The welfare cost of imperfect consumer information: Evidence from a differentiated product market*

Imke Reimers
Cornell University and ZEW

Christoph Riedl
Northeastern University

Joel Waldfogel
University of Minnesota, NBER, ZEW

November 5, 2026

Abstract

Imperfect information can lead consumers to make purchases they regret and to forgo beneficial opportunities, but the associated welfare losses have largely escaped researchers' attention. We develop a framework in which the opportunity set is larger as the consumer has better pre-purchase information. Using unusual post-purchase video game usage data, we show that the benefit of most status quo usage could have been obtained with substantially less expenditure. A simple Cobb Douglas calibration delivers substantial consumer welfare losses relative to full information, which are robust to various alternative assumptions about prices paid for games and time that users would play unowned games.

*We thank conference and seminar participants at Cornell, Minnesota, Iowa State, Arizona, Berkeley, the International Industrial Organization Conference, the Munich Summer Institute, the Paris Workshop on Platforms and Recommender Systems, the ZEW ICT conference, Toulouse School of Economics, Sogang University, the 2025 AEA meetings, Berry Fest, Münster University, and Liverpool University. We also thank Ahmad Ahmadi, Avery Haviv, David Liu, Xuan Teng, Roland Rathelot, Luca Sandrini, and John Horton for providing insightful discussant comments.

Introduction

The availability of a wide variety of differentiated products delivers substantial benefits to diverse consumers. Yet, the varied features of products – and varied tastes of consumers – can make it hard for consumers to know which products they might like. If consumers had full information prior to purchase, then their ex post valuations would match their ex ante expectations. No purchases would lead to regret, and no beneficial purchases would be missed, so that purchase decisions would deliver all of the consumer surplus potentially available. With imperfect information, by contrast, some current purchases may disappoint their buyers while some worthwhile options may be missed; and realized consumer surplus will fall short of potential. Exploring this shortfall – the deviation between potential consumer surplus and actual surplus – is the goal of the paper.

While a large literature on differentiated products examines the supply-side inefficiency of entry and product availability, welfare shortfalls on the demand side have received less attention, even though imperfect consumer information is assumed to be widespread.[1] A substantial body of research explores theoretical consequences of consumers' imperfect information, and a sizable empirical literature measures impacts of improvements in pre-purchase information on the incremental consumer surplus (CS) realized by purchases.[2] Yet, the magnitude of surplus forgone relative to full information in current consumption remains unexplored.

The absence of attention to this topic arises for a self-reinforcing combination of theoretical and empirical reasons. Researchers can typically only see purchase decisions, which are understood as instances in which consumers' valuations exceed prices. Researchers have

[1] See, for example, Mankiw and Whinston (1986), Dixit and Stiglitz (1977), Spence (1976), and Anderson et al. (1995).

[2] Central theory papers on the role of information include Nelson (1970), Weitzman (1979), and Stigler (1961). The empirical literature on the welfare increment from better pre-purchase information includes Allcott (2013), Train (2015), and Reimers and Waldfogel (2021). See also Section 1.1 below.

relatively few opportunities to see post-purchase usage that might raise the possibility of disappointment and regret.

This paper offers a simple, implementable framework for analyzing the welfare forgone by consumers because of imperfect pre-purchase information. We summarize consumers' choice opportunities with a budget constraint whose shape depends on the quality of pre-purchase information. Better-informed consumers face a more "bowed-out" budget constraint allowing greater consumption – and consumer surplus – with the same resources. Because the utility of time spent may differ across games, we measure consumption with weighted hours, where the weights are inferred from purchase behavior. We employ this framework to explore the welfare lost to imperfect information using unusually rich data on post-purchase usage of video games. Not only do we observe which consumers purchased each of the products, we also see the amount of time they spent using them and, for a subset, whether they recommend them. The data, from the video game platform Steam, include information on 50,000 consumers' cumulative usage of 100 games.

We undertake two kinds of empirical exercises. First, we present descriptive evidence suggesting welfare losses due to imperfectly informed status quo decision making. Using only data on owned games, we document that consumers commonly purchase games they end up playing very little; and we offer evidence that these choices are regretted. Still more gains, beyond those from avoiding regret, would be possible if consumers were aware of products they did not purchase but would have enjoyed. We use data on playtimes for owned games, and matrix factorization predictions (along with assumed prediction errors), to create measures of playtime for each remaining user-game combination. Consumers could experience large savings by avoiding disappointing status quo purchases. The vast majority (90 percent) of an average consumer's status quo playtime could be achieved with less than half (38 percent) of status quo expenditure. Allowing for both avoiding regret and the discovery of previously-missed opportunities, fully informed consumers could achieve a 97

percent increase in realized playtime with current expenditure.[3] The marginal utility of hours may differ across games, but when we transform hours to the weighted hour index that enters our utility model, the gains remain large: 90 percent of the status quo value of playtime could be had for about a fifth of status quo expenditure, and full information would raise the value of playtime at current expenditure by 45 percent.

In our second empirical analysis, we implement our framework directly, calibrating a model of choice in the face of a budget constraint of all other goods against weighted playtimes, which is expanded by information. Relative to the status quo, full information raises average CS by about 130 percent of status quo expenditure while reducing expenditure by three fifths. About 44 percent of full information's effect comes from avoiding regrettable purchases; the remainder arises from taking advantage of otherwise-missed opportunities. The welfare shortfall is robust to allowing much lower prices for frequently-discounted or little-played games, as well as the possibility that non-purchased games are worse than we predict them to be. Discounting the modeled playtime of unowned games by half leaves more than three quarters of the estimated gain. The gains also remain large when allowing for heterogeneity in consumer tastes for playtime of different games.

The paper proceeds in six sections. Section 1 discusses the relevant academic literatures and the product-market context; we also provide contextual evidence of ex post regret. Section 2 introduces our tractable model of the effect of pre-purchase information on a consumer's choice among product bundles. We show how greater information expands the budget set, allowing choices that deliver higher utility, which is defined over weighted hours and all other goods. Section 3 describes the data used in this study, which include not only the games that each user has but also the time they have spent using each game. We also discuss our models of the time users would play games they do not own and our

[3]The estimated effects depend in part on the accuracy of our playtime predictions for unowned games; our estimates remain large in a variety of robustness analyses, including discounting the modeled playtime of unowned games by up to half.

evidence that consumers who spend less time playing a game are more likely to regret the purchase. Section 4 presents descriptive evidence suggestive of large welfare gains from full pre-purchase information, in both raw and weighted hours, as well as in analyses showing robustness to the measurement of prices and to the modeled playtime of unowned games. Section 5 presents estimates of the welfare effects of full information using our framework, along with the component stemming from reduced regret. We also document the robustness of the estimates to various measurement assumptions. We conclude in Section 6.

1 Background

This section provides three types of background information. Section 1.1 discusses the academic literatures relevant to this study. Section 1.2 presents information about the video game industry; and Section 1.3 discusses existing evidence on regret, both generally and in video game purchases.

1.1 Relevant literatures

Our broad question – about the effect of pre-purchase information on the welfare consequences of differentiated products markets – is related to four literatures. First, our work is related to theoretical studies of the effect of information on the welfare properties of market outcomes (Nelson, 1970; Weitzman, 1979; Stigler, 1961), including studies of inefficiencies arising in differentiated products markets (Spence, 1976; Dixit and Stiglitz, 1977; Mankiw and Whinston, 1986; Anderson et al., 1995). The literature has traditionally focused on supply-side challenges associated with the number of entering products; we are instead concerned with the demand-side efficiency consequences of information given the available products.

Second, a growing literature quantifies the effects, as well as welfare benefits, of im-

proved pre-purchase information (Reinstein and Snyder, 2005; Chevalier and Mayzlin, 2006; Chernew et al., 2008; Reimers and Waldfogel, 2021; Villas-Boas et al., 2021; Fang, 2022; Compiani et al., 2024). The exercise in Bronnenberg et al. (2015) is closer to the present study: they compare drug choices of poorly informed lay consumers with those of well-informed pharmacists. Our exercise differs in asking how consumption and surplus would change if consumers knew their own realized values before purchase, and we are able to quantify both surplus forgone from regrettable choices as well as surplus lost from failing to buy valuable products. A related literature documents purchase errors, including apparent ex post mistakes (Miravete, 2003; Allcott, 2013) and explicit evidence of regret (Skelton and Allwood, 2017; Einav et al., 2023).

Third, we draw on the literature on recommender systems (Koren et al., 2009; Bobadilla et al., 2013; Koren et al., 2021; Lee and Hosanagar, 2021) to characterize the playtime that unowned games would deliver. Specifically, we use a collaborative filter approach that uses post-purchase usage for observed consumer-product pairs to estimate usage of consumer-product pairs that are not observed. The matrix factorization approach we use is representative of prediction approaches used in practice (Koren et al., 2021); and it has outperformed more complex functions such as neural networks (Rendle et al., 2020).

Finally, there are substantial literatures on various aspects of video games, including the complementarity between consoles and games (Lee, 2013), the potential impacts of video games on social outcomes (Ward, 2010), and the relationship between work hours and video games (Aguiar et al., 2021).

