

The Informational Role of Online Recommendations: Evidence from a Field Experiment*

Guy Aridor Duarte Gonçalves

Daniel Kluver Ruoyan Kong Joseph Konstan

Abstract

We conduct a field experiment on a movie-recommendation platform to investigate whether and how online recommendations influence consumption choices. Using a within-subjects design, our experiment measures the causal effect of recommendations on consumption and decomposes the relative importance of two economic mechanisms: expanding consumers’ consideration sets and providing information about their idiosyncratic match value. We find that the informational component exerts a stronger influence – recommendations shape consumer beliefs, which in turn drive consumption, particularly among less experienced consumers. Our findings and experimental design provide valuable insights for the economic evaluation and optimisation of online recommendation systems.

Keywords: Recommendations; Recommender Systems; Information Acquisition; Field Experiment; Platforms.

JEL Classifications: D83, D47, D12, L15, M37.

**Contact:* Aridor: Northwestern Kellogg, guy.aridor@kellogg.northwestern.edu. Gonçalves: University College London; duarte.goncalves@ucl.ac.uk. Kluver and Kong and Konstan: GroupLens and University of Minnesota - Twin Cities; kluve018@umn.edu, kong0135@umn.edu, and konstan@umn.edu.

This paper was originally circulated as “The Economics of Recommender Systems: Evidence from a Field Experiment on MovieLens.” We thank Heski Bar-Isaac, Yeon-Koo Che, Brett Gordon, Ilya Morozov, Shan Sikdar, Andrey Simonov, Greg Taylor, Hal Varian, and Julian Wright for their thoughtful feedback, as well as audiences at the AFE Conference (Chicago), AIM Conference (USC Marshall), ACM EC (KCL), CESifo Economics of Digitization, Chicago Booth, Columbia, Compass Lexecon, EARIE (LUISS), National University of Singapore, NBER Workshop of Digital Economics, Northwestern Institute on Complex Systems, Northwestern Kellogg, Psychology of Technology Conference (BU), PUC Chile, UK Workshop on Digital Economics, and the Virtual Quant Marketing Seminar. Daniel Kluver and Joseph Konstan are members of the GroupLens lab, which runs and maintains the MovieLens platform; Ruoyan Kong was a member of the GroupLens lab. This study was approved under both the Columbia IRB under protocol number AAAT2659 and the University of Minnesota IRB under protocol number STUDY00010960. This work was partially supported by NSF grant CNS-2016397.

First posted draft: 25 November 2022. *This draft:* 12 December 2024.

1. Introduction

Recommendation systems (RS) are nearly ubiquitous in the digital economy. These are systems that combine data from multiple consumers and sources to produce often personalised information in the form of consumer-specific recommendations (Resnick and Varian, 1997). They have a wide set of applications: from e-commerce (Dinerstein et al., 2018) and labour markets (Horton, 2017), to curation feeds on social media platforms (Levy, 2021), to cultural goods on media streaming platforms (Holtz et al., 2020), and to the articles served on news platforms (Chiou and Tucker, 2017; Claussen et al., 2023). While these systems have been shown to increase consumption and engagement across these different contexts, their purportedly large influence has spawned a debate about their welfare implications as well as precisely what these systems ought to optimise for in order to best assist consumers (McNee et al., 2006; Kleinberg et al., 2023). The key question underlying this debate is *how* they influence consumption choices.

In this paper, we discuss a field experiment on a movie recommendation platform, MovieLens, that measures the causal effect of online recommendations on consumption and decomposes the economic mechanisms that drive their influence. As recommender systems are typically deployed in environments with large choice sets, one prominent mechanism is that they can make consumers *consider* goods, including goods that they did not know about. Furthermore, as these systems are prevalent in markets with experience goods, they can provide *information* on goods’ idiosyncratic match value. We use a within-subjects design that allows us to separately measure the effect of these mechanisms on consumption. We find that, while the consideration channel plays a meaningful role in increasing consumption, the informational channel plays a more substantial and important role in driving the effects of recommendation.

Our experimental design addresses two main challenges in measuring the causal effects of recommendations and disentangling these two channels, while being general enough to be implemented on any online platform with a recommendation system.

