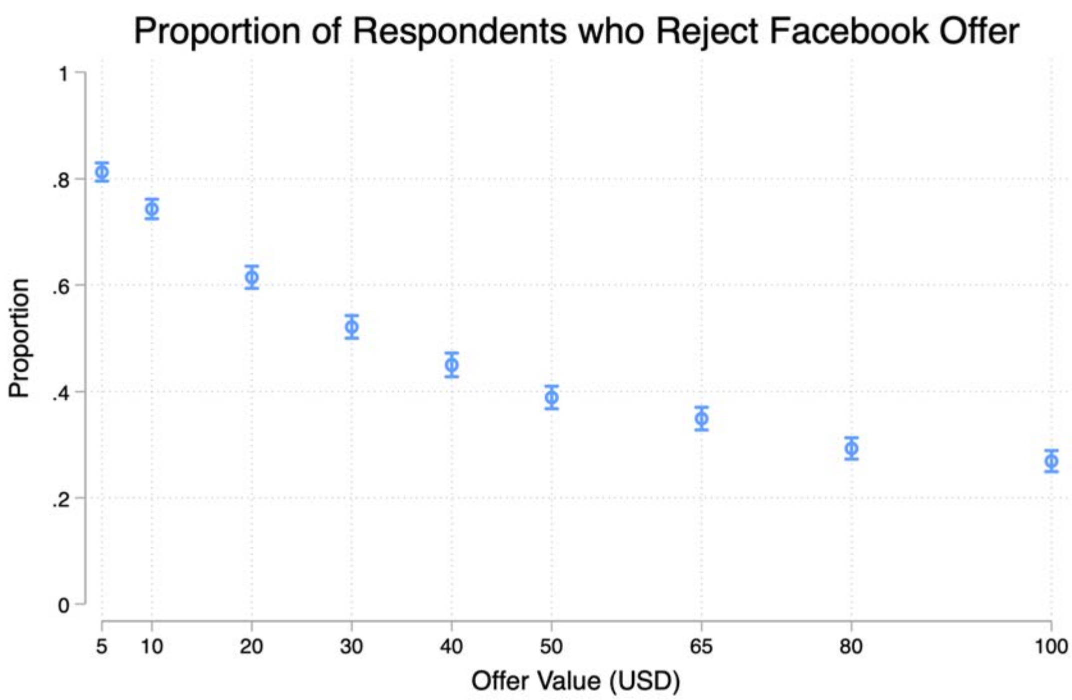


Figure 2a: Facebook offer rejection rates by offer value across all countries in the sample.

Figure 2a Notes: This figure shows, per each offer value, the proportion of respondents who reject the offer to stop using Facebook for one month. Offer values were randomly assigned. The proportion of respondents who reject the offer decreases as the offer value increases.

The median value of Facebook is the amount of money that half of respondents would accept and half would reject. We first estimate a weighted binary logit model to measure how the probability of rejecting the offer depends on its amount and then solve for the offer amount that sets the probability of rejecting to 50%. To ascertain how the value that users derive from Facebook is distributed, we first examine how the median value varies across surveyed countries (Figure 2b). The overall weighted median value is $31 overall, and ranges from $11 in Romania to $57 in Norway.

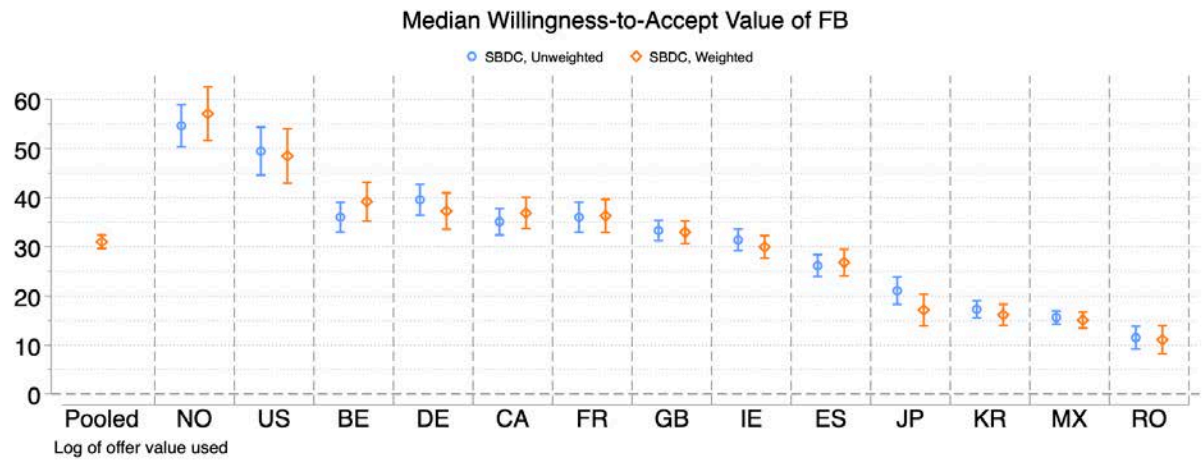


Figure 2b: Facebook median willingness to accept (WTA) by Country

Figure 2b Notes: This figure shows estimates from separate binary logit models for the median WTA value for each country. The estimates and confidence intervals in blue are unweighted, while those in orange are weighted in line with the weighting strategy described in Supplementary Materials Section 1.2. Countries are arranged in order of decreasing weighted median WTA value of Facebook.

Welfare gains from digital goods across countries

We use data on Facebook valuations to first calculate total welfare generated by Facebook across the 13 countries in our sample. We do this by multiplying the weighted median willingness-to-accept value (shown in Figure 2b) by the number of monthly active Facebook users in that country, and then multiply this number by 12 to annualize it (Appendix Figure 7).[10] Using this approach, Facebook generates a total of $246 billion in welfare across these countries (ranging from $137 billion in the US to $1.2 billion in Ireland).

[10] We assume that the willingness-to-accept (WTA) value to stop using Facebook for a year is approximately 12 times as large as the analogous willingness-to-accept for a month. Arguments can be made to support the hypothesis that willingness-to-accept increases more than proportionally or less than proportionally with respect to disconnection time. Previous research (Brynjolfsson et al., 2019a) shows that WTA to stop using Facebook for 1 month (4 weeks) is slightly more than four times WTA to stop using Facebook for 1 week, and WTA to stop using Facebook increases more than proportionally over time (from 1 month to 1 year). Therefore, our assumption might slightly underestimate the welfare generated by Facebook over a period of 12 months.

In turn, using the incentivized single-binary discrete choice (SBDC) Facebook valuations as our benchmark and applying the relative utilities across goods estimated with the best-worst scaling (BWS) method, we calibrate the valuations of other goods and estimate total welfare generated by all these 10 goods in our sample across all countries (Appendix Figure 12 shows the calibrated valuations of each digital good in each country[11]). This is done by multiplying the aggregate annual value of Facebook in each country by the utility of each digital good relative to Facebook—where the utility for Facebook is normalized to 1—obtained through the best-worst scaling (BWS) estimation. Our analysis implies that, among Facebook users, the 10 digital goods selected in this study generate a combined total of $2.52 trillion in welfare across these countries—ranging from $1.29 trillion in the US to $13 billion in Romania (Figure 3a).

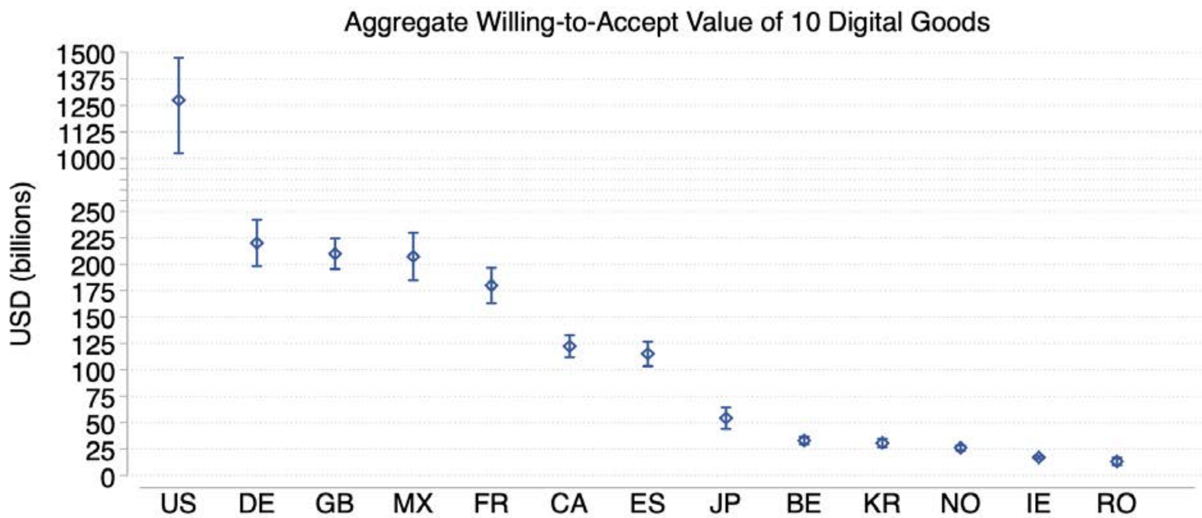


Figure 3a: Aggregate annual value of 10 digital goods by country

Figure 3a Notes: This figure presents country-level estimates of the aggregate annual value of all ten digital goods studied using relative valuations from the best-worst scaling study calibrated with incentivized Facebook valuations using the single-binary discrete choice study. Countries are arranged in order of decreasing weighted median WTA value of Facebook. The y-axis includes a discontinuity in order to visually accommodate the US estimate which is far higher than the estimate for other countries.

To explore whether the welfare gains from digital goods vary across higher-income and lower-income countries we calculate, for each country, the ratio of Facebook users' median valuations of all the digital goods as a percent of GDP per capita (Appendix Figure 8 plots these figures for Facebook and Appendix Figure 10 for all digital goods). We then regress this variable on a country's GDP per capita:

$$Valuation\ of\ Digital\ Goods\ (\%\ of\ GDPpc)_i\ =\ \beta\ w\ GDPpc_i + \varepsilon_i$$

[11] There is substantial heterogeneity in valuations of digital goods across countries. For example, Google Search is the most valued good in the US, WhatsApp is the most valued good in Mexico and YouTube is the most valued good in South Korea.

Where β denotes the parameter to be estimated, and w denotes weights that in this case are the monthly active Facebook users in each country. The rationale for including the weights is to equally consider all Facebook users in the regression.

We find a highly significant negative relationship between users' digital goods valuation as a percent of GDP per capita in a country and GDP per capita in that country (Figure 3b, see Appendix Figure 9 for an analogous plot with the valuation of Facebook alone in the Y axis). A \$10,000 increase in a country's GDP per capita is associated with a 2.09 percentage point decrease in users' valuation of the 10 digital goods relative to GDP per capita (β = -2.09, p-value < 0.001).[12] In other words, among Facebook users, the welfare gains from digital goods represent a higher share of income in lower income countries compared to higher income countries.[13]

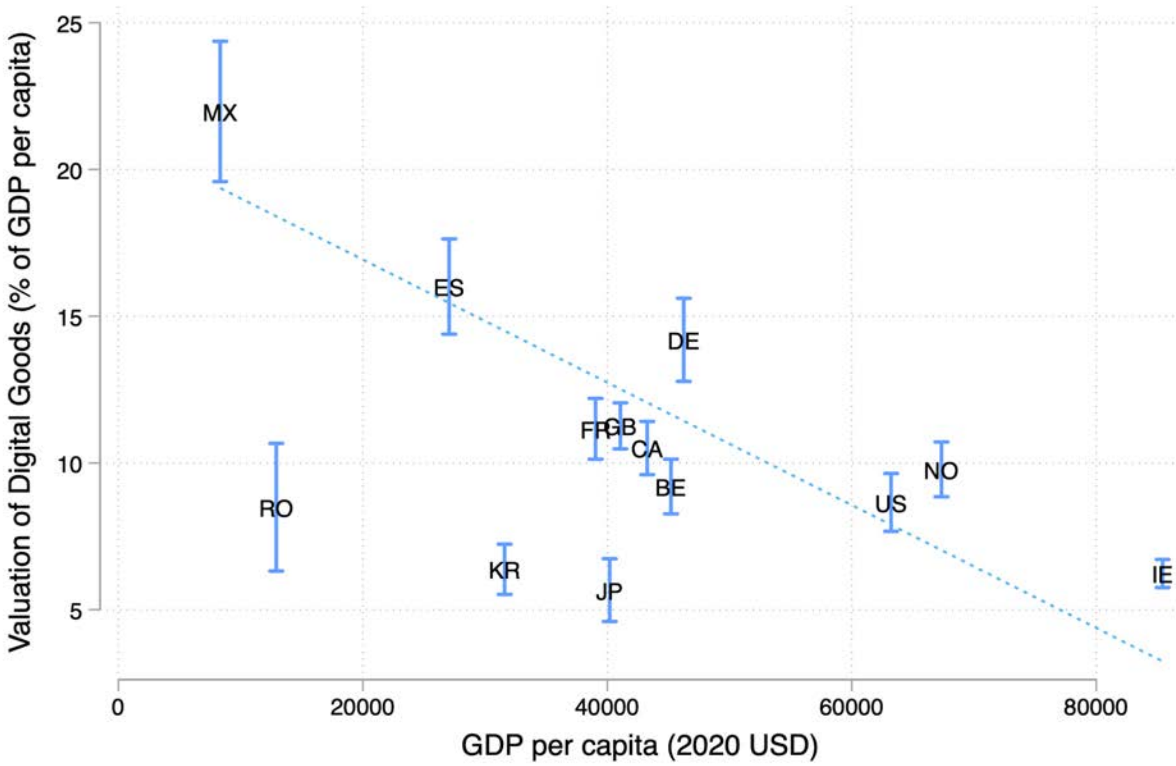


Figure 3b: Association between GDP per capita and Facebook users' valuation of 10 digital goods as a share of GDP per capita

---

[12] The unweighted regression gives a β equal to -1.11 with a p-value equal to 0.068. In Section 1.3 of the Appendix we discuss a simulation procedure that accounts for the statistical uncertainty around our estimates of the value of the 10 digital goods as a percent of GDP per capita. The simulation exercise shows that the results of the regression are robust to variations in the country estimates of the value of digital goods as a percent of GDP per capita.