1.2 Industry context

Video games attract substantial amounts of entertainment spending, as well as time use. The video game industry generated $347 billion in worldwide revenue in 2022, making it

substantially larger than the movie and music industries combined.[4] The US Bureau of Labor Statistics reports that Americans spent an average of 34.2 minutes per day playing video games during 2022.[5] Between 2014 and 2017, US men between the ages of 21 and 30 spent an average of 3.9 hours per week playing video games (Aguiar et al., 2021). US consumers spent $47.5 billion on video game content during 2022 while consumer spending overall was $9.8 trillion.[6] Hence, video games accounted for 0.49 percent of household spending, which we use to inform our expenditure share estimate in the Cobb Douglas analyses.

Video games are played on game consoles (such as the Nintendo Switch or Sony PlayStation), on phones, or on computers, where games are downloaded from digital video game distribution platforms. One of the largest such platforms, Steam, provides the setting for our analysis. Operated by Valve Corporation and founded in 2003, Steam had 33 million concurrent peak users during 2023.[7] Steam offers over 73,000 games, and revenue from game sales on Steam was $8.8 billion in 2022, about 20 percent of total US spending on video games.[8]

1.3 Evidence of ex post regret

Despite the paucity of direct evidence on consumers' ex post valuations of the products they purchase, there is evidence of regret across a variety of markets. Skelton and Allwood (2017) reports the shares of British consumers reporting regretted purchases in each of 20 product categories. They find regret in large shares of purchases. For example, 38 and 16

[4]See `https://www.statista.com/topics/868/video-games/topicOverview`. Global recorded music revenue was $31.2 billion in 2022. See `https://www.statista.com/statistics/272305/global-revenue-of-the-music-industry/`, while global movie revenue was estimated at $93.4 billion. See `https://www.ibisworld.com/global/market-size/global-movie-production-distribution/`. Global movie box office alone was $26 billion. See `https://www.imdb.com/news/ni63899899/`.

[5]See `https://www.statista.com/statistics/502149/average-daily-time-playing-games-and-using-computer-us-by-age/`.

[6]See `https://www.statista.com/statistics/252457/consumer-spending-on-video-games-in-the-us/` and `https://www.bls.gov/opub/reports/consumer-expenditures/2022/home.htm`.

[7]See `https://www.statista.com/topics/4282/steam/topicOverview`.

[8]See `https://www.statista.com/statistics/547025/steam-game-sales-revenue/`.

percent of respondents, respectively, report regretting clothing and electronics purchases in the past year. The least-regretted category is milk, whose consumers have presumably bought the product before and have good ex ante information. The pattern of results is strongly suggestive that the CS based on the ex ante valuations motivating buying behavior would understate the potential CS available, particularly for products not purchased repeatedly.[9]

Regret is also common for video game purchases. The large number of games available on the Steam platform makes it difficult for consumers to know which products they might find appealing. Social media sites feature discussions of games that consumers regret buying. A Reddit thread entitled "What's one game you regret buying?" elicited 1,700 comments. The top (most upvoted) reply was, "Probably 70% of my steam library."[10] Similar comments are shared at Quora; and many YouTube videos describe games that users regret buying.[11] Together, those comments indicate that purchase errors are not only possible but common in this context.

User recommendations provide additional, and more systematic, evidence on varying post-purchase reactions to games. Steam users leaving feedback can "recommend" or "not recommend" a game; and negative recommendations – which suggest regret – are fairly common. For example, in data that we describe below, we see 58,837 recommendations from 20,985 users, and 7,240 of the recommendations are negative. We show in Section 3.3 that negative recommendations are significantly more likely when games deliver less playtime per dollar spent, suggesting both that consumers value the time that games deliver and that game playtime, relative to price, is relevant to consumers' satisfaction from their game

[9]There is evidence of regret in other contexts as well. See `https://stradaeducation.org/value/do-you-regret-your-college-choices/` on education, `https://realestate.usnews.com/real-estate/articles/have-buyers-remorse-with-your-home-heres-what-to-do` on housing, and `https://www.forbes.com/sites/bryanrobinson/2022/12/01/5-reasons-for-boomerang-employees-and-the-great-regret-in-employment/` on labor markets.

[10]See `https://www.reddit.com/r/gaming/comments/12frdsr/whats_one_game_you_regret_buying/`.

[11]See, for example, "Video Games I Regret Buying" (`https://www.youtube.com/watch?v=l4f8CmcLJPk`) or identically titled video (`https://www.youtube.com/watch?v=WoB80aEQsnE&t=16s`).

purchases.

2 Theory

We are interested in analyzing how better pre-purchase information would affect consumers' choices among high-dimensional bundles of products. In our context, consumers are choosing which bundles, from among 100 games, to own. These decisions depend on their utility functions and their information about games. We discuss each in turn below.

2.1 Consumer utility

In general, consumers have some utility over the bundle of games, less the utility of money paid for the games:

$$u_{ij} = U\left(\mathbf{1}_{i1}, \ldots, \mathbf{1}_{iJ}\,;\, A_{0i} - \sum_{j \in J} p_j \mathbf{1}_{ij}\right),$$

where A_0 is the money the consumer has available for either games or all other goods, $\mathbf{1}_{ij}$ is an indicator that is 1 if individual i owns game j, J is the full set of available games, and $P_i = \sum_{j \in J} p_j \mathbf{1}_{ij}$ is the spending required to purchase the bundle of owned games. Tractability requires reducing the dimensionality of the problem (Chu et al., 2011; Crawford and Yurukoglu, 2012). One approach would be to assume that users derive utility from the raw sum of the hours of play time across games. Yet, users may derive different marginal utility of hours across games and, moreover, that these marginal utilities may differ across users. Accordingly, we define an index of consumer i's weighted hours: $W_i = \sum_{j \in J} \omega_{ij} h_{ij} \mathbf{1}_{ij}$, where h_{ij} denotes the amount of time that consumer i would use product j, ω_{ij} is a utility weight per usage time, and $\mathbf{1}_{ij}$ indicates whether consumer i purchased product j. Then,

$$u_{ij} = U\left(W_i;\, A_{0i} - \sum_{j \in J} p_j \mathbf{1}_{ij}\right). \tag{1}$$

This approach requires the index W which, in turn, requires the marginal utility weights ω_{ij} for transforming raw hours. We infer the game-specific values of an hour from purchase behavior in Section 4.1, and we allow these weights to differ across consumer groups in Section 4.4.

2.2 Pre-purchase information and the opportunity set

Given what a consumer knows about products prior to purchase, the consumer faces a budget constraint describing how expenditure delivers weighted playtime; and this relationship depends on the consumer's level of pre-purchase information. To derive the budget constraints, it is helpful to begin with two extreme cases, one in which consumers have no information about individual games and another in which they have full information. In the "no information" case, consumers know the average amount of weighted playtime they obtain per dollar spent, which we term ρ_i, but not the particular realized value for each game. Then a consumer spending a total of P_i would expect to receive $W_i = \rho_i P_i$ weighted hours of playtime, and they would rank products randomly. We illustrate a budget constraint resulting from such random rankings (which we term r^ϵ) in the dashed line in Figure 1, which plots the cumulative amount of weighted playtime W (x-axis) against money available for all other goods (y-axis). The consumer would face a linear expected budget constraint, with the same expected amount of money per weighted hour $1/\rho_i$ for each purchased game. While each budget constraint realization depends on the random rank order draw (r^ϵ), the budget constraints will be linear in expectation across draws.

At the other extreme, consumers have full information on the weighted hours they would play each game ($\omega_{ij} h_{ij}^T$). When armed with full information, consumer i knows that game j would deliver $\omega_{ij} h_{ij}^T$ of weighted playtime at a price of p_j. Ordering products by ascending values of $\frac{p_j}{\omega_{ij} h_{ij}^T}$, a ranking we term r^T, delivers the maximally expansive budget constraint for the consumer. The outer budget constraint in Figure 1, drawn curved to reflect a contin-

uous approximation, represents this full-information case. The shape of the consumer's full information budget constraint depends on the variability of their price per weighted hour of playtime across games, $\frac{p_j}{\omega_{ij} h_{ij}^T}$. If all games delivered the same weighted hours of playtime per dollar spent, then the full information budget constraint would be linear and indistinguishable from the no-information budget constraints. The greater the variance in a user's price per weighted hour across games, the more bowed out is their full information budget constraint.

Any ranking of products based on something other than the true $\frac{p_j}{\omega_{ij} h_{ij}^T}$ shrinks the opportunity set relative to the full information case. For example, if consumers had access to less-than-full pre-purchase information, they would not know $\omega_{ij} h_{ij}^T$ in advance but rather a prediction containing error. A consumer would then rank-order games by the price per weighted hour of use that they expected at purchase, and realized weighted playtime would ensue. Provided that the consumer's information is better than random, this delivers a realized budget constraint that lies somewhere between linearity (ignorance) and full information; and the inner, curved budget constraint in Figure 1 illustrates such a scenario. The deviation of the consumer's budget constraint from the full-information budget constraint is larger, the less accurate their information.