The first challenge is identifying the causal impact of recommendations on consumption, which is complicated by the fact that recommendations are targeted to consumers and so the experimental design needs to account for this selection while still providing high quality recommendations. A typical experimental design within this literature compares consumption between a group of consumers who receive recommendations from the recommender system algorithm and another who receive recommendations from reasonable alternative algorithms (Holtz et al., 2020; Korganbekova and Zuber, 2023). While this design works well for assessing the effect of the recommendation system, it leads to a systematic bias when measuring the effect of recommendation

as an RS should, by design, result in a treatment group with systematically higher idiosyncratic quality. Our experimental design exploits a unique aspect of our data to isolate the role of recommendation as we observe the intermediate outputs of the RS algorithm that provide us with an estimate of consumer-specific quality for each good. This motivates a within-subjects design that uses these estimates to generate for each consumer a *control* group of goods, deliberately excluded from recommendations, and a *recommendation* group of goods selected for recommendation, both of which have similar ex-ante idiosyncratic quality according to the RS estimates.

The second challenge is that consumer beliefs are typically unobserved and that recommendations simultaneously make consumers consider goods and provide information about good quality. We address the measurement issue with a belief elicitation survey that captures consumers’ beliefs about the quality of unconsumed goods and how certain they are about these assessments. In this survey consumers only see the name of the good and its movie poster – they cannot see more details about the good or the platform’s predicted rating. A byproduct of this elicitation is that it makes consumers consider the good, without providing the informational content of recommendation.¹ We exploit the variation induced from this survey to include a third *consideration-only* group of goods for each consumer, which only show up in the belief elicitation survey but not in the recommendations. By carefully selecting the set of goods to elicit beliefs about and comparing consumption frequencies across groups, we measure the causal increase in consumption due to recommendation and quantify its informational gains.²

Our experiment is conducted on a movie-recommendation platform, MovieLens, which has existed since 1997. MovieLens is noncommercial and devoted to producing helpful recommendations (it does not host movies) and features open-sourced data and algorithm implementation. Its data constitute a central benchmark in the recommender system community for the development and evaluation of new recommender system algorithms, having been used in thousands of papers.³ Specifically, the platform uses past ratings paired with a collaborative filtering algorithm (Harper and Konstan, 2015) to produce *user-specific* predicted ratings for unrated movies. These predictions are displayed to users in the first row of the platform homepage and are used to tailor the platform’s *user-specific* recommendations (“top picks”). As such, the recommendations

¹As with much of the literature on consideration sets – e.g., Masatlioglu et al. (2012) and Manzini and Mariotti (2014) – we do not distinguish between unawareness of some alternatives or working memory costs or limitations that require consumers to focus on a particular subset of the available goods.

²Throughout the paper we are agnostic to whether the informational gains come from direct inferences that consumers make as a result of being recommended an item and indirect information accrued by the fact that recommendations reduce information acquisition costs.

³Two examples in economics are Chen et al. (2010) and Rossi (2021). The vast majority of papers using MovieLens data rely on the ratings dataset to evaluate the performance of new recommendation system algorithms – see Harper and Konstan (2015) for an overview.

are personalised and considered high-quality. We maintain the high quality of recommendations during our intervention by including only the top 750 goods for each consumer as determined by the RS estimate.⁴

Our first main finding is that both consideration and recommendation induce a significant increase in consumption. Under our preferred specification, we find that consideration alone leads to a 0.2 percentage point (p.p.) increase in consumption relative to a baseline of 1.0 p.p. consumption of goods in the control group. In contrast, recommendation leads to a 1.3 p.p. increase in consumption relative to the consideration group, indicating that recommendation nearly doubles the probability of consumption relative to consideration alone.

Our second set of findings indicates that the larger increase in consumption from recommendations is primarily driven by their informational role. We show that recommendations causally influence beliefs, and that these, in turn, causally impact consumption decisions.