[13] Appendix Figure 8 shows the valuation of Facebook as a percent of GDP per capita in each country. Appendix Figure 9 shows a downward-sloping pattern—consistent with Figure 3b—between the valuation of Facebook as a percent of GDP per capita and GDP per capita.

Figure 3b Notes: The figure depicts 2020 GDP per capita (in US dollars) in the X axis and Facebook users' valuation of the 10 digital goods as a percent of 2020 GDP per capita in the Y axis. A weighted regression of Y on X (in $10,000s) with weights given by Facebook monthly active users (MAU) in each country yields a point estimate of -2.09 with a p-value equal to <0.001. An unweighted regression yields a point estimate of -1.11 with a p-value equal to 0.068. The fitted lines in the plot correspond to the weighted regression.

Welfare gains from digital goods within countries

To analyze heterogeneity in welfare gains across income levels *within* each country, we calculate valuations of Facebook and other digital goods by income levels and education. For income, we use the relative wealth index of the location of a user's residence as a proxy (Chi et al. (2022)).[14] For education, we rely on the responses that users provided in our survey.

We find that the monetary value that users derive from Facebook does not tend to vary by these indicators of material welfare. For instance, users who live in the least wealthy localities classified by the Relative Wealth Index have a value of Facebook that is similar to that of users in the highest wealth localities (Figure 4a).[15] Similarly, users with less than secondary education or with just secondary education (and who therefore tend to earn less on average) have a median value of Facebook that is not statistically distinguishable from the value that users with a college degree derive from Facebook (see Appendix Figure 6).[16]

These findings imply that the value of Facebook represents a higher *share* of their income and wealth for users who currently have lower income and wealth. Extending to other digital goods, we estimate separate conditional logit models on users who are in the bottom, middle, and top tercile of relative wealth in their respective countries in Figure 4b (Appendix Figure 11 plots this for the US alone). Interestingly, we find that the poorest and wealthiest users often have greater value for most digital goods than those in the middle of the relative wealth distribution, though the pattern is inconsistent across digital goods. For some digital goods (e.g. Google Search), users in the top tercile derive the highest absolute welfare in dollar terms while for other digital goods (e.g. TikTok) those in the lowest tercile benefit the most. The differences in valuations are rarely

[14] The relative wealth index calculated in Chi et al. (2021) estimates the relative wealth and poverty of an area at 2.4 km resolution. From the paper: "The estimates are built by applying machine-learning algorithms to vast and heterogeneous data from satellites, mobile phone networks, and topographic maps, as well as aggregated and deidentified connectivity data from Facebook. We train and calibrate the estimates using nationally representative household survey data from 56 LMICs and then validate their accuracy using four independent sources of household survey data from 18 countries."

[15] We do not have perfect information about users' locations. The more granular the geographic area, the less accurate the location predictions are. However, the accuracy of Facebook's zip-code level predictions in the US is as high as 68%—although note that the 2.4-km microregions defined in Chi et al. (2022) do not neatly correspond with zip-codes. To the extent that relative wealth index levels are geographically clustered, small imprecisions in the location predictions should not substantially impact the accuracy of the RWI tercile categorizations. That said, any noise in the relative weight indices may bias the relationship between Facebook valuations and the RWI index toward zero.

[16] We also explore heterogeneity in Facebook valuations based on home ownership and gender (Appendix Figure 6), which are both correlated with wealth. Facebook valuations do not significantly vary based on home ownership (owned vs. rented home). Women value Facebook significantly higher than men (and women's wealth is lower than men in all of our sample countries).

significantly different from each other across terciles which suggests that, on balance, these digital goods tend to lower welfare inequality within countries.

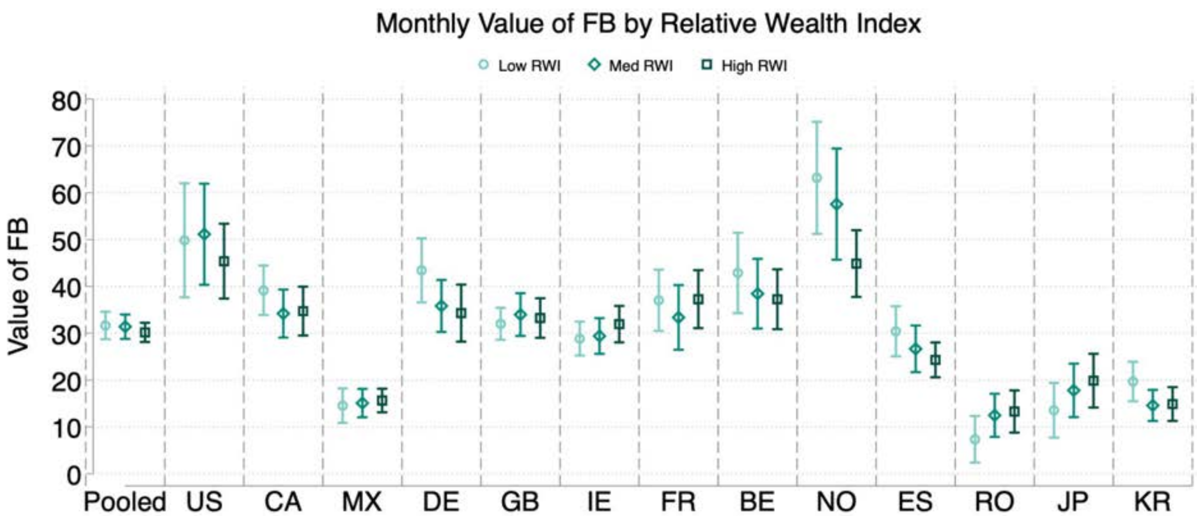


Figure 4a: Value of Facebook by relative wealth index

Figure 4a Notes: This figure shows weighted estimates from separate binary logit models for the median WTA value for each relative wealth index tercile within each country.

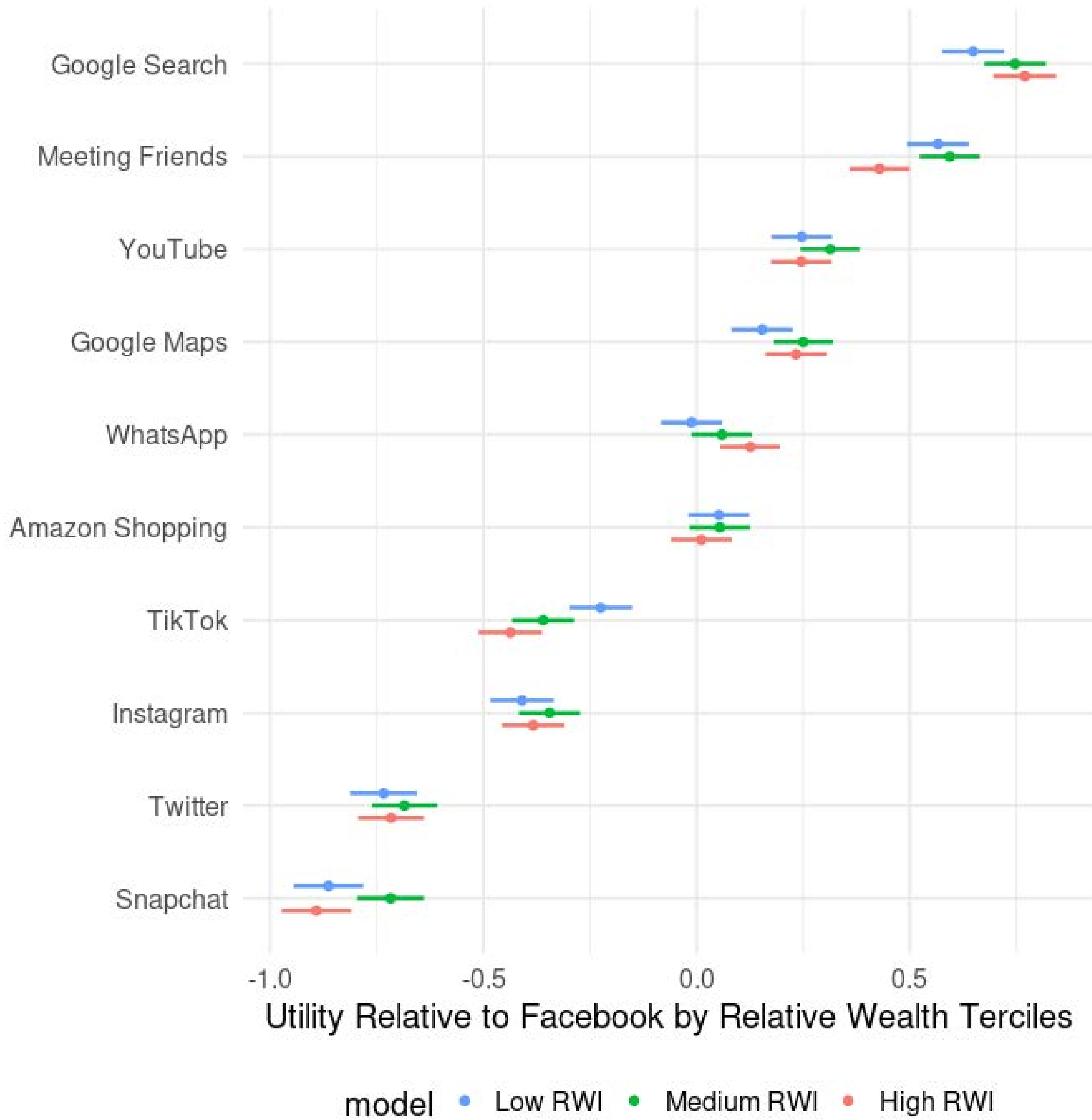


Figure 4b: Relative valuation of 10 digital goods by relative wealth index

Figure 4b Notes: This figure shows estimates from three separate conditional logit models (each containing users belonging to the low, medium and high RWI terciles within their respective countries) of the utility of each digital good relative to Facebook (which is the omitted good, with utility set at 0). These results should be interpreted in conjunction with those in Figure 4a.

Comparing consumer welfare with time spent and firm revenue

Previous research has estimated consumer welfare generated by free digital goods using measures of time spent (Goolsbee and Klenow, 2006; Brynjolfsson and Oh, 2012) and advertising revenues (Nakamura et al., 2017). Do our measures of consumer welfare capture additional information beyond measures of time spent and advertising revenues? To explore this, we compare our estimates of valuation of Facebook with time spent on Facebook. Figure 5 plots these comparisons for Facebook valuations across all the 13 countries pooled together. We split our study respondents into three terciles—low, medium, and high—based on time spent on Facebook. For each of these terciles, we calculate the median valuation of Facebook.

We find that users in the first tercile—i.e. users who spend the least time on the platform—have a median valuation of $19.72 per month. Users in the second tercile have a median valuation of $32.97 per month. Finally, users in the third tercile have a median valuation of $40.61 per month. Moving from the first tercile to the third tercile, the valuation of Facebook increases by 2.06 times while time spent increases by 12.84 times. Thus, valuation increases much more slowly than time spent, implying that value is distributed across a broad user base rather than concentrated on a few very active users.

We can also compare consumer welfare gains to revenues for the producers. Nordhaus (2005) estimated that only a small portion of the total welfare generated by technological advances in the 1948-2001 time period was ultimately captured by producers. Instead, consumers enjoyed the vast majority of the welfare gains. Our study finds results that are consistent with Nordhaus (2005): when we compare our welfare estimates to advertising revenues, we find that user value for just the Facebook app in the 13 countries studied ($284 billion) is nearly three times the global advertising revenue of Meta Platforms' ($115 billion, including Facebook, Instagram, and WhatsApp). Tadelis et al. (2023) find that each dollar spent on Meta ads leads to over three dollars in revenues for advertisers. Our findings imply that the vast majority of the welfare gains from using Facebook go to consumers and not to Facebook.