2.3 Status quo information and utility

Consumers in the status quo could have a range of possible budget constraints that fall short of full information. A ranking according to r^T maximizes the weighted playtime that each expenditure delivers, whereas a random ranking (r^ϵ) produces a linear budget constraint in expectation. A consumer's information could even be "worse than random" in the sense that their ranking could be negatively correlated with r^T.[12] To accommodate the possibility that

[12] At an extreme, a consumer who ranks products according to $-r^T$ would attain minimum weighted playtime with any level of expenditure.

status quo information might vary from full information to worse than random, we create an index that weights the random and full information rankings via

$$I(\kappa_i) = \kappa_i r^T + (1 - |\kappa_i|) r^\epsilon, \tag{2}$$

where $\kappa \in [-1, 1]$, and the consumer's resulting ranking r^I is in order of $I(\kappa_i)$.

The shape of a consumer's status quo budget constraint depends on the size and sign of κ_i. If the consumer possessed full information in the status quo, then $\kappa_i = 1$; and the consumer's status quo budget constraint would lie on the full information budget constraint. If the status quo consumer has better-than-random but less-than-full information, then $0 < \kappa_i < 1$ and the average status quo budget constraint is less bowed out than the full information budget constraint. If $\kappa_i = 0$, the consumer's budget constraint is linear with a slope based on the average price per weighted hour across games, as depicted by the dashed linear budget constraint in Figure 1. Finally, if $-1 \leq \kappa_i < 0$, then the consumer's knowledge gives rise to a budget constraint that is "bowed in" toward the origin. If consumers have access to information in the status quo, then $\kappa > 0$; and we find below that consumers have better-than-random information in the status quo.

Given the budget constraint associated with their pre-purchase information, the consumer chooses a point such that their marginal rate of substitution (MRS) equals the slope of the expected budget constraint. Figure 1 illustrates the utility-maximizing weighted hours choices with full information (Point B) and with a less informed budget constraint (Point A). In what follows, we estimate κ_i for each user to develop empirical characterizations of the status quo and full information budget constraints, along with models of utility that deliver both status quo and counterfactual choices and their welfare effects.

3 Data and playtime measures

3.1 Data

The main data for this study include information on 50,000 Steam users and 100 popular Steam games. For each user, we observe which of these games they own, as well as the cumulative number of hours they have spent playing each of the games, as of our data collection between May 9 and 16 of 2021.[13] The underlying data include 192,137 users, who collectively owned 33,844 distinct games. From these, we chose the 100 most popular games (by number of users who bought the game) with positive prices. To ensure sufficient playtime data observations, we restrict attention to users who purchased at least 20 of those popular games; and we randomly selected 50,000.

We collected historical price data for each game from `https://steampricehistory.com` on November 15, 2023, and we construct game prices by averaging across each game's daily prices. We also use the 25^{th} and 10^{th} percentiles of each game's price across days for robustness analyses in Sections 4.4 and 5.4 that allow more frequently discounted games to have been purchased at lower prices. Using our baseline price measure, prices average \$19.47 across the 100 games. Our users have spent an average of \$685.8 and own an average of 35.71 games.[14] These games provide them with 2,144.9 hours of cumulative playtime; and users play each owned game an average of 60.1 hours.

We also observe individual reviews left by 20,985 of our users, for a total of 58,837 instances of "recommended" or "not recommended" feedback on the Steam platform.[15] We

[13] See `https://developer.valvesoftware.com/wiki/Steam_Web_API`, which we used to obtain lists of owned games and their playtime for players with publicly visible profiles.

[14] While Steam allows users a two-week window to return games played less than 120 cumulative minutes, the games we observe in users' libraries are games they still possess long after the return window closed. Still, for the sake of conservatism, we undertake robustness analyses below in which games played for less than 30, 60, or 120 minutes are assumed to have been obtained without payment.

[15] We collected these data from `https://steamcommunity.com/profiles/[steamid]/recommended` during June of 2024.

use these data, in Section 3.3, to validate our assumption that users who play games for less time, in relation to the price they pay, are more likely to regret those purchases. Finally, we observe user join years for robustness checks and information from `howlongtobeat.com` for validating average playtime.

3.2 Modeling true playtime

Answering our research questions requires measures of true, realized playtime h_{ij}^T for each user and game, including those not currently owned. This section discusses how we create measures of true playtime by first using matrix factorization to obtain playtime predictions (h_{ij}^P, in Section 3.2.1), and then adding errors to mimic realizations, as explained in Section 3.2.2.

3.2.1 Predictions of playtime from matrix factorization (h_{ij}^P)

Our users own about a third of the 100 games on average, so the data matrix for generating predictions is sparse.[16] One can imagine a variety of approaches to filling in the prediction matrix. Our preferred prediction approach is to create a collaborative filter using matrix factorization, and we show below that our approach outperforms other prediction techniques.

Our implementation of the collaborative filter follows the approach of Koren et al. (2009). We employ a matrix factorization model that maps both users and products to a joint latent factor space with $k = 100$ dimensions. Each product j is associated with a vector $m_j \in \mathbb{R}^k$, and each user i is associated with a vector $n_i \in \mathbb{R}^k$. The elements in m_j (n_i) capture the extent to which a product (user) possesses those latent factors. We estimate the fitted values of the log playtime that a user would derive from a given game from the inner products in that latent space, $\hat{h}_{ij} = m_j^T n_i$.

[16] This has been formulated as a matrix completion problem, in which missing elements of the user-item matrix have to be predicted from limited historical data as not every user has interacted with every item (Jannach et al., 2016).

Estimation of the predicted playtime values proceeds in two steps. First, we tune the hyperparameters of our prediction model to minimize the regularized squared error between observed hours and the product $m_j^T n_i$ on the set of owned games, without overfitting. These hyperparameters are the number of iterative model fitting steps, the step size by which parameters change, and the penalty for parameters that are very large or very small.[17] We tune the hyperparameters on a training set that includes all but one owned game per user, which we save for the validation set.[18] From this validation set, we further reserve 10 percent of observations as a final test set. All reported test-set fit statistics are computed on data that are not used at any stage of hyperparameter tuning process.

In the second step, we refit the model using the hyperparameters with the lowest validation-set error from above. We do this ten times, holding out a different tenth of the sample each time. This gives us out-of-sample playtime predictions for every owned and unowned user-game combination.

Our latent factor estimation uses stochastic gradient descent optimization (Funk, 2006). Although the modeling approach includes no observable characteristics of games nor users, the large number of latent factors implicitly captures these types of variation. In the video game context, the 100 latent factors we estimate reflect game and user characteristics, such as demographics and tastes. Because they are based on users' game usage, rather than observed user or game characteristics, they might also capture less well-defined dimensions such as pace of game play (real time vs. round based), the visual style of the game (realistic vs. cartoon), or dimensions that cannot be interpreted at all.

How well do our out-of-sample predictions perform? Table 1 shows the root mean squared errors (RMSE) from a sequence of prediction approaches, from the global average (using the

[17] We fine-tune the model's hyperparameters by independent uniform random sampling of parameter values (Bergstra and Bengio, 2012).

[18] As is typical in the recommender literature, we perform leave-one-out cross validation. Because there is no natural order to individuals' interactions with products, we selected one game at random for each individual as validation data.

average value of playtime across users and games as the common prediction) to matrix factorization with 100 factors.[19] The RMSE measures for the out-of-sample validation sets run from 0.722 with the global average to 0.614 with our preferred, "sophisticated" prediction approach. Our preferred approach substantially outperforms the alternative techniques.[20]

3.2.2 Measures of true playtime

Measuring the full welfare effects of information requires measures of true playtime h_{ij}^T. We directly observe realized playtime for the owned games; the additional task is to create realistic measures of playtime for unowned games, where playtime is not observed. One might be tempted to directly use the sophisticated predictions as estimates of true playtime, but doing so would understate the dispersion of realized playtime across unowned games. Instead, we need to realistically model the errors in the predictions.

Our main approach estimates h_{ij}^T for the unowned games as the prediction (h_{ij}^P) plus an error drawn from the empirical distribution of errors for owned games, as follows. First, we sort all observations by the prediction h_{ij}^P. Then, for unowned games, we use the realized error associated with the owned observation with the next-highest value of h_{ij}^P. Matching each unowned game to the owned game with the next-higher h_{ij}^P is a simple, nonparametric, rank-based way to borrow a residual from a similar point in "prediction space." It is a form of right-continuous nearest-neighbor/quantile-mapping of residuals that preserves the monotone ordering by h^P, captures heteroskedasticity in errors, avoids parametric assumptions, and prevents optimistic interpolation. Because of the way that the errors are assigned to predicted playtimes h_{ij}^P, our predictions of realized playtimes (h_{ij}^T) reproduce the same

[19] We discuss all prediction approaches, including their fits, in more detail in Appendix Section A.