In order to causally identify how changes in beliefs drive consumption, we leverage the fact that our experimental intervention induces exogenous variation in recommendations. Realising that such variation constitutes randomised information provision (Haaland et al., 2023) is the basis of our identification strategy, relying on an instrumental-variable approach. We find that a one-point increase in expected match value and a decrease in uncertainty (both on a 1-5 scale) lead to an 8.2 and 12.6 percentage point increase in consumption, respectively.

Having shown that beliefs drive consumption, we then turn to examining how recommendations affect beliefs. We find that recommendations reduce uncertainty by 0.063 and shift expected match value assessments closer to the platform’s predicted rating by 0.015. Paired with the earlier estimates on the relationship between beliefs and consumption, these changes in beliefs account for much of the observed increase in consumption. Additionally, less experienced consumers – those with shorter consumption histories – are more uncertain in their assessments and experience a larger causal increase in consumption from recommendations. This demonstrates that the informational role of recommendations operates as a function of the experience the consumer has, with newer consumers benefiting the most.

These results have important implications for understanding the welfare consequences of online recommendations. Specifically, they imply that recommendations do more than manipulate consideration sets: they provide consumer-specific match value information in markets with experience goods. This suggests that, so long as the information provision from the recommendation

⁴The focus is thus on estimating the effect of recommendations on consumption where the set of recommendations comprises “high quality” recommendations coming from the platform’s recommendation algorithm and not randomly selected movies, which would not contain an informational component.

system is unbiased,⁵ recommendations improve consumer welfare largely through making them more informed about their match values, and not through expanding consideration sets. This suggests that the increasing trend of online platforms to introduce bias into their recommendation systems by steering consumers to more profitable goods can mainly hurt consumer welfare to the extent that it erodes information provision (Armstrong et al., 2009; Hagiu and Jullien, 2011; Calvano et al., 2022).

Understanding the economic mechanisms through which recommendations influence choice is crucial not only for assessing their welfare implications but also for guiding the design of recommendation systems aimed at maximising consumer welfare. While the idea that recommendations primarily function as information provision is economically intuitive, it contrasts with a dominant focus in the literature on expanding consumers’ consideration sets as the hallmark of “good” recommendations (Castells et al., 2015; Kaminskas and Bridge, 2016; Steck, 2018). Our findings suggest that recommendations designed to provide match value information about considered goods, as emphasised by the serendipity evaluation criterion (Kotkov et al., 2016), are more effective at driving consumption and delivering what are often perceived as “better” recommendations. Moreover, our findings validate the importance of collecting economically motivated belief data, beyond traditional consumption data, as a tool for online platforms to directly measure and optimise for the informativeness of recommendations.⁶

1.1. Related Literature

This paper contributes to a burgeoning literature on recommender systems and the impact of recommendations.

The Impact of Recommendation on Consumption. Recent literature has examined whether recommendation systems impact consumption patterns. Senecal and Nantel (2004), Das et al. (2007), Freyne et al. (2009), Zhou et al. (2010), Claussen et al. (2023), Holtz et al. (2020), and Donnelly et al. (2024) show that recommendation systems increased consumption in hypothetical choices in a lab experiment, Google News, a social network, a news website, YouTube, Spotify, and Wayfair respectively. These papers collectively provide evidence that personalised recommendations, relative to a non-personalised benchmark, meaningfully increase consumption. There has subsequently been a significant amount of work measuring the aggregate effects of recommendation systems on the types of goods that get consumed (Fleder and Hosanagar, 2009; Nguyen et al.,

⁵As is the case in our context since the platform is noncommercial.

⁶In a companion paper, we provide a scalable procedure for collecting such data, enabling their integration into platform recommender systems (Aridor et al., 2024).

2014; Brynjolfsson et al., 2011; Hosanagar et al., 2013; Lee and Hosanagar, 2019; Holtz et al., 2020; Korganbekova and Zuber, 2023) with relevant implications for market competition and product variety, as well as the emergence of filter bubbles and echo chambers.