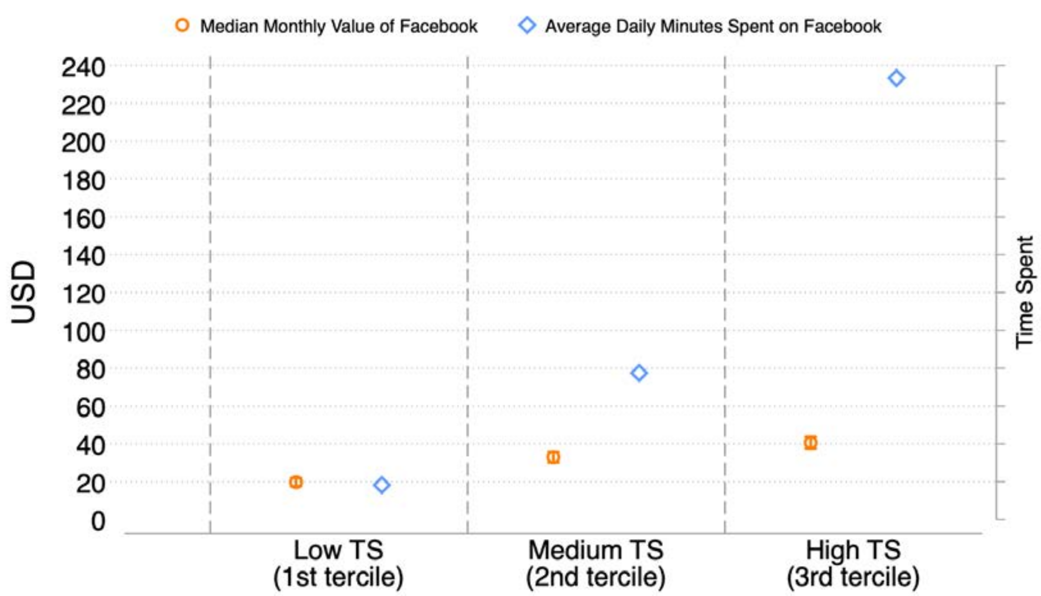


Figure 5: Comparing monthly valuation of Facebook with daily time spent

Discussion

Digital goods generate large benefits for consumers, but because most of these goods are free to use, these benefits are largely invisible in standard government statistics such as GDP and productivity. In this paper, we provide estimates of the value digital goods create for users in 13 countries around the world by conducting large-scale incentivized online choice experiments on representative samples of nearly 40,000 people.

We find that the 10 selected digital goods across the 13 countries generate more than $2.5 trillion in aggregate consumer welfare per year, roughly equivalent to 6% of GDP in these countries. We also find that lower-income individuals and countries disproportionately benefit from these digital goods. These findings suggest that digital goods reduce inequality in welfare within and across countries.

Our approach is subject to a number of limitations. Compared to GDP, which can be measured with high precision, our estimates are relatively noisy. We are confident in our qualitative findings that these digital goods create trillions of dollars of value and reduce welfare inequality, but the exact values are not precisely estimated. The large sample size in our study partly mitigates this problem via the law of large numbers, but there may remain systematic biases in our estimates for a variety of reasons.[17] Relatedly, we study a particular sample of countries and of digital goods. While a pattern is evident within this large sample, different effects may occur for other sets of countries and goods. While our sample account for a substantial fraction of the likely value of global GDP and of value from digital goods, the out-of-sample implications can best be addressed by simply expanding the sample.

Furthermore, even when the valuations we obtain are accurate, they may reflect irrational choices that are not in the consumers' genuine self-interest due to "digital addiction" (Allcott et al., 2022) or other errors in judgment (Kahneman et al., 1982). In addition, these goods may create positive or negative externalities on other people — from shared memories and connections to misinformation and polarization — which means that the total welfare gains are not necessarily equal to the sum of individual valuations. While these concerns are important, it should be noted that the same concerns apply to standard measures of GDP, which also reflect consumer values which may be irrational or omit important externalities. Future work should seek to address these concerns, and online choice experiments can also be used to quantify these externalities (e.g., Bishop et al., 2017; Collis and Eggers, 2022).

Having demonstrated the feasibility of running massive online choice experiments to estimate valuations of multiple goods across multiple countries, future work can expand this line of research in at least three dimensions: i) More goods, including non-digital goods like breakfast cereal, improved healthcare, or cleaner water, ii) More countries or regions, and iii) More respondents per item (which will increase the precision of our estimates). Furthermore, by

[17] See Brynjolfsson et al. (2019b) for a more detailed discussion of potential biases and shortcomings of massive online choice experiments.

conducting online choice experiments such as this one at a regular cadence, e.g. annually or even more frequently, and with consistent methods, we can better understand not only levels but also changes in welfare as the basket of goods and other variables change over time.

This paper provides a first step toward systematically estimating welfare using massive online choice experiments. Since the contribution of digital goods to welfare is likely to continue to grow in the twenty-first century, establishing a reliable baseline will provide a foundation for understanding the magnitude and nature of future changes in the economy.

Methods

We surveyed 39,717 Facebook users in 13 countries on the Facebook internal survey platform. The two main components of the survey were (i) a hypothetical willingness to accept measurement for various digital goods using best-worst scaling, and (ii) an incentivized willingness to accept measurement for Facebook using single binary discreet choice. The survey sample for each component was weighted to be representative of the population of active Facebook users in each country. The survey invitation was shown to respondents at the top of their Facebook Feed (Appendix Figure 1).

Sampling & Weighting

We recruited our sample by sending in-app survey invitations to Facebook users. All 18+ Facebook users in these 13 countries who had been active on the Facebook platform in the month before the start of the survey, and whose accounts had been created at least 30 days ago, were eligible to be included in the sampling frame. The 13 countries included in our study are the United States, Canada, Mexico, Germany, United Kingdom, Ireland, France, Belgium, Norway, Spain, Romania, Japan, and Korea. We selected these 13 countries based on a combination of research interest and the availability of survey pool resources in each country. We use inverse probability weighting models to account for selection in our sample, such that all our estimates are representative of the population of Facebook users in each country. Our ability to weight results to match populations of Facebook users is an improvement over existing work. For all analyses, we account for both unit and item non-response bias. For analyses in which we pool data across countries, we also account for differential probability of inclusion in our sample using design weights. This is necessary since the probability of inclusion in our sample is different across countries. The Appendix provide further details of the weighting methodology.

Measuring Relative Welfare from Digital Goods

We used a best-worst scaling (BWS) methodology to measure the relative willingness to accept for stopping the use of 10 digital goods or not meeting friends in person for one month. The digital goods include social media / messaging tools (Facebook, Twitter, Instagram, WhatsApp, Snapchat, and TikTok), and other digital tools (Google Search, Google Maps, YouTube, and Amazon Shopping). Using a balanced-incomplete block design, we generated 70 questions with the 11 items mentioned above plus 10 monetary amounts. We dropped 7 questions that included only monetary amounts, resulting in 63 questions from which we presented 3 random questions to each respondent. Within each question, respondents were asked to indicate their 'best' and 'worst' options from among three options presented to them. These questions were presented to respondents at the end of the survey.

The sample size for best-worst scaling questions is 18,443. A total of 23,752 respondents[18] answered best-worst scaling questions, out of which 18,443 remained after applying quality checks (see below for details). We calculate separate weights for best-worst scaling respondents using a similar methodology to the one outlined in Appendix Section 1.2 to ensure that the best-worst scaling respondents are representative of the Facebook population in each country. Appendix Figure 2 shows an example BWS task.

Calibrating Welfare Gains Using Facebook

In the same survey of Facebook users from 13 countries, we used an incentivized single binary discreet choice method to elicit respondents' willingness to accept to stop using Facebook for 1 month (Appendix Figures 3 and 4). Respondents were asked: "Would you be willing to stop using Facebook for one month in exchange for X?", where X was chosen randomly from a set of 9 monetary values from $5 to $100. We clarified to the respondents that they could be randomly selected for their choices to be fulfilled, and if so, they would actually be eligible to receive the offer amount if they deactivated their Facebook account for a month.

Respondents were given offers in their own currency. For instance, if a respondent in France was chosen to receive an offer equivalent to $50, they were given an offer of 45 Euros, which was equivalent to US$50 at the time. These offers were incentivized, and participants had to agree to a set of terms and conditions drafted by legal experts to be in compliance with local laws in each country. The amounts paid ranged from US$5 to US$510. A total of approximately US$14,400 were paid out to 113 participants. Each choice made by respondents of whether to accept or reject an offer had a 2% probability of being selected. This probability was not known to respondents.

The Appendix provides additional details about our methods as well as supplementary analyses.

References

Allcott, H., Braghieri, L., Eichmeyer, S., Gentzkow, M. (2020). The Welfare Effects of Social Media. *American Economic Review, 110(3),* 629-676.

[18] The sample size for the BWS questions is slightly lower than the sample size for the Facebook SDBC question due to survey attrition. The BWS questions appeared later in the survey. Respondents could stop answering the survey at any time.

Allcott, H., Gentzkow, M., & Song, L. (2022). Digital addiction. *American Economic Review*, *112*(7), 2424-63.

Becker, G. M., M. H. DeGroot and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9(3),* 226-232.

Bishop, R. C., Boyle, K. J., Carson, R. T., Chapman, D., Hanemann, W. M., Kanninen, B., ... & Scherer, N. (2017). Putting a value on injuries to natural assets: The BP oil spill. *Science*, *356*(6335), 253-254.

Brynjolfsson, E., & Collis, A. (2019). How should we measure the digital economy? *Harvard Business Review*, *97*(6), 140-148.

Brynjolfsson, E., Collis, A., & Eggers, F. (2019a). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, *116*(15), 7250-7255.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., & Fox, K. J. (2019b). *GDP-B: Accounting for the value of new and free goods in the digital economy* (No. w25695). National Bureau of Economic Research.

Brynjolfsson, E., & Oh, J. (2012). The attention economy: Measuring the value of free digital services on the internet.

Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, *119*(3), e2113658119.

Collis, A., & Eggers, F. (2022). Effects of restricting social media usage on wellbeing and performance: A randomized control trial among students. *Plos one*, *17*(8), e0272416.

Coyle, D., & Nakamura, L. (2022). Time Use, Productivity, and Household-centric Measurement of Welfare in the Digital Economy. *International Productivity Monitor*, *42*, 165-186.

Fischer, C. (2017). Inequality Is About Access to Public Goods, Not Income. *Boston Review*. April.

Goolsbee, A., & Klenow, P. J. (2006). Valuing consumer products by the time spent using them: An application to the Internet. *American Economic Review*, *96*(2), 108-113.

Hanani, H. (1961). The existence and construction of balanced incomplete block designs. *The Annals of Mathematical Statistics*, *32*(2), 361-386.

Jones, C. I., & Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *American Economic Review*, *106*(9), 2426-2457.

Kahneman, D., Slovic, S. P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. *Cambridge University Press.*

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). Best-worst scaling: Theory, methods and applications. *Cambridge University Press.*

Lowensohn, J. (2012) Apple Features worth up to $100 to Samsung Buyers, Witness Says, *CNET*, August 10. https://www.cnet.com/tech/tech-industry/apple-features-worth-up-to-100-to-samsung-buyers-witness-says/

Masood, E. (2022). GDP is getting a makeover—what it means for economies, health and the planet. *Nature (News Feature)*, 611, 224-226.

Nakamura, L. I., Samuels, J., & Soloveichik, R. H. (2017). Measuring the Free Digital Economy within the GDP and productivity accounts. *FRB of Philadelphia Working Paper.*

Nordhaus, W. D. (1996). Do real-output and real-wage measures capture reality? The history of lighting suggests not. In *The economics of new goods* (pp. 27-70). University of Chicago Press.

Nordhaus, W. D. (2005). Schumpeterian profits and the alchemist fallacy. *Yale Economic Applications and Policy Discussion Paper*, (6).

Slottje, D. J. (1987). Relative price changes and inequality in the size distribution of various components of income: A multidimensional approach. *Journal of Business & Economic Statistics*, *5*(1), 19-26.

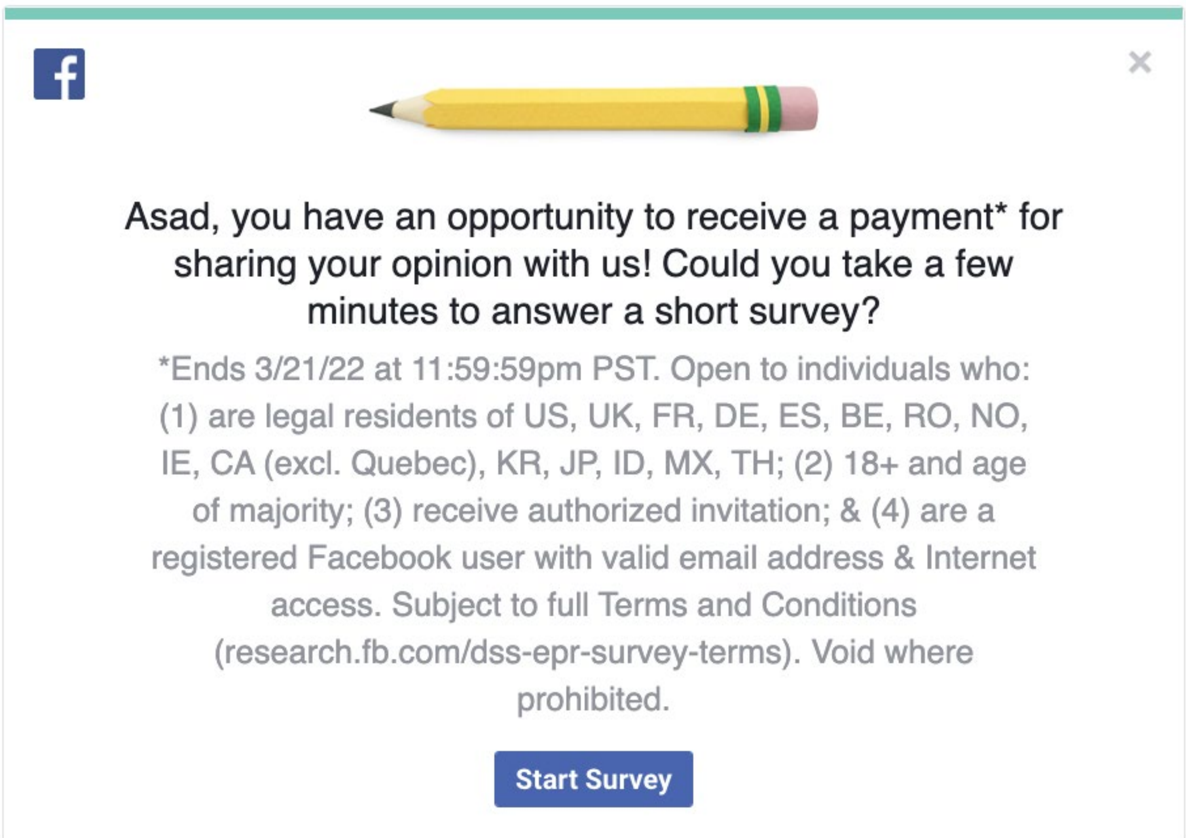
Tadelis, S., Hooton, C., Manjeer, U., Deisenroth, D., Wernerfelt, N., Greenbaum, L., and N. Dadson (2023) Learning, Sophistication, and the Returns to Advertising: Implications for Differences in Firm Performance, NBER Working Paper.