[20] Achievable RMSE values are often quite compressed (Koren, 2009) but there is evidence that even small improvements in RMSE terms can have a significant impact on the quality recommendations. To put this number in perspective: The equivalent improvement of personalized recommendations via matrix factorization ($k = 50$) over product-level averages on the Movielens 100K dataset – a common benchmark dataset used in the recommender literature – and predicting star ratings on a 1-5 scale is 8.1 percent (Adomavicius and Zhang, 2012).

correlation of predicted playtime and the error for both owned an unowned games.[21]

As an alternative to our baseline "borrowed errors" approach, we also draw errors for unowned games from a parametric distribution fitted to the errors of owned games. We fit a two-component normal mixture to the observe errors, which are skewed and fat-tailed. Each simulated error is drawn from the distribution of the fitted mixture: we convert a standard normal draw into a percentile and take the mixture's value at that percentile. We induce the normal draws to be correlated with the prediction, so that the correlation between the log prediction and the simulated error matches its value among owned games (-0.0185). The parametric errors enter robustness analyses in Sections **??** and 5.

We report summary statistics for prices as well as our measure of realized playtimes in Table 2, which reports averages of user-by-game observations for owned games, unowned games, and total games. Prices average $19.47 per game, with little difference between owned and unowned games. By contrast, owned games deliver more playtime than unowned games. Using our baseline error estimates, owned games deliver an average of 60.1 realized hours of playtime, while unowned games deliver an average of 41.5 hours. This suggests both that status quo consumers have some information and that they value playtime. Moreover, the average amounts of time that sample users spend playing are highly correlated, across games, with the amounts of time that general users spend playing the games, according to `HowLongtoBeat.com` (HLTB). For example, the log observed playtime and the log HLTB time, including the main story and extras, are correlated at 0.80.

3.2.3 Do we overestimate playtime for unowned games?

Unowned games may be unowned because users have private information – about how little they would use them – that our predictions lack. If so, then our measure of the welfare effect from games overlooked in the status quo could overstate the gains. Our predictions

[21] Unlike linear regressions, which produce errors that are orthogonal to predictions, the deviations of realized playtime from the matrix factorization predictions can be correlated.

incorporate at least some of this difference: a regression of the log of our predicted playtime measure on individual and game fixed effects, along with an unowned indicator, gives a coefficient of -0.150 (0.021) on the unowned indicator. That is, unowned games are predicted to be played substantially (13.9 percent) less than owned games.

Still, the predictions may insufficiently shade the playtime that would ensue if users purchased them. To test this, we would ideally compare playtime for users choosing to own games with playtime for users who got games without paying for them. If the owners obtaining the games for free played the games less than the predictions implied, then we would want to shade predictions to avoid overestimating benefits.

A separate game purchase dataset allows a comparison resembling the ideal test: We compare the realized playtime for full-price game purchasers to those obtaining the same game at a discount. We regress the log of observed playtime in the first 56 days after purchase on the discount share as well as purchase week and game fixed effects. The discount share coefficient is -0.207 (0.049) , indicating that playtime is 18.7 percent lower for games discounted 100 percent. We cannot reject the hypothesis that the reduced playtime when free equals the reduced playtime for unowned games in our main dataset (p-value = 0.28). Hence, the lower playtime in our realized playtime predictions shades enough, allaying concerns that our predicted playtime for unowned games overstates how much they would be played. Still, in robustness checks we allow for additional playtime shading of up to 50 percent.

3.3 Playtime, user reviews, and regret

Our model assumes that informed consumers would buy games yielding more weighted hours per dollar before others and therefore that deviations from that pattern – purchases of games yielding little weighted playtime per dollar spent – would be regretted. We observe positive and negative reviews from 20,985 users. We take "not recommend" reviews as indicators of regret; and of the 58,837 reviews that we observe for owned user-game combinations in

our data, 12.6 percent (7,420 reviews) are not recommended. We explore the relationship between playtime per dollar spent and regret by estimating regressions of measures of non-recommendation on the logarithm of playtime per price for owned games.

Table 3 shows coefficients from these regressions. The first four columns include only observations for which a user left "recommended" or "not recommended" feedback and includes only users leaving both positive and negative reviews. The columns vary in their use of fixed effects. Regardless of whether game or user fixed effects are included, playtime per dollar spent bears a negative and significant relationship to the tendency to not recommend the game. Columns (3) and (4) include game fixed effects, and the coefficients are even larger in size than those in columns (1) and (2), providing some evidence that games may differ in the utility they provide per playtime. Columns (5) and (6) use the same users but also include user-game observations without a recommendation, and the dependent variable is a regret index (1 for not recommended, 0 for no review, and -1 for recommended). Column (5) reports a linear model with game and user fixed effects, while column (6) reports an ordered probit without fixed effects. In both columns, the coefficients are negative and significant. We take the estimates in Table 3 as evidence that games that are little-played in relation to their prices are regretted and, by extension, that a function of weighted playtime less expenditure is a reasonable representation of utility.

4 Descriptive welfare evidence

Using our measures of true playtime, we document the additional raw and weighted hours of playtime that status quo expenditures could buy if users had better information. We do this in three parts. First, Section 4.1 infers game-specific values of an hour from purchase behavior and so constructs the weighted hours W of Section 2. Section 4.2 uses both owned and unowned games to calculate the additional playtime that full pre-purchase information

would allow at status quo expenditure using raw hours as the usage measure underlying utility. We then compare these numbers to measures based on both weighted hours in Section 4.3. Section 4.4 explores robustness on the descriptive welfare measures.

4.1 From hours to weighted hours

We infer the hours weights from the relationship between the tendency for users to own games, in relation to the hours the games deliver and their prices. To the extent that people value an hour of play similarly across games, they will be more willing to buy games delivering more hours, all things equal. But this correlation is not perfect, and we view ownership deviations as reflections of variation in utility of playtime. To explore this variation, we aggregate the data to the game level. Define s_j as the share of our users owning game j and H_j as a measure of the (unweighted) average hours delivered by game j. We postulate that consumer i's utility for game j depends on a function of the average hours the game delivers, its price, a game-specific unobservable ξ_j, and an extreme value error; and we estimate this model as a simple logit, using levels and logs of average hours played, via

$$\ln\left(s_j/(1-s_j)\right) = \beta_0 + \beta \ln(H_j) + \alpha p_j + \xi_j. \tag{3}$$

We estimate that $\hat{\beta} = 0.446$ (0.066) and $\hat{\alpha} = -0.0167$ (0.0044). Figure 4 depicts the relationship between s_j and $\ln(H_j)$ across all games using "true" hours of playtime across all users. The evident positive relationship supports the idea that users purchase games for the hours that they deliver, which itself lends support to aggregation of hours across games. Yet, the points in Figure 4 are not precisely on a line. Although this is partly because the figure does not account for prices, unobserved game ownership tendencies (ξ_j) vary additionally for a variety of possible reasons, including different marginal utilities of playtime across games. A game that is owned more than its hours and price predict is assumed to deliver more value

per hour. We use this idea to infer game-specific hours weights. In particular, we compute the weights ω_j that solve $\beta \ln(H_j) + \xi_j = \beta \ln(\omega_j H_j)$. Hence, $\omega_j = \left(e^{\xi_j}\right)^{1/\beta}$. The weights associate high value to status quo choices: they attribute ownership differences that prices and hours do not explain to each game's value of an hour. The inter-quartile range for ω_j runs from 0.51 to 1.66, and the median is 0.82. We use weighted hours W as the unit of playtime in what follows, where $W_i = \sum_{j \in J} \omega_{ij} h_{ij}^T$.

By attributing each game's entire error to its utility weight, our weighted hours measure rationalizes status quo holdings as optimal. This, in turn, reduces the gain from full information relative to the status quo and therefore produces conservative measures of welfare gain from full information. Below we calculate descriptive measures using both weighted and raw hours.

4.2 Information and raw hours of playtime

The data on games that users currently own provide a useful first glimpse into the regrettable nature of status quo consumption. Consumers in our sample spend an average of $685.8 to purchase games delivering a total of 2,144.9 hours of playtime. Here, we calculate the maximum unweighted hours that each consumer might have achieved from subsets of their owned games at any level of expenditure by ordering the games they had purchased by p_j/h_{ij}^T, then summing the realized cumulative playtime. The solid line in Figure 2 shows the ensuing average realized playtimes for consumers if they had full information. As the second vertical line indicates, users could achieve the vast majority (90 percent) of status quo playtime with about 38 percent of initial expenditure, on average, or at a 62 percent "discount." The last 10 percent of playtime costs consumers an average of roughly 15 times more, per hour, than the first 90 percent. It is of course conceivable that consumers value the time spent playing marginal games more highly than their prices per hour, so Figure 2 is merely suggestive of regret at this point.

The full effect of information depends not only on the regrettable purchases avoided but also on the beneficial purchase opportunities missed in the status quo. We calculate the overall effect of information, relative to the status quo, by comparing average hours of realized playtime delivered by status quo choices against two extreme alternatives that status quo expenditure could produce, random game choices and choices made with full information. Figure 3 summarizes the resulting calculations. First, randomly chosen bundles exhausting status quo budgets deliver on average 23.6 percent fewer hours of playtime than status quo choices, confirming that consumers, on average, have useful information in the status quo. Second, a consumer armed with full information prior to purchase could on average nearly double status quo playtime (a 97.2 percent increase).