Our paper differs from both of these streams of literature by identifying the effect of *recommendations* produced by a recommender system on consumption via good-level randomisation and by examining the mechanisms through which recommendations operate. Existing work compares consumption between personalised and non-personalised (i.e., popularity-based) recommendations. This experimental design is natural for assessing the impact of the recommendation algorithm on behaviour, but *not* for isolating the role of or discerning the mechanisms for what the recommendation itself is doing.⁷ Our experimental design instead focuses on isolating the role of the recommendation itself by having a control and a consideration-only group of goods *that would have been recommended*, which allows us both to discern the causal effect of recommendation and to disentangle its mechanisms. Indeed, the beliefs data we collect allows us to explain why recommendations may influence consumption – crucial for understanding its welfare effects and for guiding design. Finally, this allows us to explore heterogeneity in the effects of recommendation across consumers.

Two related papers to ours within this stream of literature are Kawaguchi et al. (2021) and Chen et al. (2023). Kawaguchi et al. (2021) decompose recommendation effects into attention and utility in a vending machine context. Our approach differs in two key ways: we directly measure information provision by eliciting consumer beliefs, and we focus on online platforms, where mechanisms differ significantly from the time-pressure dynamics emphasised in vending machines. A more recent paper, Chen et al. (2023), studies the effects of recommendations on consumer search patterns using a large-scale field experiment that changes the quality of recommendations and measures search intensity. In contrast, our work directly measures the informational gains from recommendations and separates these gains from their effects on consideration, providing a clearer understanding of their causal impact by controlling for selection rather than relying on exogenous variation in quality.

Recommender System Evaluation. Our work contributes to the computer science literature on evaluating recommendation quality and determining which recommendations to present to consumers. Early studies in this field recognised that goods with the highest predicted ratings are not always the most useful recommendations (McNee et al., 2006). Since then, various metrics

⁷This is since a well-implemented recommender system should, by design, result in a treatment group with systematically higher idiosyncratic quality relative to the control group, which biases measurement of the effect of recommendation itself.

have been proposed to assess recommendation quality by combining predicted consumer ratings with their past consumption history. These metrics can be categorised into two groups: those that define good recommendations as ones that expand consumers’ consideration sets, such as coverage and novelty (Kaminskas and Bridge, 2016; Castells et al., 2015), and those that prioritise providing “unexpected and useful” information on match value, such as serendipity (Kotkov et al., 2016). The rationale behind these metrics is that they implicitly incorporate theories about how recommendations influence consumer decision-making processes. Our work contributes to this literature by formalising how recommendations influence decision-making through an economic lens and by using a field experiment to quantify the relative importance of these mechanisms, ultimately guiding the choice between different evaluation metrics.

Other Marketing Tools. There is a connection between the economic mechanisms that drive the effect of online recommendations and other marketing tools, such as advertising and consumer reviews. We defer the full discussion of the relationship between these to [Appendix C](#), but broadly argue that, despite sharing common mechanisms, recommendations are generated via an economically distinct process which leads to a relative difference in the importance of different mechanisms.

2. Framework and Hypotheses

2.1. Framework

This section outlines a simple theoretical framework for modelling consumption decisions in the presence of recommendations. Throughout, we focus on the case of experience goods since it represents a common setting in which recommendation systems are deployed and is a good match for the environment of our experimental intervention. Our experiment will allow us to test both assumptions and implications of the model.

Consumers. We consider a consumer i who, in each period $t = 0, 1, \dots$, makes a choice c_t , consisting of a good chosen from a finite set of goods X or an outside option. We denote consumer i ’s consumption history at time t by $C_{i,t} := \{c_{i,\ell} \in X, \ell \leq t\}$, and assume that consumers choose each good at most once.

Match Value and Consumers’ Beliefs. Different consumers may value the same good differently. Each good x has an idiosyncratic match value $v_{i,x}$ for each consumer i , reflecting their subjective assessment of its idiosyncratic value. Prior to consuming a good, consumers are uncertain about its match value. We model this by assuming $(v_{i,x})_x$ are jointly normally distributed,

with mean $(v_{i,x}^b)_x$ and covariance $\Sigma_i = [\sigma_{i,x,x'}^b]$, where $\sigma_{i,x,x'}^b \geq 0$. Upon consuming a good, consumer i fully resolves uncertainty regarding its value, $v_{i,x}$.