Supplementary Materials

1 Materials and Methods

1.1 Survey Invitation

The survey invitation was shown to respondents at the top of their Facebook Feed. Here is an example of the survey invitation:



Appendix Figure 1: Survey invitation

1.2 Weighting

We use inverse probability weighting models to account for selection in our sample, such that all our estimates are representative of the population of Facebook users in each country. Our ability to weight results to match populations of Facebook users is an improvement over existing work.

For all analyses, we account for both unit and item non-response bias. For analyses in which we pool data across countries, we also account for differential probability of inclusion in our sample using design weights. This is necessary since the probability of inclusion in our sample is different across countries.

1. Design weights:

The probability of inclusion in our sample is different across countries. Our weighting strategy adjusts for the different probability of selection into the sample by country (and treatment / control condition) due to the uneven sample allocation across countries:

Weight for responses from country i = 1 / (Number of users included in the sampling frame from country i / monthly active user population from country i)

The weights obtained using this approach can be interpreted as the number of monthly active users (i.e. population) represented by a user included in the sampling frame (i.e. sample).

2. Unit non-response weights within country:

We adjust the weights to account for unit non-response by modeling the probability that a user in the country's sampling frame responds to the survey as a function of observable user characteristics. For each country, the target population is the sampling frame. To account for the possibility that the response model differs across countries, we model this probability separately by country.

Weight = 1 / Est. Pr(user starts survey)

Following internal standard practice by survey scientists, we use the following user characteristics in unit non-response weighting models: Gender, Primary Phone OS, whether the user has a profile picture, age (quartile bins), friend count (quartile bins), the number of days within the last 28 days that the user was active, an indicator for whether the user was active for all days within the last 28 days, and time since the user created their account (quartile bins).

3. Item non-response weights for the valuation questions

We account for item non-response by modeling the probability that a user in the strata who started the survey responds to the relevant survey question (i.e. answered at least one of the best-worst scaling questions, or answered the Facebook incentivized valuation question). For each strata, the target population is the set of users from that cluster who started the survey. To account for the possibility that the response model differs across countries, we model this probability separately by country.

Weight = 1 / Est. Pr(user answers question | user starts survey)

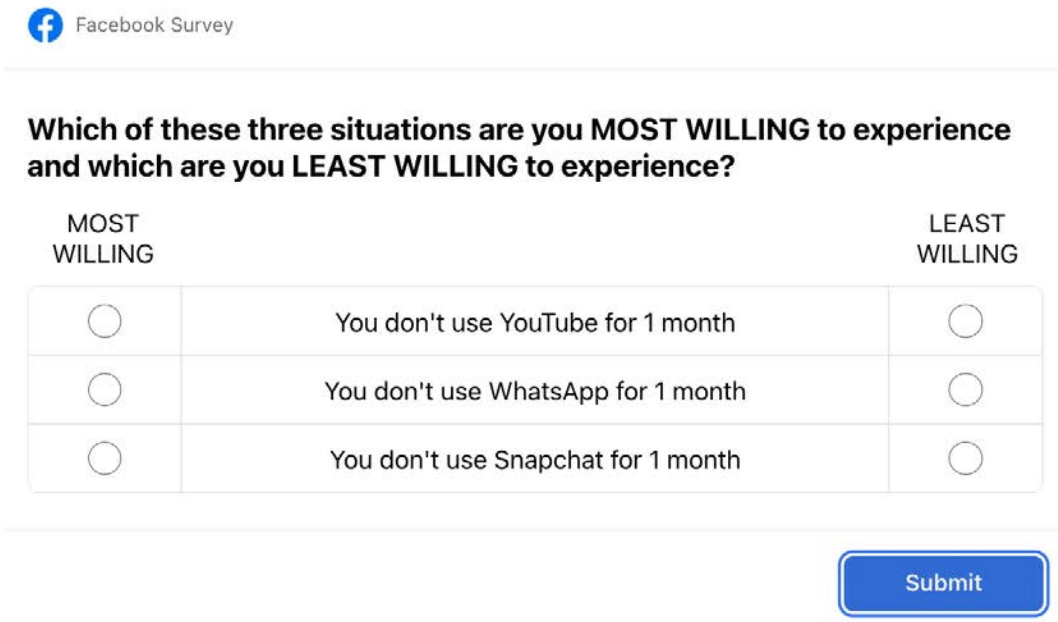
1.3 Simulation Procedure to Account for the Statistical Uncertainty Around the Valuation Estimates

In Figure 3b in the main text, we regress Facebook users' median valuation of the 10 digital goods as a percent of 2020 GDP per capita (in the Y axis) on 2020 GDP per capita (on the X axis). We weight country observations according to the number of monthly active users on Facebook. However, our estimates of the users' valuation of digital goods as a percent of GDP per capita are themselves uncertain. To account for this uncertainty, we ran a simulation analysis. The simulation exercise shows that the results of the regression are robust to variations in the country estimates of the value of digital goods as a percent of GDP per capita.

In the simulations, we draw different values of the Y variable (Facebook users' valuation of the 10 digital goods as a percent of GDP per capita) for each country from a Normal distribution with mean and variance matching each country's point estimate and confidence intervals around Y. For each realization, we run a weighted regression that is analogous to the regression in the main text. We then compute the percent of such regressions that yielded a negative and significant β coefficient. For the weighted regression, we find that 100 percent of such regressions result in a negative and significant β. For the unweighted regression, as the number of simulation realizations grows, the percent of significant β coefficients converges around 17.1 percent. These two results are not surprising insofar as the p-value associated with the weighted regression in the main text is 0.000, the p-value associated with the unweighted regression in the main text is 0.068, and the confidence intervals around the Y estimates are fairly tight.

1.4 Best-Worst Scaling

We used a best-worst scaling methodology to measure the relative willingness to accept for stopping the use of 10 digital goods or not meeting friends in person for one month. Below is an example question:



Appendix Figure 2: Example BWS task

To estimate the relative value that respondents place on these digital goods, we fit a conditional logit model on these best-worst scaling responses. The response data is structured such that there is a separate row for each possible response pair (i.e. each possible way in which respondents could have indicated their best and worst options). Since each question includes three choices, there are six possible rows for a single response to a best worst scaling question. Separate variables for each of the 21 items (except for 1 'excluded' category) are included as independent variables in the model. For each row, the best and worst responses for that row are given the values 1 and -1 respectively under the maxdiff model (with the other items being assigned the value 0). The dependent variable in these models is a dichotomous response variable taking on the value 'TRUE' if the row corresponds to the respondent's choice for that question and 'FALSE' otherwise.

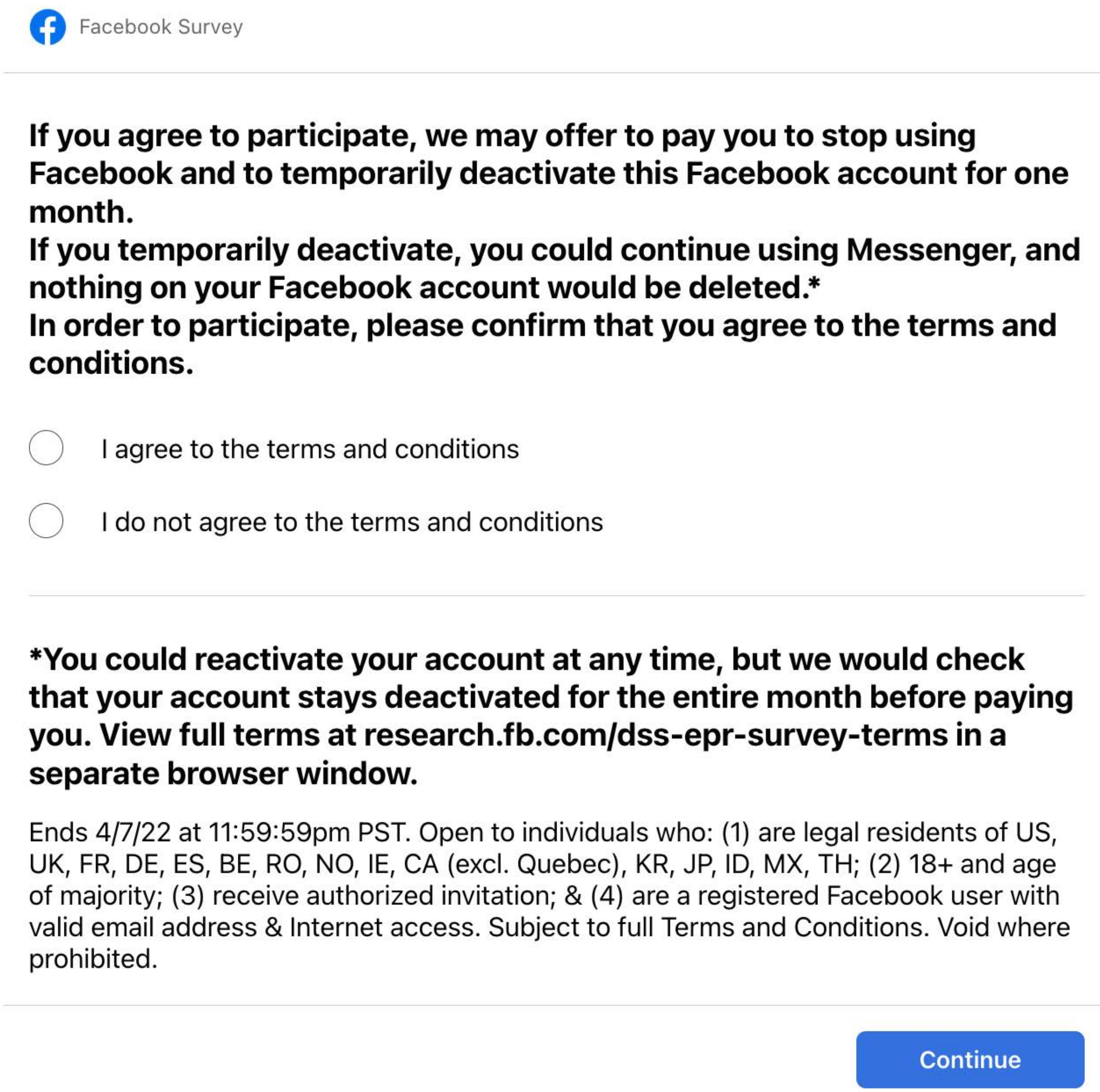
We employ an attention check embedded in the research design to exclude respondents who did not rationally evaluate the relative utility of monetary amounts. This attention check appears in questions that have two monetary amounts and one digital good as the two options. Out of 23,752 respondents, 16061 (68%) answered such a question. Out of these, 10,752 (67%) passed the attention check and 5309 (33%) failed. We exclude these 5309 respondents from our analysis, which leaves 18,443 respondents.

1.5 Incentivized Deactivation Experiment to Calibrate Welfare Gains Using Facebook

In the same survey of Facebook users from 13 countries, we used an incentivized single binary discreet choice method to elicit respondents' willingness to accept to stop using Facebook for 1 month. Respondents were asked: "Would you be willing to stop using Facebook for one month in exchange for X?", where X was chosen randomly from a set of 9 monetary values from $5 to $100. We clarified to the respondents that they could be randomly selected for their choices to be fulfilled, and if so, they would actually be eligible to receive the offer amount if they deactivated their Facebook account for a month.

Respondents were given offers in their own currency. For instance, if a respondent in France was chosen to receive an offer equivalent to $50, they were given an offer of 45 Euros, which was equivalent to US$50 at the time. These offers were incentivized, and participants had to agree to a set of terms and conditions drafted by legal experts to be in compliance with local laws in each country. The amounts paid ranged from US$5 to US$510. A total of approximately US$14,400 were paid out to 113 participants. Each choice made by respondents of whether to accept or reject an offer had a 2% probability of being selected. This probability was not known to respondents.

A total of 379 respondents were selected to deactivate their account. Out of these, 170 (45%) attempted to deactivate, based on log data. 113 (30%) successfully deactivated and were paid. There may be multiple reasons why compliance with deactivation was not higher. Deactivation was a difficult five-step process, involving at least 8 clicks. Furthermore, we were only able to contact respondents for deactivation via email. Respondents may not have checked their emails in time, they may have missed them, or they may have been filtered as spam.



Facebook Survey

If you agree to participate, we may offer to pay you to stop using Facebook and to temporarily deactivate this Facebook account for one month.
If you temporarily deactivate, you could continue using Messenger, and nothing on your Facebook account would be deleted.*
In order to participate, please confirm that you agree to the terms and conditions.