We do an analogous calculation in expenditure rather than hours terms. The better the information the consumer has, the less costly is the achievement of status quo hours. The calculation points to large welfare benefits of better information relative to the status quo. Full information allows status quo playtime to be achieved with just 17 percent of status quo expenditure – a $592 reduction in spending relative to the status quo. A consumer buying random games would need to spend 25 percent more than in the status quo for the same playtime. Hence, the average welfare benefit of full information to consumers is at least $540. Of course, the actual welfare effect of better information arises not only from a cost reduction for the achievement of status quo hours but also from users' informed choice of how many hours to purchase.

4.3 Descriptive welfare with weighted hours

The calculations above based on hours of playtime assume that an hour of playtime has equal value across games. Yet, hours are not utility. The right panels of Figures 2 and 3 repeat the exercises above with weighted rather than raw hours. While weighted hours are on a different scale from hours, ratios to the status quo are still comparable. Among owned

games, a user could achieve 90 percent of status quo weighted playtime with 21.4 percent of expenditure, a 79 percent discount (dashed line in Figure 2).

Figure 3 shows how many weighted hours status quo spending can produce with random choices compared with full information. A randomly-chosen bundle would deliver 53 percent of status quo weighted playtime (vs 76 percent in hours), and full information would delivers 45 percent more weighted playtime than the actual bundle at status quo expenditure (vs 97 percent more in hours). Weighted hours make the status quo look better, which raises the status quo benefit over random choices but reduces the benefit of full information. Yet, the proportionate gain from full information gain remains large. In expenditure terms, status quo weighted playtime could be had with 18 percent of status quo expenditure under full information (17 percent in hours), and a random bundle would cost 65 percent more than the status quo for the same weighted playtime.

4.4 Robustness

The findings above suggest substantial gains from better information. We investigate this further by asking how the descriptive results change with various alternative assumptions. These include differing prices that users pay for games, the possibility that unowned games would deliver fewer hours than our predictions indicate, weighted hours based on utility weights varying across user groups, the possibility that little-played games were obtained at no charge, and the exclusion of games that might have been obtained as parts of bundles.

Table 4 compares our baseline estimates of the weighted hours from random and full information expenditure, relative to the status quo, to various alternatives. Steam games are sometimes discounted. For some games, the median daily price equals its 25^{th} or 10^{th} percentile price; for others – those that are discounted – the 10^{th} and 25^{th} percentile prices are lower. As the first panel of Table 4 shows, replacing the days-weighted mean price with the 25th or 10th percentile of each game's price across days (with a minimum of \$0.49),

with the weights re-estimated for each price, delivers similar descriptive results essentially unchanged. Status quo expenditure could deliver 51 to 55 percent of status quo weighted hours with random purchases and 41 to 46 percent beyond status with full information.

The second panel of Table 4 explores impacts of shading our hours predictions for unowned games. If hours were between 25 and 50 percent below our predictions, then randomly-chosen status quo spending could produce 46 to 50 percent of status quo weighted hours. Shading reduces the full information increase in weighted hours, to 33 percent with 25 percent shading and a 22 percent increase with 50 percent shading. Recall that the model's predictions of unowned game usage is substantially less than the usage differential for games obtained without charge (relative to paid games), suggesting that relatively little shading – and certainly not 50 percent – is needed.

The third panel of Table 4 explores the impacts of ω_{ij} weights that vary across groups of users. Groups explored include those in 5 groups according to join years, as well as 8 groups identified by k-means ownership clusters. Resulting estimates are again similar. Random allocation delivers 42 to 51 percent of status quo weighted hours, while full information delivers 33 to 36 percent above the status quo.

Steam games are sometimes acquired in bundles. We explore this in two ways. First, the fourth panel of Table 4 explores the impact of treating little-played games as if they had been obtained without charge, using three cutoffs: 30, 60, and 120 minutes of cumulative playtime. Random allocation delivers 55 to 57 percent of status quo weighted hours, while full information delivers 53 to 64 percent above the status quo.

Second, if games in our sample were obtained inadvertently (and free of additional charge) when users obtained other games, we might overestimate the welfare shortfalls in status quo consumption if these games are not eliminated because their users used them less than the time cutoffs above. We explore this by removing possibly-bundled game titles from the

analysis altogether.[22] The last panel of Table 4 explores how the descriptive results change if we remove 14 games that may have been acquired as parts of bundles. Results are very similar to the baseline.

5 Welfare estimation

The descriptive results suggest substantial welfare forgone in status quo consumption, relative to full information. Yet, the calculations are rough in that they neither allow endogenous selection of expenditure and playtime, nor do they quantify the welfare benefit of information in dollar terms. To analyze the welfare effects of improved information in a theory-consistent way, we need two major ingredients. First, in Section 5.1 we calculate status quo and full-information budget sets relating weighted hours to all other goods, which we obtain using the framework from Section 2. Second, Section 5.2 uses a Cobb Douglas utility function over weighted hours and all other goods for selecting the utility-maximizing points on the respective budget constraints. In Section 5.3, we present estimated welfare benefits of full information, both overall and from avoiding regretted purchases. Section 5.4 reports the robustness of the estimates to the measurement of prices and to the modeled playtime of unowned games. Finally, Section 5.5 decomposes the change in CS into components associated with information as opposed to cross-game variation in weighted playtime per dollar of game price.

5.1 Information and the expected budget constraint

The budget constraint that each consumer faces depends on their information, which is embodied in the model through the parameter κ_i in Equation (2). In particular, we model each consumer's rank ordering of games as dependent on an index which is a convex combination

[22] See Appendix A for determination of which games may be parts of bundles.

of the random and full information ranks, $I(\kappa_i) = \kappa_i r^T + (1 - |\kappa_i|)r^\epsilon$, where $\kappa \in [-1, 1]$, and r^T orders games by price per weighted hour.

We simulate the index based on 50 draws of the random ranking r^ϵ for each user and each value of $\kappa \in \{-1, -0.9, \ldots, 0.9, 1\}$. For each user and κ, we compute the average weighted hours of playtime that each user's status quo expenditure would deliver, $\tilde{W}_i(P_i; \kappa)$; and we choose the κ_i that minimizes the distance between the average simulated playtime $\tilde{W}_i(P_i; \kappa)$ and the status quo playtime W_i. The median value of κ is 0.4, and the inter-quartile range runs from 0.2 to 0.5.

Each draw gives a user's budget constraint. We calculate the slope of the status quo budget constraint for each user using the ratio of the average game price to the average weighted hours based on the 50 draws for the marginal game (the game purchase where simulated cumulative expenditure is closest to status quo expenditure). The resulting user-specific slopes of the status quo budget constraints ($p_W(\kappa)$) imply that for the median user an additional weighted hour costs \$0.32 at users' status quo bundles. The inter-quartile range varies between \$0.17 at the 25$^{\text{th}}$ percentile and \$0.57 at the 75$^{\text{th}}$.

5.2 Cobb Douglas calibration

We estimate welfare effects using a Cobb-Douglas calibration where utility depends on both weighted hours of playtime and consumption of all other goods as $U = W^a A^{1-a}$, where a is informed by expenditure data.[23] We observe the status quo weighted hours choice for each user, which (suppressing i subscripts) we now denote by W_0; and the κ derivation above delivers the status quo budget constraint slope at W_0. This allows us to infer the status quo levels of "all other goods," or AOG level A_0 and therefore to calculate status quo utility, as follows. User utility maximization occurs where their marginal rate of substitution

[23] We report Cobb Douglas estimates based on $a = 0.05$. Results are nearly identical with $a = 0.1$ or $a = 0.005$.

($MRS = aA_0/(1-a)W_0$) equals the slope of the budget constraint, $p_W(\kappa)$. Rearranging terms, we solve

$$A_0 = \frac{1-a}{a} W_0 \, p_W(\kappa)$$

for each user. Status quo utility is then $U_0 = W_0^a A_0^{1-a}$.

In order to generate the full-information budget constraints, we first need the point where the status quo and full information budget constraints meet the AOG axis (the user's income level). We calculate this as the sum of A_0 and status quo expenditure, or $I = A_0 + P_0$. Then, given I, we create the full-information budget constraints by sorting all products by their price per weighted realized hour of playtime, and we calculate the optimal choices for each user as their utility-maximizing points along these budget constraints. The value of information, relative to status quo choices, is the amount of money the user would need to forgo from an informed state to bring their utility to the status quo level.

5.3 Welfare benefits of full information

While the full welfare gains we will calculate arise from better information about both purchased and non-purchased games, we begin with a simpler calculation of the gains from avoiding purchases that users regretted, and we then report the full welfare results from all games. Table 5 presents welfare effect estimates. The top panel – "actual" – describes status quo hours, weighted hours relative to the status quo, game purchases, and expenditure.

The first row below "full information" presents the no-regret results. We solve the utility models against budget constraints that include only the games purchased in the status quo. By construction, expenditure, games purchased, and hours decrease. With full information about already-owned games, users would buy an average of 10.5 of the 35.7 games they had purchased in the status quo, reducing their expenditure by $504.0, or by about 73 percent. Because consumers are just eliminating games, their playtime must fall; but the average

weighted hours reduction is only about 5 percent. Given the large expenditure reductions and small weighted playtime reductions, consumer surplus rises by an average of $398.0, or by more than half of status quo expenditure.