Consideration and Exposure. Consumers may not consider all goods at all times. We denote consumer i 's consideration set at time t as $\Gamma_{i,t}$, which – following, e.g., [Goeree \(2008\)](#) and [Masatlioglu et al. \(2012\)](#) – we take to represent the subset of goods the consumer is considering. This can be due to limitations in working memory or simply being unaware of all options; similarly to existing literature,⁸ we do not take a stance on how the consumer's consideration set is formed. We focus instead on how recommendations may affect it.

Platforms manipulate consideration sets by prominently displaying specific goods and forcing consumers to pay attention to them. We consider such manipulations as $e_{i,x,t} = 1$ to denote the situation in which the recommender system forces exposure of consumer i to good x at time t . Forced exposure implies consideration ($e_{i,x,t} = 1 \implies x \in \Gamma_{i,t}$), but not the converse: goods that consumers were forcibly exposed to are a subset of the goods in their consideration set. This mechanism is thus similar to what has been considered in examining the effects of advertising (see, e.g., [Goeree, 2008](#)).

Learning and Recommendation. Recommender systems typically affect consumers' information beyond the formation of their consideration set. Since recommendations correspond to prominently featuring a subset of goods as being of high predicted match value ("top picks for you"), being recommended can itself signal a higher (match) value. Indeed these recommendations are driven by the recommender system's personalised estimate of the value of each good x for consumer i , denoted $v_{i,x}^p$, which is also often displayed. We will term such signal the *platform's predicted value*. We model this direct information provision and denote a recommendation for good x to consumer i at time t by $r_{i,x,t} = 1$ and assume that the platform discloses its consumer-idiosyncratic predicted value $v_{i,x}^p$. For simplicity, $v_{i,x}^p$ is modelled as a Gaussian signal centred on $v_{i,x}$.

In addition to providing direct information about the value of a good, recommendations can have a further informational effect. Specifically, due to prominent positioning, recommendations reduce information acquisition costs by making it easier to learn about the good's attributes. We consider both of these components as the informational aspect of recommendation and do not decompose their relative importance.

⁸A recent literature focuses on models of consideration set formation which give rise to probabilistic consideration of goods, e.g., [Caplin et al. \(2019\)](#), [Cattaneo et al. \(2020\)](#), and [Abaluck and Adams-Prassl \(2021\)](#). We do not consider specific mechanisms through which consideration sets are refined, such as through shortlisting ([Manzini and Mariotti, 2014](#)), or costly search ([Honka et al., 2017](#)).

Choice. Consumer i evaluates match value according to an increasing and concave utility function $u_i : \mathbb{R} \rightarrow \mathbb{R}$, capturing the consumer's attitudes toward uncertainty. We assume the utility associated with the outside option available at time t , $u_{i,o}$, is distributed according to some distribution F , independently across match values and across periods, and known to the consumer before making their choice at time t . Then, given their information at time t , consumer i at time t chooses to maximise current expected utility, i.e., $c_{i,t} \in \arg \max_{y \in \Gamma_{i,t} \cup \{o\}} \mathbb{E}_{i,t}[u_i(v_{i,y})]$.

2.2. Hypotheses

Recommendations affect consumption through multiple informational channels. One important issue is that, by recommending good x to consumer i , the platform is also exposing the consumer to that good and therefore forcing consideration ($r_{i,x,t} = 1 \implies e_{i,x,t} = 1 \implies x \in \Gamma_{i,t}$).

Our first hypothesis is that forcing consideration has on its own a positive impact on consumption, but recommendations affect it further:

Hypothesis 1. *Forcing consideration of a good increases its consumption; recommending it increases consumption further.*

Our remaining hypotheses consider whether recommendations' effect on consumption can be partly explained by their role in directly providing information about the goods' match value. Underlying our theoretical framework is the assumption that recommendation affects consumption by affecting consumers' beliefs, which in turn explain consumption patterns. We decompose our analysis in two steps.