- I agree to the terms and conditions
- I do not agree to the terms and conditions

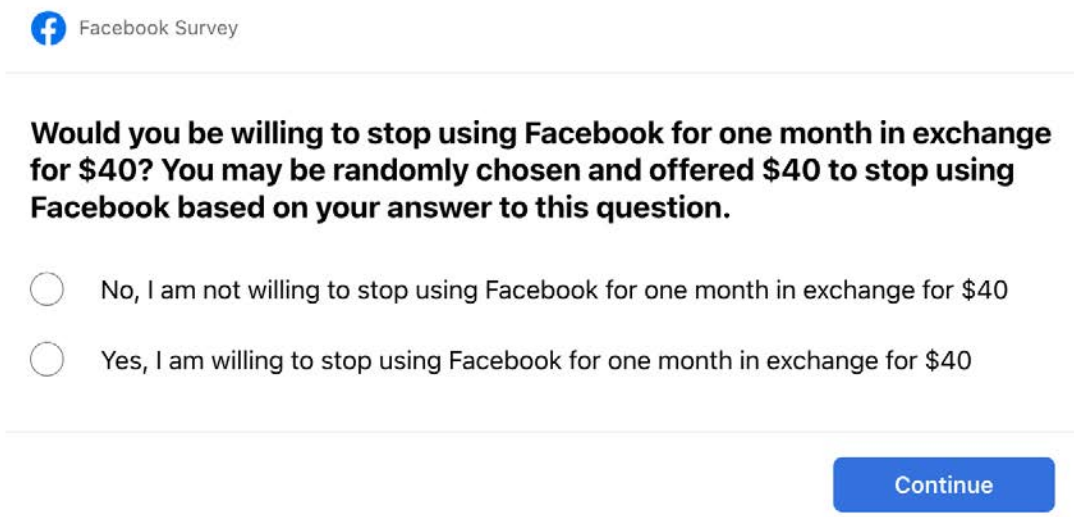
***You could reactivate your account at any time, but we would check that your account stays deactivated for the entire month before paying you. View full terms at research.fb.com/dss-epr-survey-terms in a separate browser window.**

Ends 4/7/22 at 11:59:59pm PST. Open to individuals who: (1) are legal residents of US, UK, FR, DE, ES, BE, RO, NO, IE, CA (excl. Quebec), KR, JP, ID, MX, TH; (2) 18+ and age of majority; (3) receive authorized invitation; & (4) are a registered Facebook user with valid email address & Internet access. Subject to full Terms and Conditions. Void where prohibited.

Continue

Appendix Figure 3: Terms and conditions screen

Here is an example of the offer made to respondents after they had agreed to the terms and conditions:



Appendix Figure 4: Example Facebook offer screen

To arrive at the median willingness-to-accept value of Facebook, we first estimate a weighted logit model to ascertain how the probability of rejecting the offer depends on the amount of the offer:

$$P(Reject\ Offer) = \Lambda(\beta_0 + \beta_1\, Offer)$$

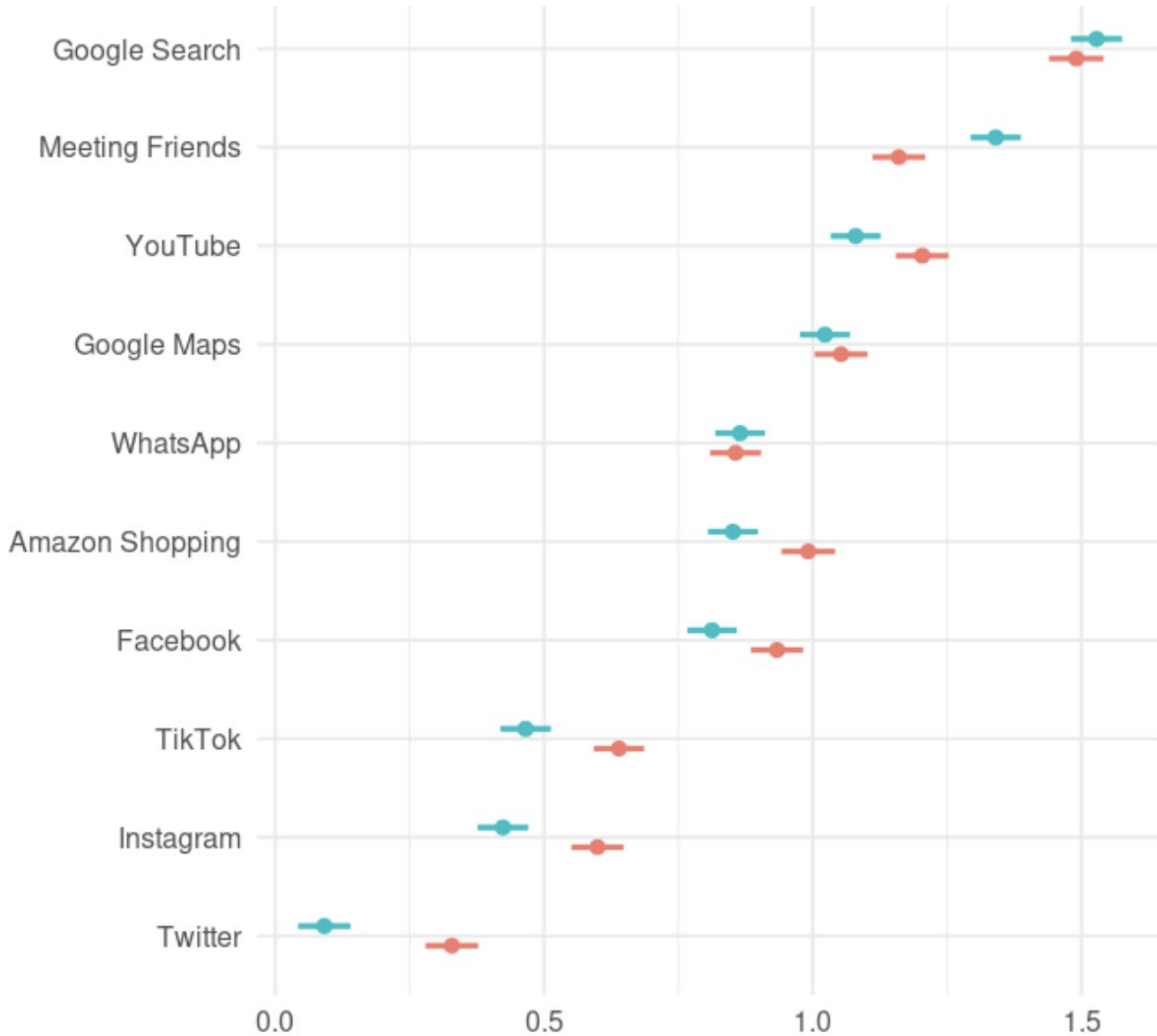
where Offer is equal to the randomly assigned offer amount from $5 to $100, and RejectOffer is an indicator equal to 1 when a respondent answers that they are not willing to stop using Facebook in exchange for that offer amount.

To find the median, we solve for the offer amount that sets this probability equal to 0.5, which yields:

$$Offer* = -\beta_0 / \beta_1$$

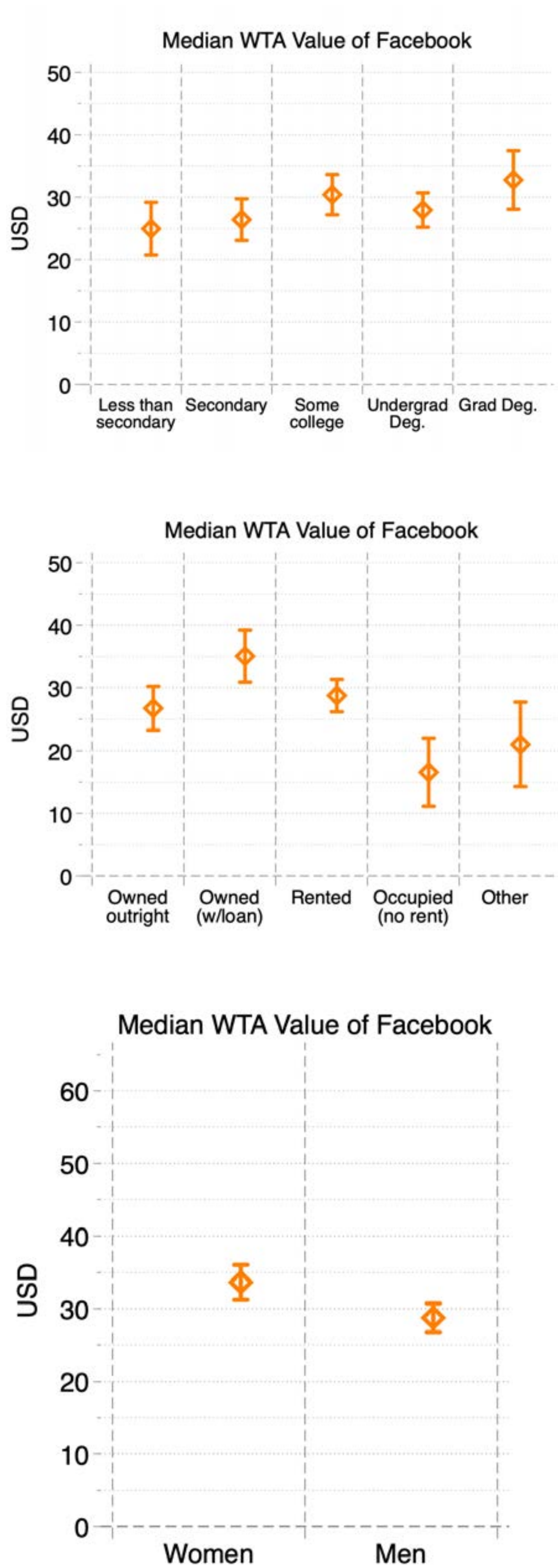
We use bootstrapped standard errors for the confidence interval estimation.

2 Additional Results and Robustness Checks

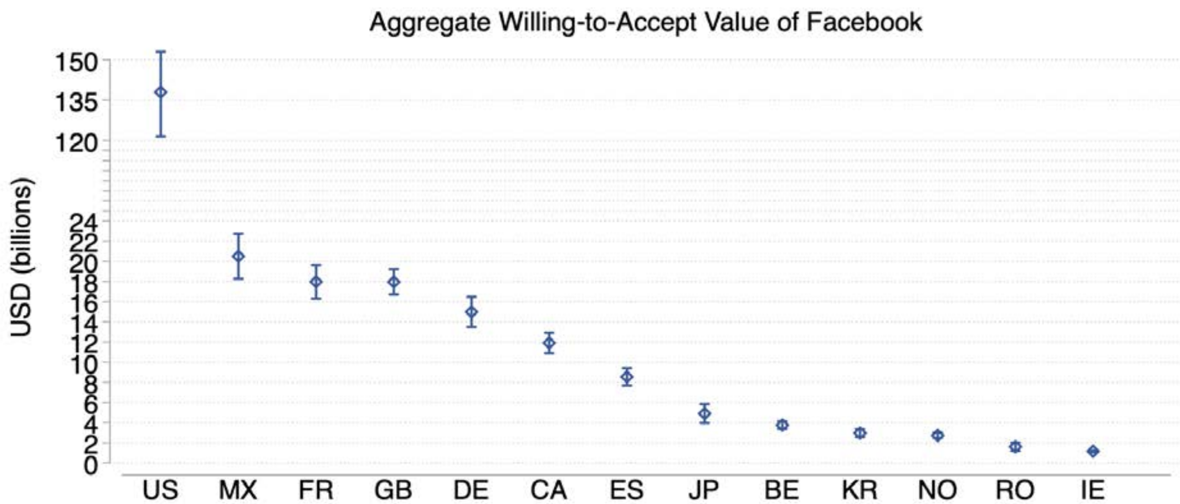


Appendix Figure 5: Relative disutility from stopping use, estimated using a conditional logit model. Snapchat is the omitted category.

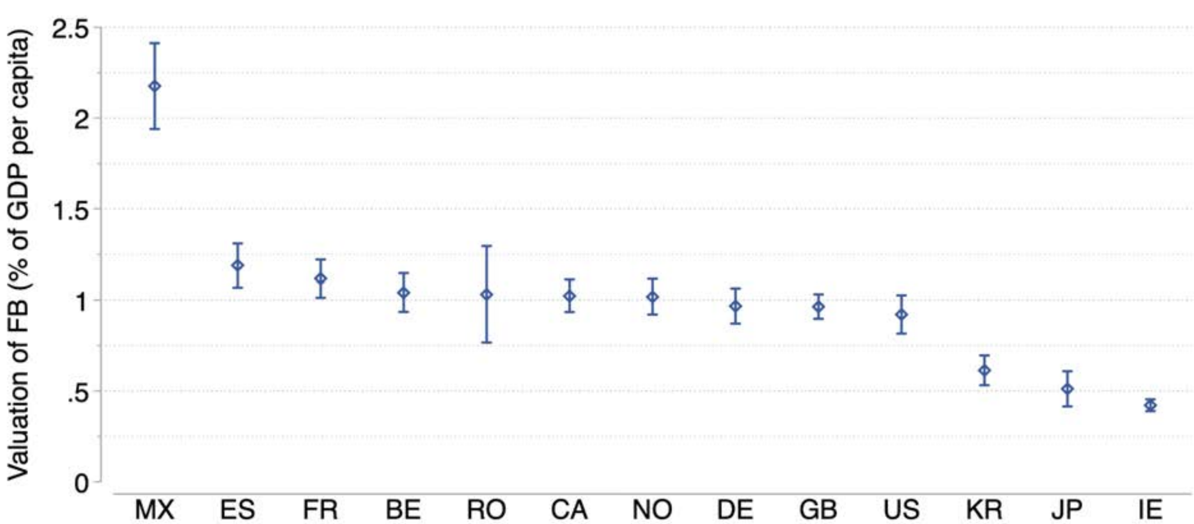
Figure Notes: The figure compares BWS estimates under (i) the full sample and (ii) an alternate model including only the observations containing comparisons of non-monetary items. The full sample excludes respondents who failed attention checks. Please see Appendix Section 1.4 (Measuring Relative Welfare from Digital Goods) for a description of the attention checks. This figure shows that the results are robust to the exclusion of non-monetary values.