The "overall effect" row of Table 5 reports the welfare effects of full information about all games, including those that the consumer might have missed in the status quo. While avoiding regret by itself reduces both purchases and hours played, full information about all games raises weighted hours played substantially. Weighted hours rise by roughly a third above the status quo level. At the same time, expenditures (and games purchased) decline by about three fifths. Full information raises consumer surplus by an average of $912.44, or by 133 percent of the $685.8 in status quo expenditures. These estimates of the change in CS are roughly 2.3 times the effects from eliminating regret alone. Hence, full information would allow substantially higher utility with much less expenditure, relative to the status quo; and 56 percent of the overall gain stems from the purchase of otherwise-unknown games.

5.4 Robustness of the welfare estimates

Just as we explored the effect of different assumptions on pricing and playtime on descriptive results in Section 4.4, we do the same here for welfare estimates. Table 6 reports the full-information gain and the share of the ΔCS from avoiding regret from various robustness exercises.

The first panel ("Price") explores the effect of alternative pricing assumptions, that users bought games at their 10^{th} and 25^{th} values rather than the mean over the observed period. While status quo expenditure is perforce smaller with lower prices, the welfare effects remain quite similar. The ratio of ΔCS to baseline expenditure is 1.44 with 25^{th} percentile prices and 1.35 with 10^{th} percentile prices, and the respective shares from avoiding regret are 41.6 and 42.3 percent.

The second panel shows the effect of assuming the hours of unowned games are 25 and 50

percent below our basic predictions. This exercise leaves baseline expenditure unchanged but reduces the $\Delta CS/_{\text{expenditure}}$ ratio to 1.22 for 25 percent shading and 1.07 for 50 percent shading. The share of the welfare gain from avoiding regret varies between 46 percent with 25 percent shading and 49 percent with 50 percent shading. Even with very substantial shading – far beyond what is suggested in Section 3.2.3 – the welfare gains from full information remain 107 percent of status quo expenditure.

The third panel explores the impact of group-specific game utility weights (ω_{ij}) for different groups of consumers $i \in$ groups G. These include five groups by when users joined the platform and 8 groups that were identified via k-means clustering according to the games they own. Here, again $\Delta CS/_{\text{expenditure}}$ remains substantially above 1, while the share of the gain from avoiding regret remains nearly half.

The fourth panel of Table 6 explores the impact of treating little-played games as free rather than having been purchased at their average price. Assuming that users paid less for games leads to conservative estimates of welfare gains from full information. Using cutoffs of 30, 60, and 120 minutes for deeming games to have been free reduces $\Delta CS/_{\text{expenditure}}$ slightly below the baseline 1.33, to between 1.27 (with the 30 minute cutoff) and 1.22 for the 120 minute cutoff. Because we treat some purchased games as free, the share of the gain from avoiding regret is slightly smaller: between 41.9 percent for the 30 minute cutoff and 37.9 percent for the 120 minute cutoff.

The last panel of Table 6 shows how the welfare measures change with the exclusion of the 14 games sometimes acquired in bundles. Results remain very similar to the baseline.

We conclude that our baseline estimates are robust to a wide range of alternative assumptions and that the welfare gain from full information would (unless we have substantially overestimated playtime for unowned games) raise welfare by more than status quo expenditure.

5.5 Decomposing the change in CS

In our modeling framework, changes in expenditure and CS depend on two factors. First, consumers can be differently informed in the status quo. Users with poor information – low or negative κ_i and therefore less bowed-out status quo budget constraints – get larger benefits from additional information, all else constant. Second, users for whom the price of weighted hours varies substantially across games (i.e., with a higher variance of $\omega_j h_{ij}^T/p_j$) have more to gain from additional information. Empirically, both matter; but the variation in κ gives rise to much more of the variation in welfare outcomes.[24] Because we have an estimate of ΔCS for each individual, we can also explore how it varies with user experience on the platform, measured by their join year. The change in CS, relative to expenditure, is higher for users with less time on the platform. The ratio ΔCS/expenditure is roughly 2.0 for users joining in 2015 and 2016 while it averages 1.4 for users joining between 2002 and 2011, and it rises almost monotonically across join years from 2011 to 2017.

6 Conclusion

Differentiated products can deliver substantial value to heterogeneous consumers, but only if people know which products to purchase. Lack of post-purchase usage data has diverted attention from the possible problems of regret and missed opportunities in differentiated product choices. Using novel data on post-purchase usage of video games, we document that status quo purchases deliver usage that falls far short of what full information would allow; and we present evidence that games that are little-played in relation to their prices are regretted by their buyers. Descriptive analysis shows that full information would allow

[24] We run the regression $\Delta CS_i = \nu_0 + \nu_1 \kappa_i + \nu_2 \sigma_i(\omega_j h_{ij}^T/p_j) + \varepsilon_i$, where the σ term is the user-specific standard deviation of $\omega h/p$. We estimate $\hat{\nu}_1$ = -1206.0 (se=11.15) and $\hat{\nu}_2 = 0.0115$ (0.0003). The resulting coefficients imply that a one-standard deviation increase in κ changes ΔCS by 3.1 times the change for a one-standard deviation increase in our measure of the variation in $\omega h/p$.

consumers to obtain more playtime, and playtime that would be substantially more valuable to them, with status quo expenditures.

To measure the welfare effects of better information with and explicit two-part framework. First, we develop a model of how information affects choice sets; and second, we create a tractable model of consumer choices of product bundles, in which the marginal utility of time using each product may vary. Using a Cobb Douglas calibration, we find that status quo consumption forgoes a great deal of potential benefit. In our setting, full information could raise CS by about 130 percent of status quo expenditure while cutting expenditure by three fifths. Our finding of a large effect of full information is robust to a wide range of alternative assumptions about the prices users paid, our predictions of playtime for unowned games, and varying marginal utilities of each game across user groups.

The potentially undesirable supply-side effects of entry and product availability in differentiated product markets are well understood, both theoretically and empirically. We believe that the demand-side shortfalls in consumer surplus from imperfect information also merit attention. The welfare shortfalls might arise in a variety of contexts in which heterogeneous consumers choose among differentiated products. Given suitable data – on post-purchase usage and user satisfaction – documenting the shortcomings of status quo consumption for other contexts is a fruitful area for further study.

References

Adomavicius, G. and J. Zhang (2012): "Stability of recommendation algorithms," *ACM Transactions on Information Systems*, 30, 1–31.

Aguiar, M., M. Bils, K. K. Charles, and E. Hurst (2021): "Leisure luxuries and the labor supply of young men," *Journal of Political Economy*, 129, 337–382.

Allcott, H. (2013): "The welfare effects of misperceived product costs: Data and calibrations from the automobile market," *American Economic Journal: Economic Policy*, 5, 30–66.

Anderson, S. P., A. De Palma, and Y. Nesterov (1995): "Oligopolistic competition and the optimal provision of products," *Econometrica: Journal of the Econometric Society*, 1281–1301.

Bergstra, J. and Y. Bengio (2012): "Random search for hyper-parameter optimization." *Journal of Machine Learning Research*, 13.

Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez (2013): "Recommender systems survey," *Knowledge-Based Systems*, 46, 109–132.

Bronnenberg, B. J., J.-P. Dubé, M. Gentzkow, and J. M. Shapiro (2015): "Do pharmacists buy Bayer? Informed shoppers and the brand premium," *The Quarterly Journal of Economics*, 130, 1669–1726.

Chernew, M., G. Gowrisankaran, and D. P. Scanlon (2008): "Learning and the value of information: Evidence from health plan report cards," *Journal of Econometrics*, 144, 156–174.

Chevalier, J. A. and D. Mayzlin (2006): "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research*, 43, 345–354.

Chu, C. S., P. Leslie, and A. Sorensen (2011): "Bundle-size pricing as an approximation to mixed bundling," *The American Economic Review*, 263–303.

Compiani, G., G. Lewis, S. Peng, and P. Wang (2024): "Online search and optimal product rankings: An empirical framework," *Marketing Science*, 43, 615–636.

Crawford, G. S. and A. Yurukoglu (2012): "The welfare effects of bundling in multichannel television markets," *American Economic Review*, 102, 643–685.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic competition and optimum product diversity," *The American Economic Review*, 67, 297–308.

Einav, L., B. Klopack, and N. Mahoney (2023): "Selling Subscriptions," Tech. rep., National Bureau of Economic Research.

Fang, L. (2022): "The effects of online review platforms on restaurant revenue, consumer learning, and welfare," *Management Science*, 68, 8116–8143.

Funk, S. (2006): "Netflix update: Try this at home," .

Jannach, D., P. Resnick, A. Tuzhilin, and M. Zanker (2016): "Recommender systems—beyond matrix completion," *Communications of the ACM*, 59, 94–102.

Koren, Y. (2009): "Collaborative filtering with temporal dynamics," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 447–456.

Koren, Y., R. Bell, and C. Volinsky (2009): "Matrix factorization techniques for recommender systems," *Computer*, 42, 30–37.

Koren, Y., S. Rendle, and R. Bell (2021): "Advances in collaborative filtering," *Recommender systems handbook*, 91–142.