First, we test whether beliefs are sensibly related to consumption. In particular, one would expect that consumers are more likely to choose goods they believe have a higher expected value, $v_{i,x,t}^b := \mathbb{E}_{i,t}[v_{i,x}]$, which in our framework corresponds to utility being increasing in value. On the other hand, if consumers are uncertainty-averse (concave utility), then, all else equal, they are more likely to choose goods about whose match value they are less uncertain. We then have:

Hypothesis 2. *Goods with higher expected match value and lower uncertainty are more likely to be consumed.*

Then, we test the effect of recommendations on consumers' beliefs. In our theoretical framework, we assumed that the recommendation of good x to consumer i at a given time provides a noisy signal of the true match value, that is, $v_{i,x}^p = v_{i,x} + \sigma_{i,x}^p \epsilon_{i,x}^p$, where $\epsilon_{i,x}^p \sim N(0,1)$.⁹ This not only

⁹The recommendation system provides recommendations that are among the platform's best guesses for the consumer's preferred good and as such we expect updating to mostly be in the positive direction. However, this is not

leads to the consumer being more certain about their valuation of good x but also drives their expected value toward the signal. We then posit the following:

Hypothesis 3. *Recommending a good (i) makes consumers less uncertain and (ii) drives their beliefs towards the platform’s predicted match value.*

3. Experimental Design

In order to study if and how recommendations impact consumption we conduct an experimental intervention on a movie recommendation online platform, MovieLens. Our intervention has two main features: (i) we generate random variation in recommendations to study their causal effect on consumption, and (ii) we elicit belief data about good match value prior to consumption to examine the informational mechanisms through which recommendation acts. In this section, we provide background information on the platform and describe our experimental procedures.

3.1. Background on the Recommendation Platform

MovieLens is a movie-recommendation platform created in 1997. It is used by consumers to obtain information about movies as well as personalised movie recommendations based on their ratings. The platform has been widely used, and its movie ratings data are a central benchmark in the recommender system community for the evaluation of new recommender system algorithms.¹⁰

The platform’s home page displays movies organised by categories in rows, with the very first one showing eight “top picks”, the platform’s top recommended movies for the user. Movies are set in a grid fashion, with their poster, title, and the platform-predicted rating for the user.¹¹ The platform-predicted ratings are personalised to each user and obtained through a sophisticated collaborative filtering algorithm combining the user’s and others’ past ratings, reviews, and other metadata (Harper and Konstan, 2015). When hovering over a movie title, users see its genres, their platform-predicted rating, and the average and number of community ratings. Subsequent rows correspond to recent releases, and other categories of potential interest (e.g., “favorites from the past year” or “new additions”).

Consumers use the platform to find movies to watch and to rate the movies after watching. Clicking on a movie’s page provides access to detailed information about the movie, including its

always the case and, for generality, we allow updating to go in either direction.

¹⁰For instance, the search expression “MovieLens dataset” or “MovieLens data” returns over 9,000 entries on Google Scholar, whereas “Netflix dataset” or “Netflix data” – which includes both proprietary data and the well-known public access Netflix Prize competition (Bennett and Lanning, 2007) – returns less than half the number of entries.

¹¹Screenshots of the platform’s interface are included in the [Online Appendix C](#).

trailer, synopsis, cast, associated tags (Vig et al., 2012), and similar movies. It does not provide consumption opportunities, as it does not host movies to stream nor does it direct users to other platforms. This allows us to study consideration effects of recommendations separately from search costs, since recommendations do not affect potential frictions in finding or accessing a particular target consumption good. Furthermore, the platform is free to use and noncommercial, and it is incentive-compatible for users to truthfully report their ratings, as truthful information improves the platform’s recommendation quality.

The recommendation system used by the platform is of high quality and ideal from both a user’s and a researcher’s perspective for several reasons. First, its open-source ratings data are extensively used to develop and evaluate high-quality recommendation algorithms, which ultimately support the quality of the recommendation algorithms MovieLens deploys. Second, the set of algorithms used is transparent, as these are open source and constitute canonical implementations of widely used item-item or singular-value-decomposition collaborative filtering algorithms (see Ekstrand et al., 2011). Finally, MovieLens operates as a noncommercial platform, aligning its recommendation focus with user satisfaction rather than platform profitability. This user-centric approach alleviates concerns about potential biases in user perception of recommendations.