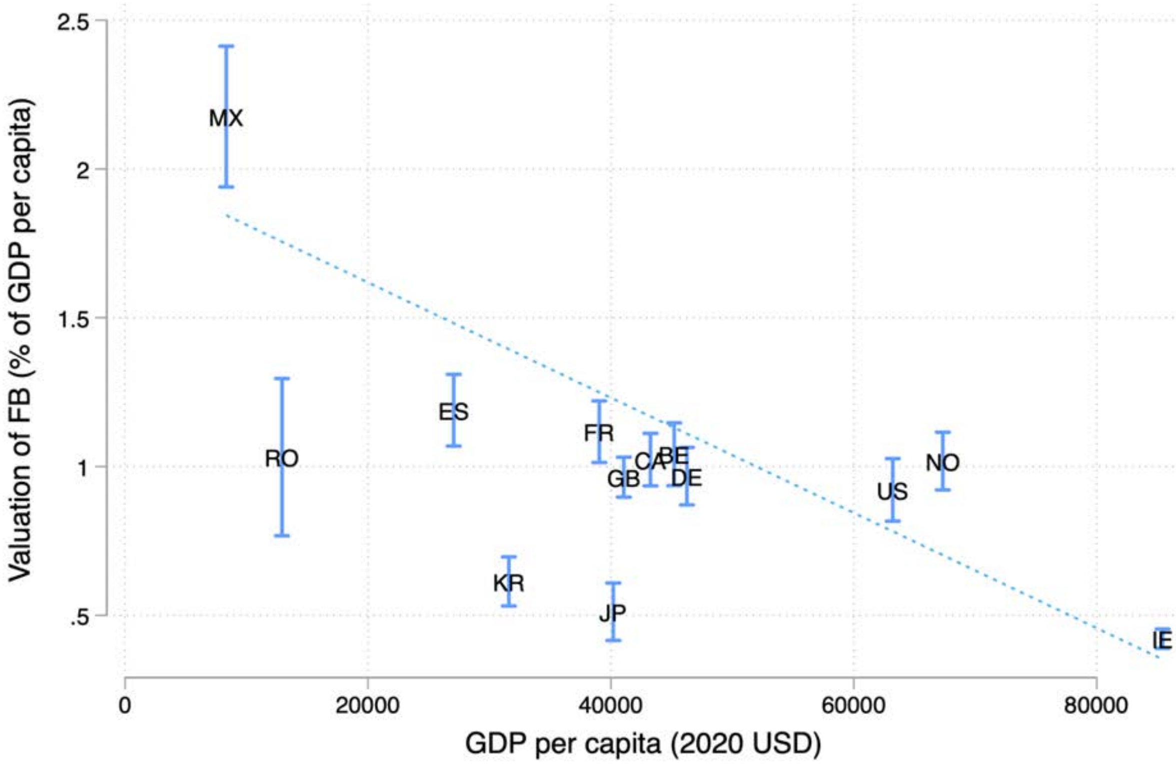
Appendix Figure 6: Median Value of Facebook by Education, Home Ownership & Gender



Appendix Figure 7: Facebook aggregate annual value by country from incentivized single binary discrete choice experiments

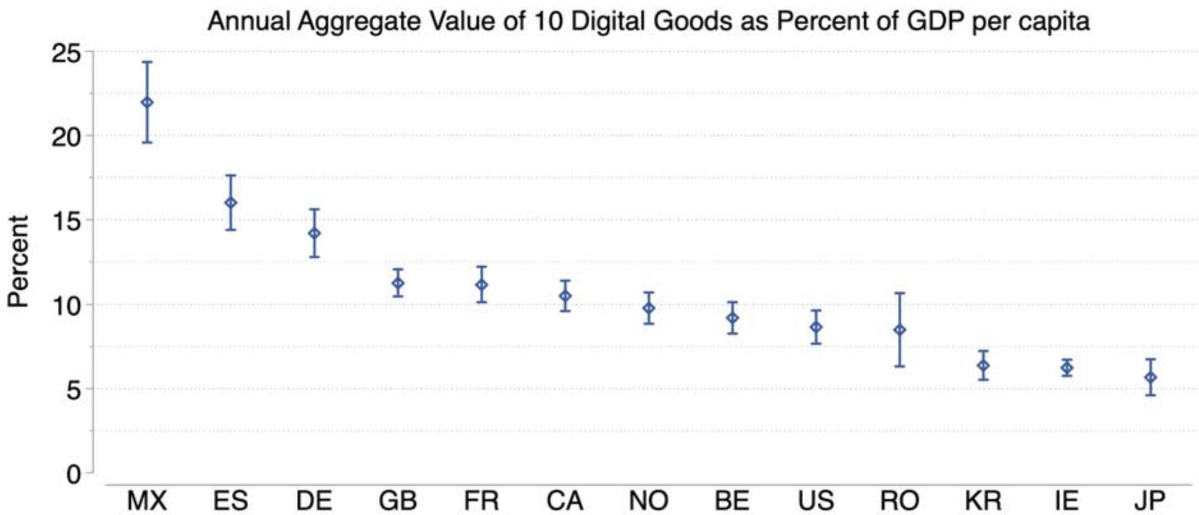


Appendix Figure 8: Valuation of Facebook as a share of GDP per capita

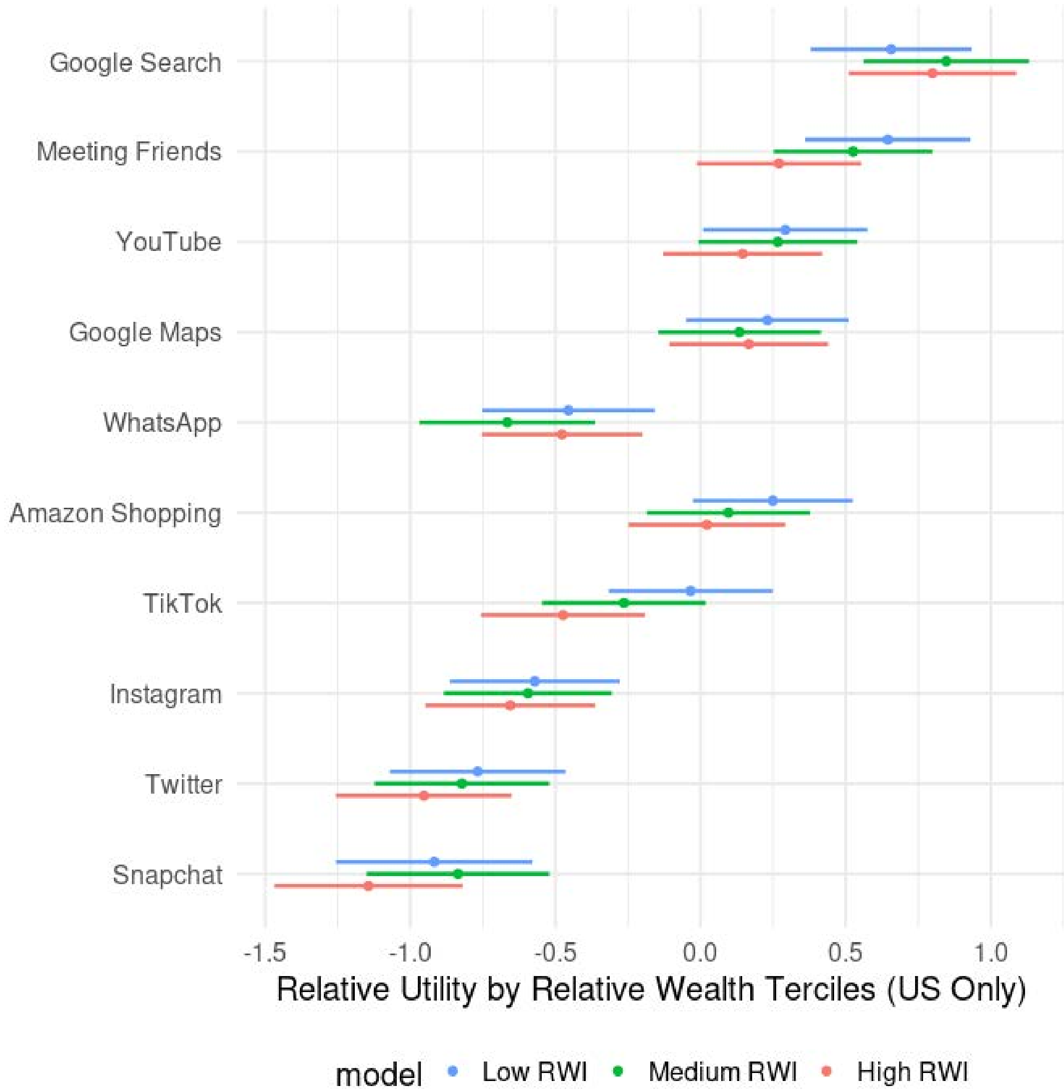


Appendix Figure 9: Association between GDP per capita and valuation of Facebook as a share of GDP per capita

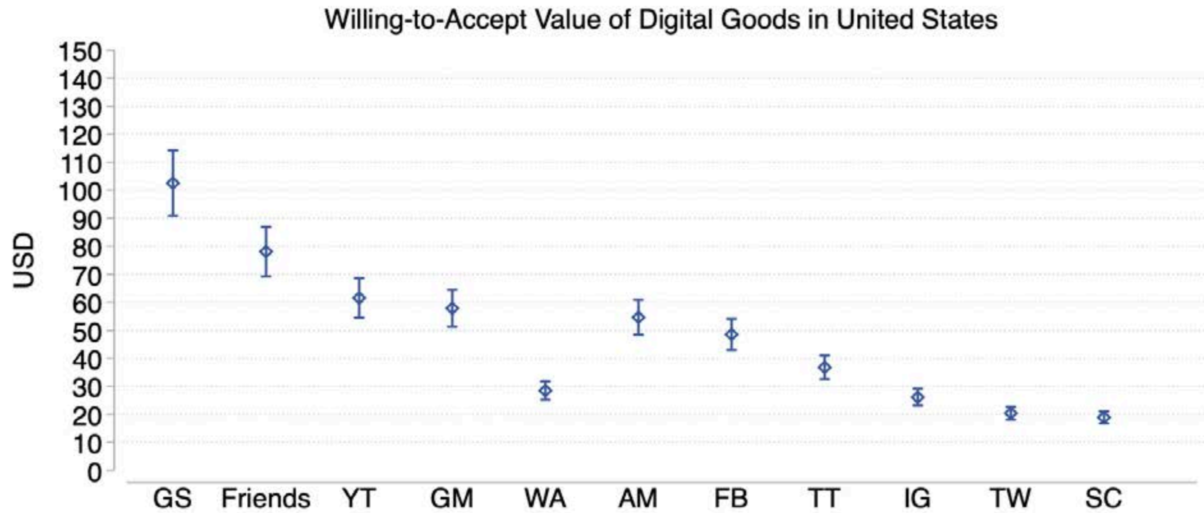
Appendix Figure 9 Notes: The figure depicts 2020 GDP per capita (in US dollars) in the X axis and user valuation of Facebook as a percent of 2020 GDP per capita in the Y axis. A weighted regression of Y on X (in \$10,000s) with weights given by Facebook monthly active users (MAU) in each country yields a point estimate of -.19 with a p-value equal to 0.001. The unweighted regression yields a point estimate of -.12 with a p-value equal to 0.028. The fitted lines in the plot correspond to the weighted regression.