Lee, D. and K. Hosanagar (2021): "How do product attributes and reviews moderate the impact of recommender systems through purchase stages?" *Management Science*, 67, 524–546.

Lee, R. S. (2013): "Vertical integration and exclusivity in platform and two-sided markets," *American Economic Review*, 103, 2960–3000.

Mankiw, N. G. and M. D. Whinston (1986): "Free entry and social inefficiency," *The RAND Journal of Economics*, 48–58.

Miravete, E. J. (2003): "Choosing the wrong calling plan? Ignorance and learning," *American Economic Review*, 93, 297–310.

Nelson, P. (1970): "Information and consumer behavior," *Journal of political economy*, 78, 311–329.

Reimers, I. and J. Waldfogel (2021): "Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings," *American Economic Review*, 111, 1944–1971.

Reinstein, D. A. and C. M. Snyder (2005): "The influence of expert reviews on consumer demand for experience goods: A case study of movie critics," *The journal of industrial economics*, 53, 27–51.

Rendle, S., W. Krichene, L. Zhang, and J. Anderson (2020): "Neural collaborative filtering vs. matrix factorization revisited," in *Proceedings of the 14th ACM Conference on Recommender Systems*, 240–248.

Skelton, A. C. and J. M. Allwood (2017): "Questioning demand: a study of regretted purchases in Great Britain," *Ecological Economics*, 131, 499–509.

Spence, M. (1976): "Product differentiation and welfare," *The American Economic Review*, 66, 407–414.

Stigler, G. J. (1961): "The economics of information," *Journal of political economy*, 69, 213–225.

Train, K. (2015): "Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples," *Journal of choice modelling*, 16, 15–22.

Villas-Boas, S. B., C. Bonnet, and J. Hilger (2021): "Random utility models, wine and experts," *American Journal of Agricultural Economics*, 103, 663–681.

Ward, M. R. (2010): "Video games and adolescent fighting," *The Journal of Law and Economics*, 53, 611–628.

Weitzman, M. L. (1979): "OPTIMAL SEARCH FOR THE BEST ALTERNATIVE." *Econometrica*, 47.

Tables and figures

Table 1: RMSE for different prediction models

Approach	RMSE			Details
	Train	**Validation**	**Test**	
Collab. Filtering	**0.295**	0.599	0.614	$k = 100$
	0.311	**0.595**	**0.612**	$k^* = 55$
	0.440	0.606	0.623	$k = 50$
	0.463	0.617	0.633	$k = 10$
	0.523	0.623	0.645	$k = 5$
Game & user FEs	0.589	0.611	0.627	
Game & user char's	0.616	0.623	0.640	Random Forest (sampsize=5k)
Game & user char's	0.625	0.630	0.645	All characteristics from below
Game FEs	0.629	0.634	0.648	Identical to game char's which already fully saturate the item fixed effect space
Game char's	0.629	0.634	0.648	6 attributes & 95 tag dummies
Game char's	0.684	0.688	0.701	6 attributes
User char's	0.703	0.708	0.720	7 attributes & 117 country dummies
User char's	0.704	0.708	0.721	7 attributes
Global average	0.706	0.711	0.722	

Note: Prediction error (RMSE) for training, validation, and test sets for a data subset of 39,130 users for whom user-level characteristics are available and 100 games with at least partial game-level level data. The models are estimated with OLS unless stated otherwise. Game and user attributes used in the models are described in detail in Appendix Section **??**, and missing observations were imputed with means.

Table 2: Summary statistics

	by ownership status		total
	owned games	non-owned	all games
Price	19.20	19.62	19.47
	(11.54)	(12.29)	(12.03)
Hours played	60.14	41.50	48.16
	(254.29)	(199.06)	(220.58)
N	1,784,919	3,215,081	5,000,000

Note: Means and standard deviations for expenditure and playtime for owned and unowned games (Columns 1 and 2), as well as for all games in the dataset (Column 3). All summary statistics are at the user-game level, and playtime is measured in hours per game.

Table 3: Playtime and expressions of user regret

	"not recommended" among reviewed				regret index	
	(1) None	(2) User	(3) Game	(4) Both	(5) Both	(6) Probit
ln(mins per dollar)	-0.0262^{***}	-0.0439^{***}	-0.0491^{***}	-0.0683^{***}	-0.0357^{***}	-0.0979^{***}
	(0.00166)	(0.00198)	(0.00170)	(0.00202)	(0.00052)	(0.00178)
N	21,342	21,342	21,342	21,342	147,410	147,410

Note: Regressions of user recommendation measures on time played per dollar spent. Columns 1-4 use only user-game observations with a recommendation and from users who left both positive and negative reviews. The dependent variable is 1 for negative recommendations. Column 1 includes no fixed effects, Column 2 includes user fixed effects, 3 includes game fixed effects, and 4 includes both. Columns 5 and 6 include the non-recommended game observations as well (for users leaving both positive and negative reviews), and the dependent variable is an index: 1 for "not recommended," 0 for no review, and -1 for "recommended." Column 5 reports a linear model with user and game fixed effects, and Column 6 reports an ordered probit without fixed effects.

Table 4: Robustness of descriptive results to measurement assumptions

	weighted playtime relative to status quo	
	random bundle	full information
Weighted baseline (mean price)	0.533	1.449
Price		
25th percentile price	0.513	1.405
10th percentile price	0.548	1.462
Games played less than the cutoff treated as acquired free		
30 minutes	0.554	1.534
60 minutes	0.561	1.575
120 minutes	0.570	1.635
Excluding 14 suspected bundle games	0.542	1.452
Modeled playtime of unowned games reduced by		
25 percent	0.495	1.333
50 percent	0.456	1.218
Weights specific to groups of users		
Join cohort (5 groups)	0.506	1.357
Ownership clusters (8 groups)	0.415	1.330
Normal-mixture errors for unowned games	0.531	1.478

Note: Weighted playtime relative to status quo weighted playtime for randomly chosen bundles (first column) and bundles chosen under full information, when spending equals status quo expenditure. The first row shows baseline levels consistent with Figure 3. The *Price* panel (Panel a) uses different assumptions on the price paid for each game. Panel b shades the modeled playtime for unowned games by 25 and 50 percent. Panel c allows the ω weights to vary not only across games but also across consumers, where consumers are either divided by the year they joined the platform, or via k-means clustering by game ownership. Finally, Panel d treats games played less than a certain time threshold as obtained for free.

Table 5: Welfare results

	hours	weighted hours W (actual = 100)	# games	game expenditure	ΔCS
Actual	2144.9	100.0	35.71	685.81	
Full information					
Avoiding regret	1671.5	94.6	10.50	181.79	397.98
Overall effect	2921.5	131.1	14.98	255.17	912.44

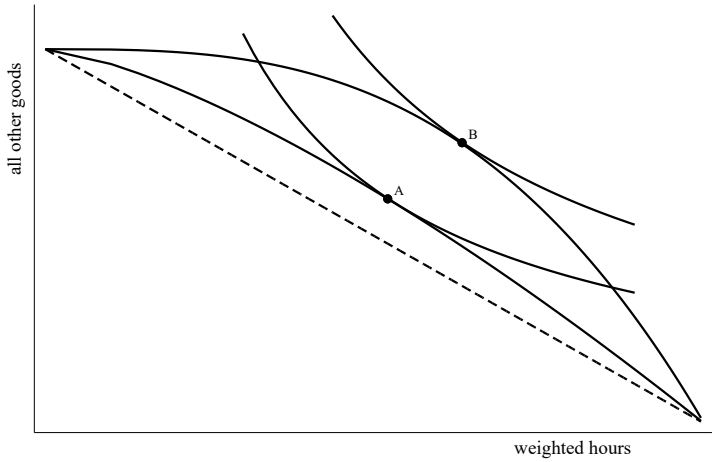
Note: Simulations of the effects of full information on purchase and usage of, as well as consumer surplus from, all games in the dataset. These simulations quantify the combined value of both avoiding regrettable status quo purchases and finding otherwise-unknown games. The first row shows status quo values of hours and weighted hours indexed to 100, the number of games purchased, and game expenditure; and the remaining rows show these measures, as well as the change in consumer surplus, from full information simulations. The second – "avoiding regret" – row shows results when only considering already-owned games. The last row shows overall results when allowing consumers to purchase the best among all games in the dataset. All figures are per-person averages.