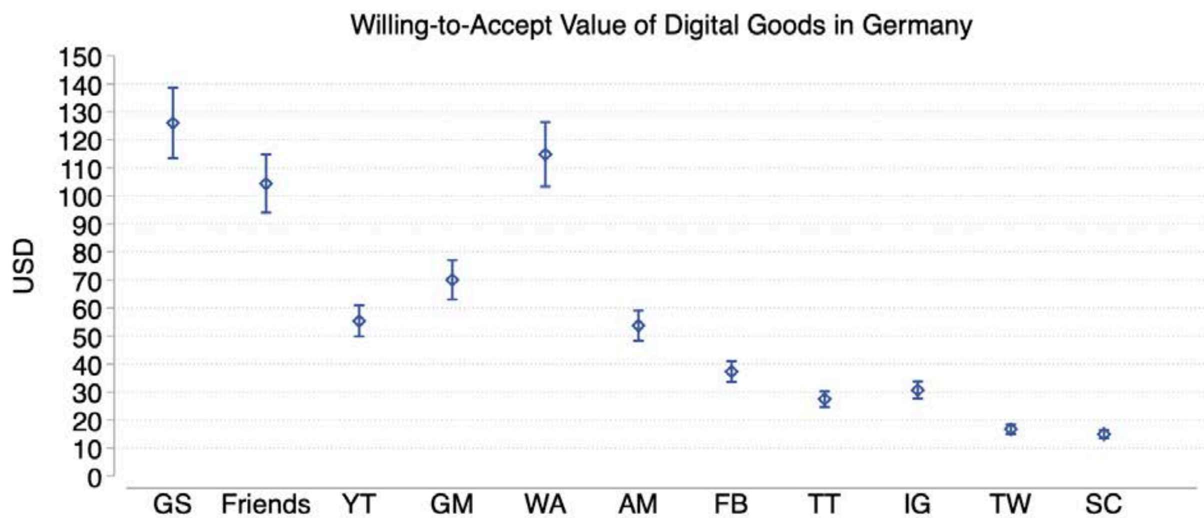


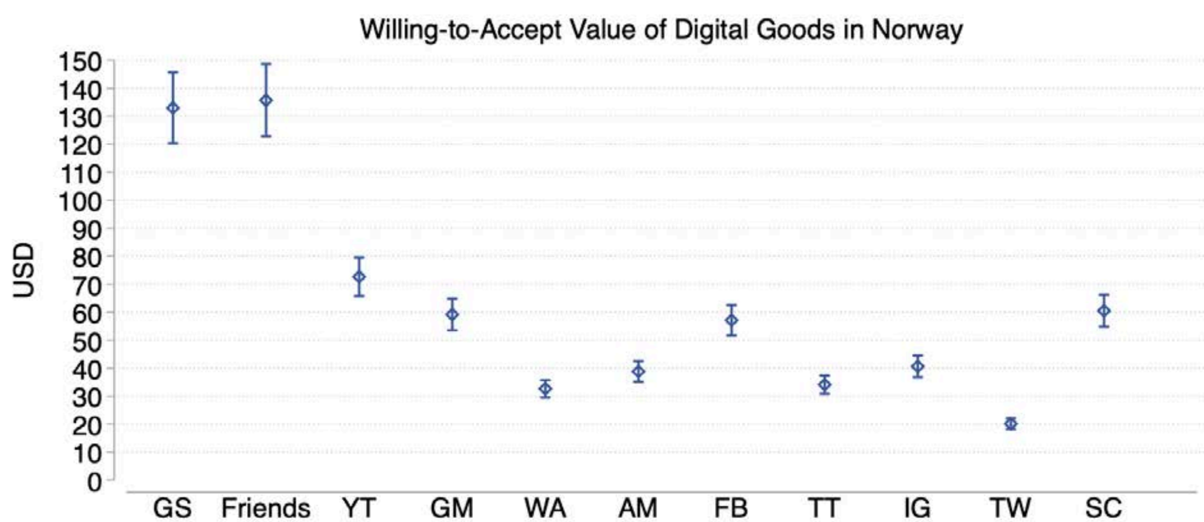
Appendix Figure 10: Aggregate annual value of 10 digital goods as share of GDP per capita



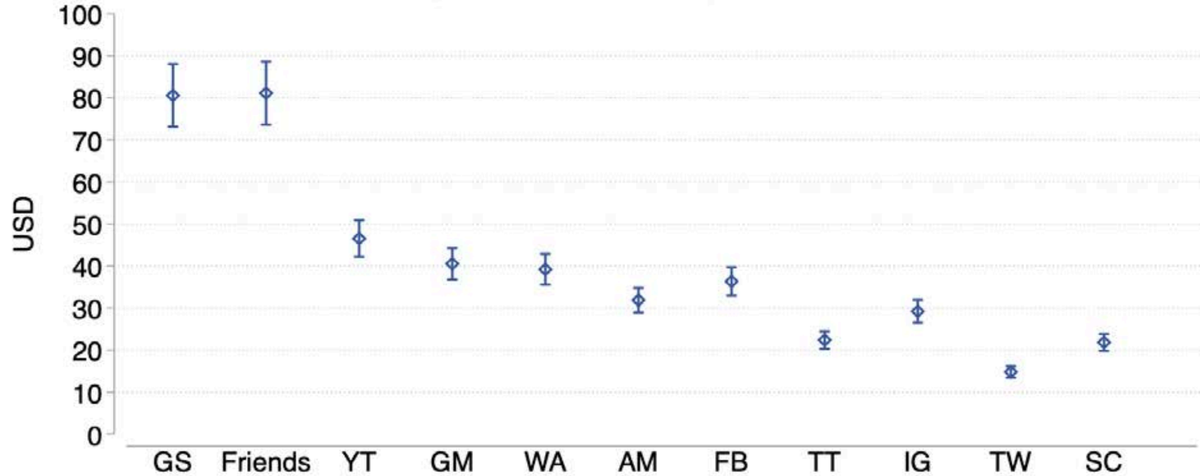
Appendix Figure 11: Relative utility by relative wealth index, with Facebook as the omitted category (US Only)


Willing-to-Accept Value of Digital Goods in United States
USD
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC

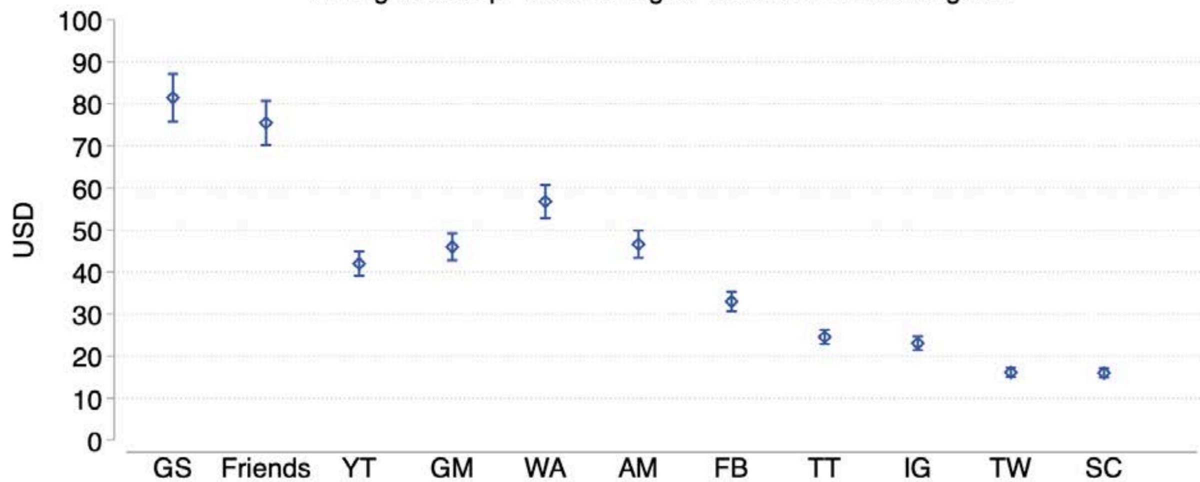

Willing-to-Accept Value of Digital Goods in Germany
USD
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC


Willing-to-Accept Value of Digital Goods in Norway
USD
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC

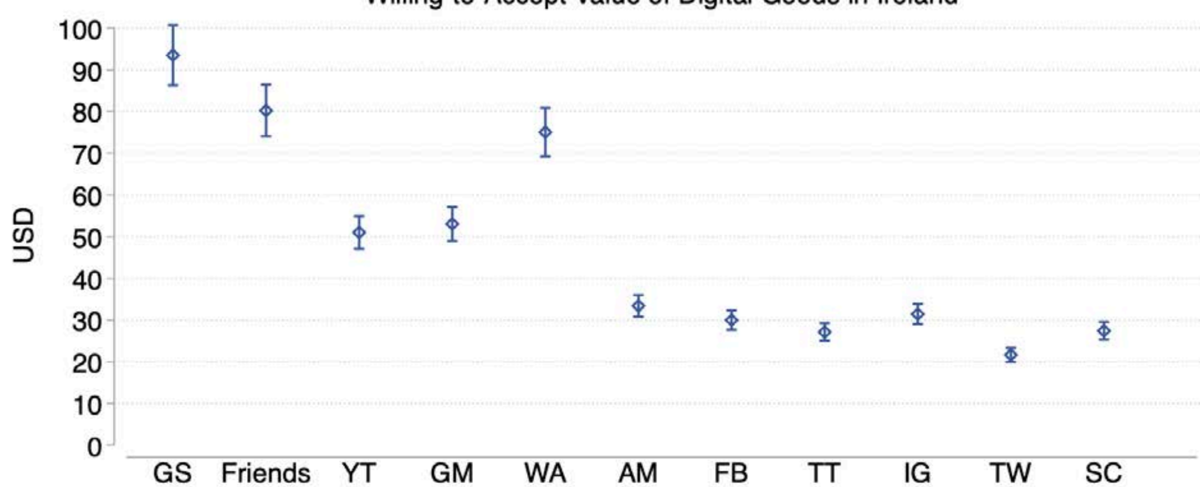
Willing-to-Accept Value of Digital Goods in France

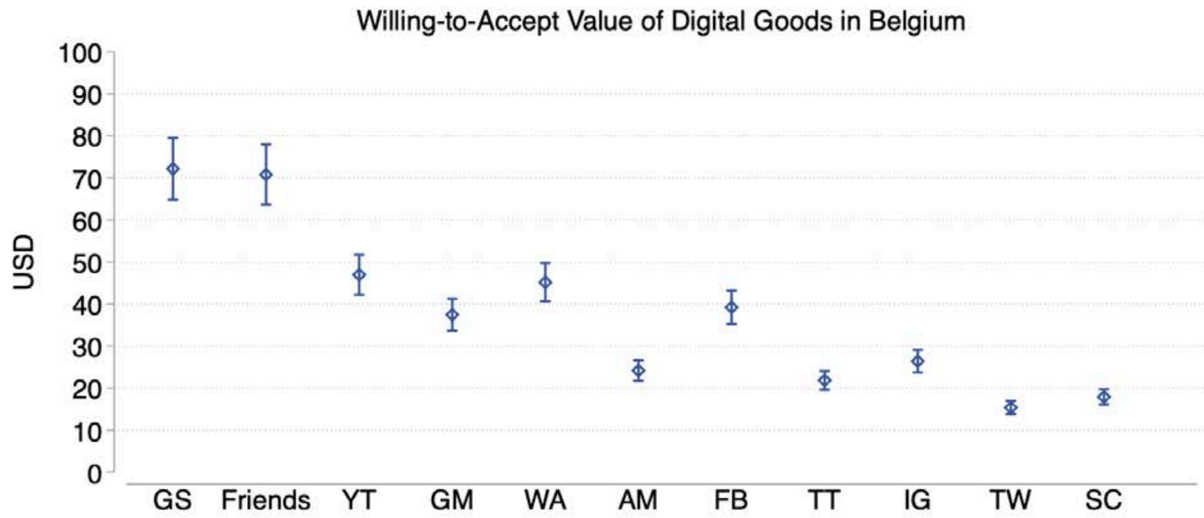


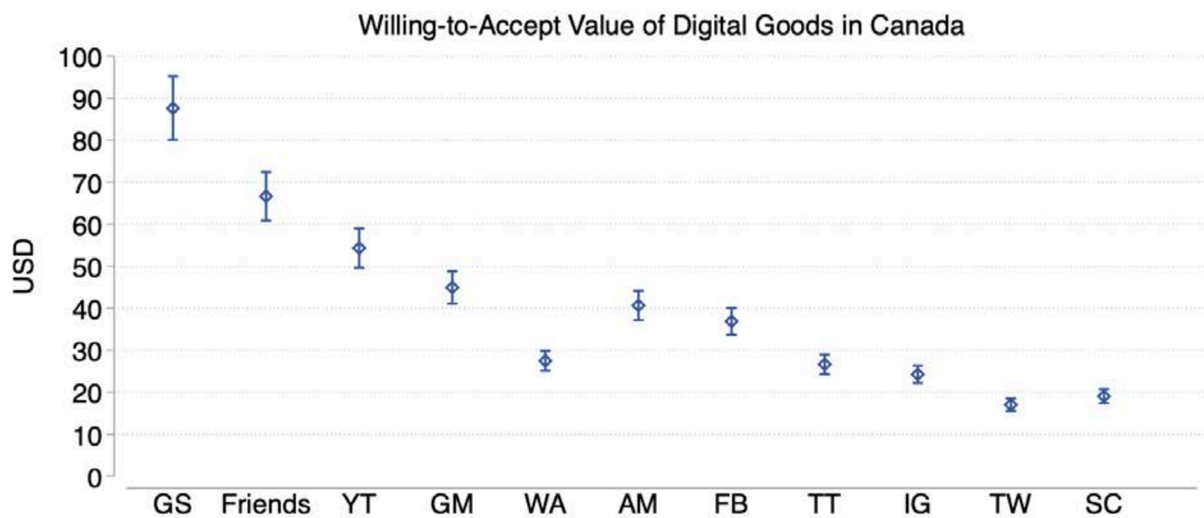
Willing-to-Accept Value of Digital Goods in United Kingdom

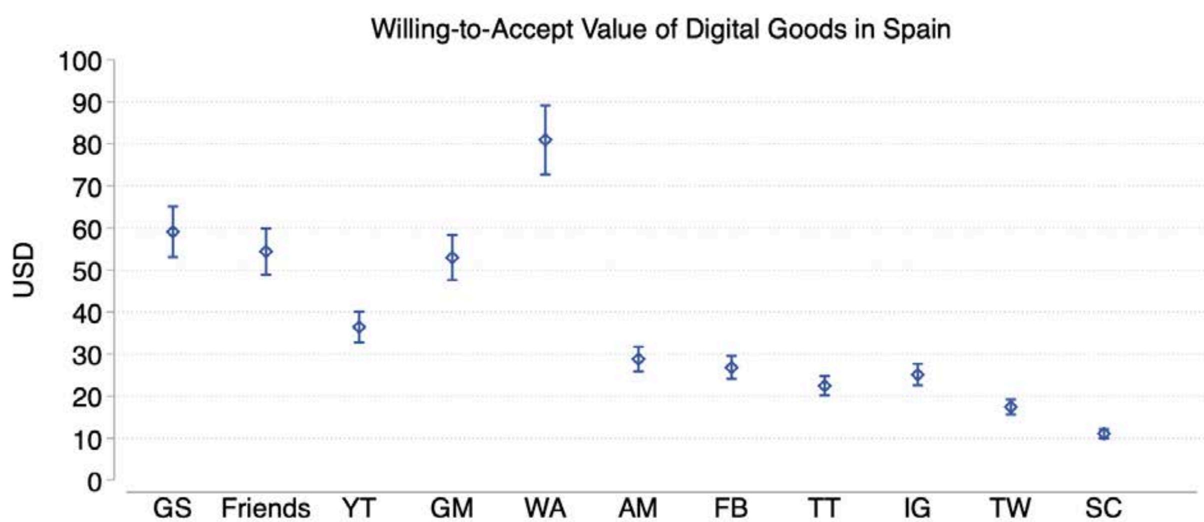


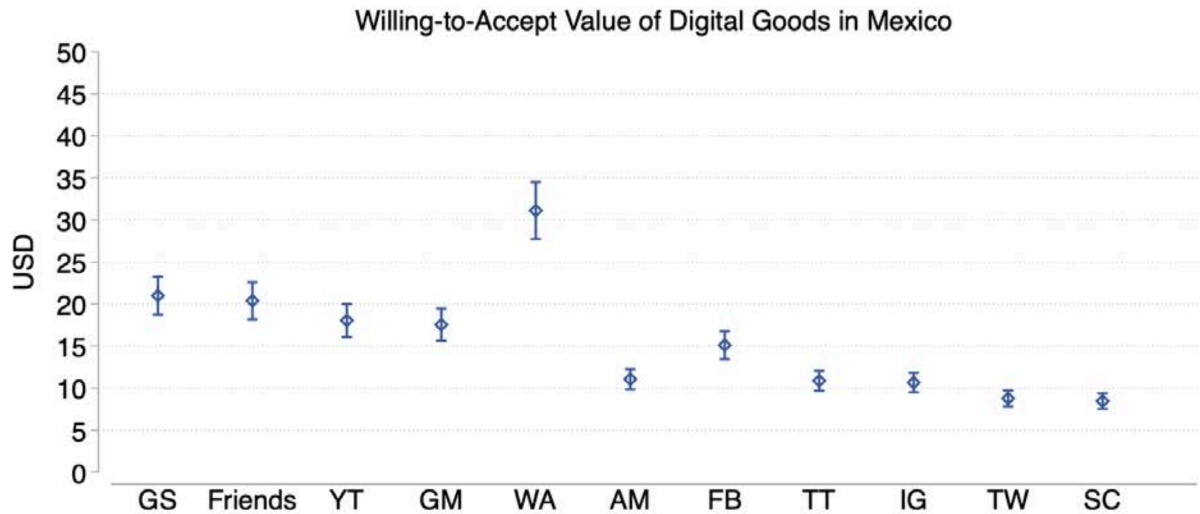
Willing-to-Accept Value of Digital Goods in Ireland

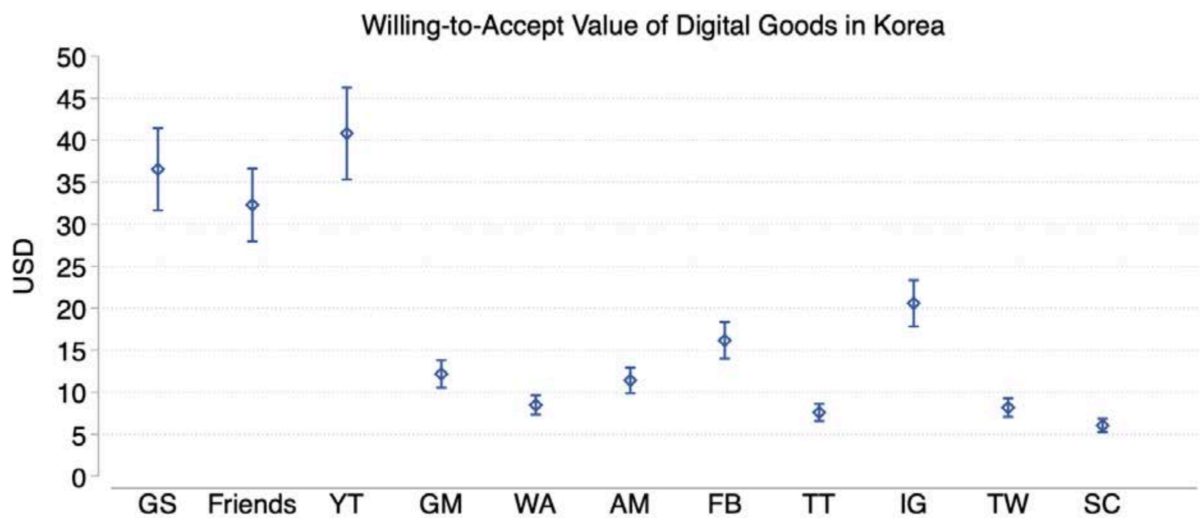


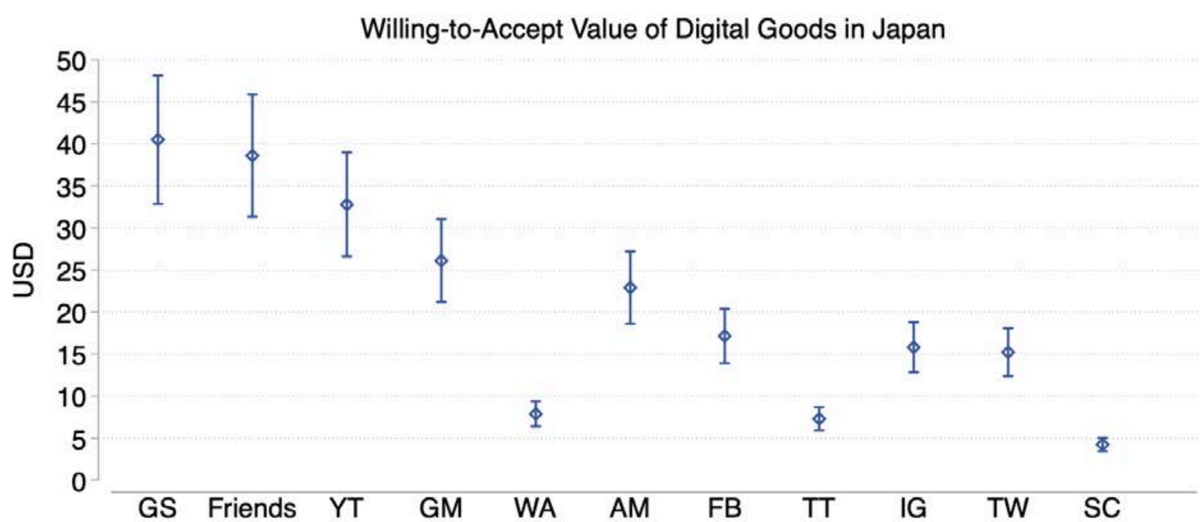

Willing-to-Accept Value of Digital Goods in Belgium
USD
0
10
20
30
40
50
60
70
80
90
100
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC

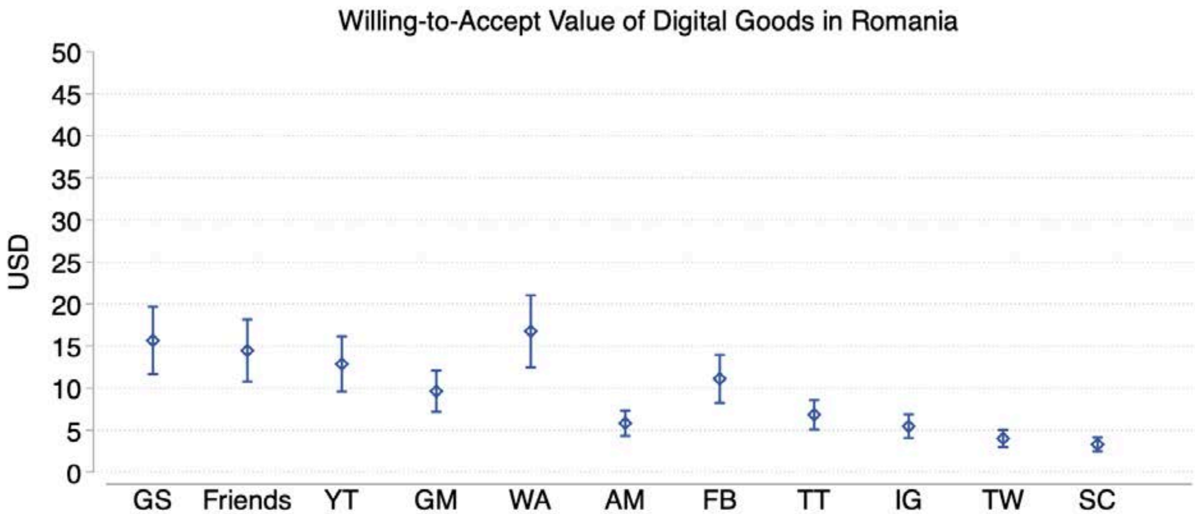

Willing-to-Accept Value of Digital Goods in Canada
USD
0
10
20
30
40
50
60
70
80
90
100
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC


Willing-to-Accept Value of Digital Goods in Spain
USD
0
10
20
30
40
50
60
70
80
90
100
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC

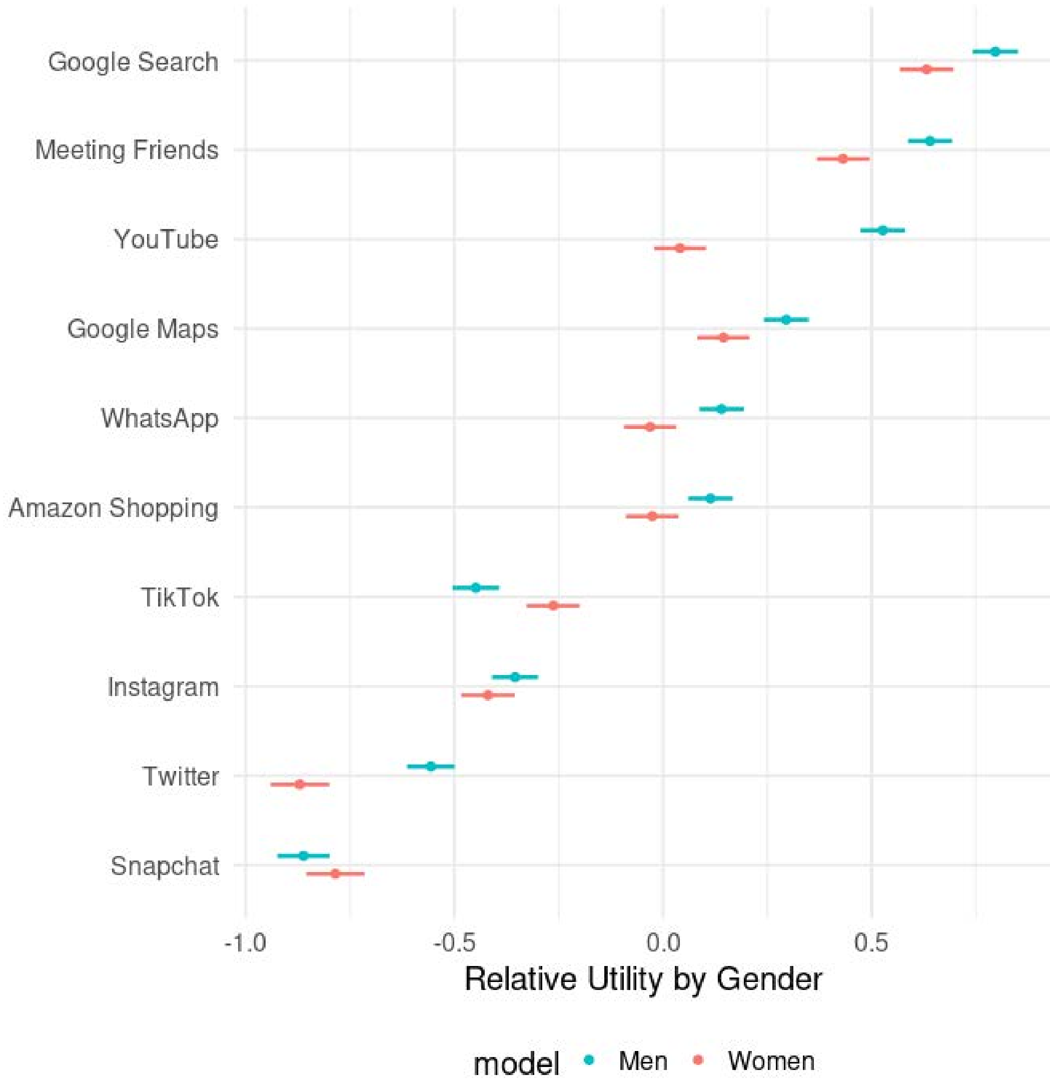

Willing-to-Accept Value of Digital Goods in Mexico
USD
50
45
40
35
30
25
20
15
10
5
0
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC


Willing-to-Accept Value of Digital Goods in Korea
USD
50
45
40
35
30
25
20
15
10
5
0
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC


Willing-to-Accept Value of Digital Goods in Japan
USD
50
45
40
35
30
25
20
15
10
5
0
GS
Friends
YT
GM
WA
AM
FB
TT
IG
TW
SC



Appendix Figure 12: Implied Median WTA Value of Each Digital Good by Country



Appendix Figure 13: Relative utility of digital goods by gender. Facebook is the omitted category. This figure should be interpreted in conjunction with Appendix Figure 6 which shows SBDC valuations of Facebook by gender