Table 6: Robustness of the welfare estimates to measurement assumptions

	status quo spending ($)	mean ΔCS / mean spending	share of ΔCS from avoiding regret
Weighted baseline (mean price)	686	1.33	43.6%
Price			
25th percentile price	647	1.44	41.6%
10th percentile price	323	1.35	42.3%
Games played less than the cutoff treated as acquired free			
30 minutes	642	1.27	41.9%
60 minutes	613	1.25	40.5%
120 minutes	567	1.22	37.9%
Excluding 14 suspected bundle games	623	1.26	45.5%
Modeled playtime of unowned games reduced by			
25 percent	686	1.22	45.6%
50 percent	686	1.07	49.0%
Weights specific to groups of users			
Join cohort (5 groups)	686	1.34	43.8%
Ownership clusters (8 groups)	686	1.18	46.6%
Normal-mixture errors for unowned games	686	1.38	42.9%

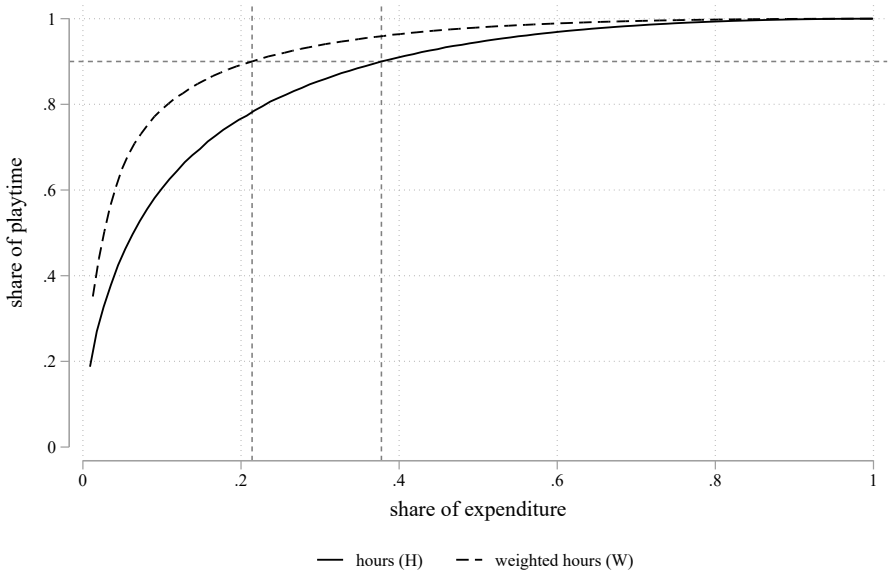
Note: Weighted hours throughout, with the weights re-estimated for each price. Each cell re-estimates the budget-constraint slopes and information parameters of Section 5.1; $a = 0.05$. Full information has every user gaining in every cell. Status quo expenditure is $634 at the median under the mean price, $596 under the 25th percentile price and $302 under the 10th percentile price; percentile prices are floored at $0.49. The mean full-information gain at the mean price is $912, $839, and $735 at 0, 25, and 50 percent shading.

Figure 1: Pre-purchase information, hours of playtime, and expenditure on games



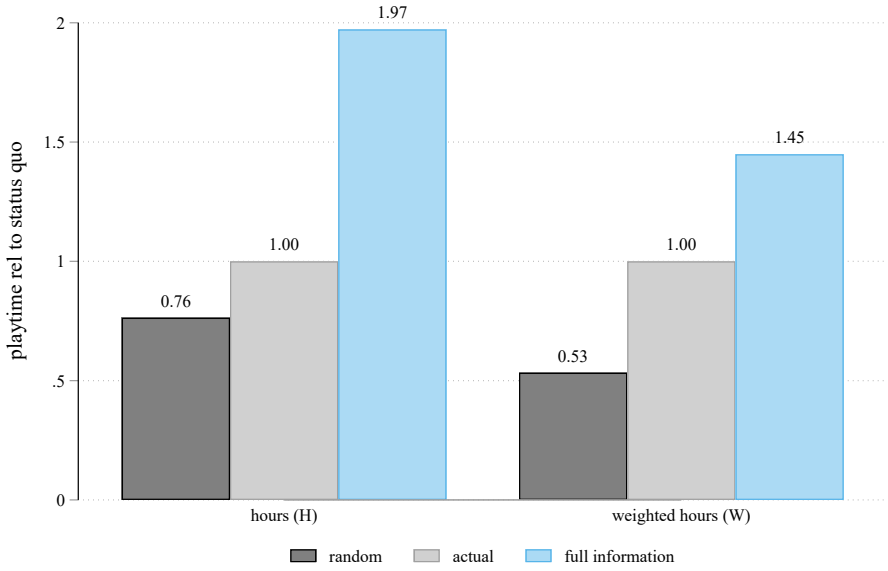
Notes: The figure depicts two budget constraints for weighted hours of playtime (x-axis) vs all other goods (y-axis). The outer budget constraint reflects full information about games' playtime and prices, while the inner curved budget constraint reflects imperfect but better-than-random information on games' playtime in relation to their prices. A fully informed consumer maximizes utility by choosing point B, while the imperfectly informed consumer maximizes utility by choosing A. The less accurate the information the consumer has, the less "bowed out" the budget constraint. A consumer with no information about games would face the (dashed) straight-line budget constraint.

Figure 2: Potential for regret



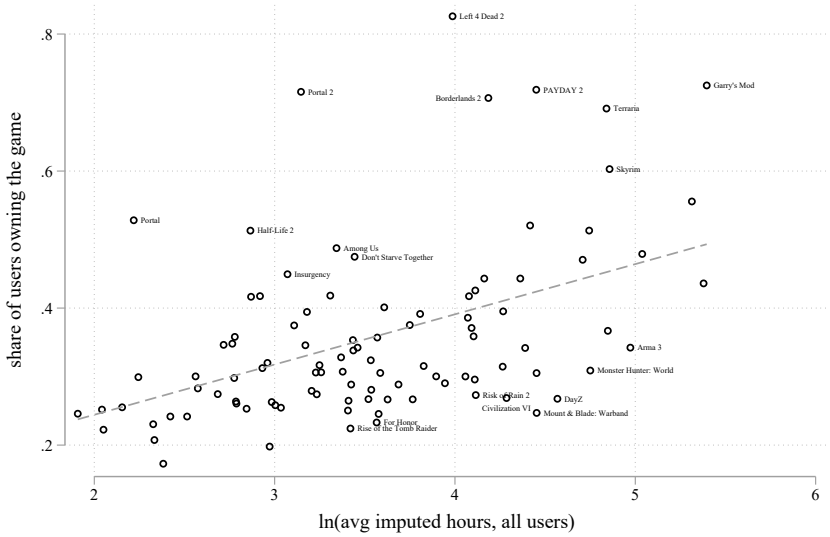
Notes: This figure plots the share of status quo playtime delivered by shares of status quo expenditure on owned games with full information, meaning that games are ordered by realized playtime per dollar spent. The solid line shows the share of raw hours and indicates that users could on average attain 90 percent of status quo playtime with about 38 percent of status quo expenditure on owned games. The dashed line provides the analogous calculation for weighted hours: 90 percent of status quo weighted playtime is reached with 21 percent of expenditure. The two vertical lines illustrate these shares.

Figure 3: Potential for welfare gains playtime at status quo expenditure



Notes: The bars show the average cumulative playtime, relative to the status quo, that status quo expenditure could purchase under different information assumptions, in hours (left) and in weighted hours (right). For example, the leftmost bar indicates that random selection of games would deliver 76 percent of status quo playtime, while full information would allow the achievement of 97 percent more playtime. In weighted hours, random choice delivers 53 percent of the status quo, and full information allows 45 percent more.

Figure 4: Ownership, playtime, and game-specific utility weights (ω_j)



Notes: The figure plots the share of users owning a game against the natural log of the average "true" playtimes (h_{ij}^T) across all users, for all 100 games in the sample.

A Games acquired in bundles

Some games enter users' libraries as parts of bundles rather than individual purchase decisions. Our main analysis excludes games receiving no playtime and therefore deals with games acquired in bundles directly for users who never use those games. By determining which game titles are frequently acquired this way, we can remove them for all users for a robustness analysis that avoids possible bundled acquisitions.

We identify candidate games by measuring elevated probabilities of joint ownership across game pairs in the libraries of our 50,000 users. For each pair of games j and k, we compute the share s of j's owners who also own k, and we compare it with the expected share e given only how many games those owners have, calculated within 20 library-size groups. Because a very popular partner leaves little room for s to exceed e, we measure the link as $(s-e)/(1-e)$, the share of the possible excess that is realized. This equals zero when ownership is what library size predicts and one when every owner of j also owns k. We flag j if this measure reaches 0.5 for any partner. This approach identifies 14 games for exclusion: Counter-Strike, Counter-Strike: Condition Zero, Half-Life, Half-Life 2, Half-Life 2: Deathmatch, Half-Life 2: Episode One, Left 4 Dead, Portal, Witcher: Enhanced Edition, Witcher 2, Borderlands: The Pre-Sequel, Borderlands GOTY, Rise of the Tomb Raider, and Don't Starve.

B Appendix tables and figures

Table A.1: shading of unowned games implied by discounted and free acquisitions

	difference in	implied shading (%)	
	ln(minutes)	whole gap	net of g
Acquisitions vs full-price purchases, within game (56 days after purchase)			
Deepest quartile of discounts (75–95%)	−0.251 (0.041)	22	10
Discounts of 85% or more	−0.394 (0.057)	33	22
100% discount (linear in the discount)	−0.207 (0.049)	19	6
Deepest quartile, with user fixed effects	−0.313 (0.050)	27	15
Free, normally paid (newer games)	0.012 (0.067)	−1	0
Unowned vs owned games in our data (mtrue, user and game fixed effects)			
Ownership gap g	−0.150 (0.021)	14	

Note: A: $(100(1 - e^{b})$, the whole gap. B: $100(1 - e^{\min(b-g,0)})$, the part of the gap not already in the ownership gap g.