3.2. Experimental Intervention

The platform provides a natural setting to test our hypotheses. We take ratings as a measure of realised good match value and the platform’s predicted rating as a noisy signal; we will use rating and (match) value interchangeably. We take the platform’s ‘top picks’ category as a recommendation of the specific goods listed there.

To understand recommendations’ informational impact, we regularly survey consumers’ beliefs. We identify the effect of recommendations by inducing exogenous randomness and comparing outcomes for recommended goods to those which *would* otherwise be as likely to be recommended.

At the start of the experiment, consumer i ’s 750 goods (not previously consumed) with the highest platform-predicted value are split into three sets: a *control set*, a *consideration-only set*, and a *recommendation set*. We only elicit beliefs about match value for goods in the consideration-only and the recommendation sets, and we restrict recommendations to goods in the latter. We control for consumer’s idiosyncratic tastes by employing stratified block randomisation: we randomly assign the n -th, $(n + 1)$ -th, and $(n + 2)$ -th goods with the highest platform-predicted value to each of the three sets.

Every day that the consumer enters the platform, we elicit their beliefs about the match value of 10 goods not previously consumed.¹² After completing the survey, the consumer is taken to the platform’s home page, in which they are presented with the set of eight recommended goods (their “top picks”). The following procedure summarises how we choose the goods for belief elicitation and recommendation.

Procedure 1 (Belief Elicitation and Recommendation).

- Step 1. Elicit good consumption occurring since the previous visit and remove the consumed goods from the control, consideration-only, and recommendation sets.*
- Step 2. For both the consideration-only and recommendation sets, select uniformly at random two out of the top eight goods, resorted by platform-predicted value plus i.i.d. Gaussian noise.*
- Step 3. Elicit beliefs about the four goods selected in Step 2 in the current period, the four selected in the previous period, and two more selected uniformly at random from the consideration-only set.*
- Step 4. Recommend the top 8 goods in the recommendation set as per the sorting generated in Step 2.*

The algorithm underlying the predicted rating remained the same. Recommendations are therefore credible, as these refer to goods with higher predicted value on average (Steps 2 and 5). As recommended goods are of high predicted value, our sampling procedure elicits beliefs on goods of similar predicted value in the consideration-only set (Step 3). This, together with having beliefs about goods being elicited in two subsequent periods (Step 4), enables us to identify how recommendations impact beliefs. By randomly selecting two goods from the consideration-only set to elicit beliefs on (Step 4), we learn consumers’ beliefs about goods for a broader part of the domain.

Finally, we note that this procedure also allows us to isolate the effects of forced consideration. Specifically, we leverage the unavoidable fact that eliciting beliefs about a good’s match value requires exposing consumers to that good to identify the effect of consideration by comparing outcomes for goods in the consideration-only set and those in the control set.¹³

¹²The survey can be deferred to the following visit to the platform. The set of platform recommendations remains the same until the survey is completed.

¹³We highlight that, via our elicitation survey, our consideration manipulation unambiguously makes the consumer aware of and actively consider the good in question, whereas on the platform consumers may overlook or ignore some recommendations and thus not actively consider them (Zhao et al., 2016). Hence, our estimates likely represent an upper bound on the effect of consideration.

3.3. Measurements

In this section, we provide details on the belief data and consumption measurement.

Beliefs about Match Value. We elicit consumers’ beliefs about match value through a survey they are asked to fill out whenever they return to the platform. In the survey, we elicit beliefs about the match value of 10 movies, selected according to the procedure described in [Section 3.2](#).

For each of the 10 selected movies, the consumer was asked whether or not they had watched it. If they had watched it, we asked for their rating – corresponding to $v_{i,x}$, the idiosyncratic realised match value – and an approximate date of when they watched it. If they had not watched it, we elicited both their expected rating and a measure of uncertainty about their reported expectation.

The expected rating serves as a measure of expected match value, corresponding to $v_{i,x,t}^b$, and is measured on the same scale as the one used to rate movies in the platform (a 10-point Likert scale, ranging from 0.5 to 5). Our measure of uncertainty $\sigma_{i,x,t}^b$ is proxied by asking how sure participants are of their reported expected rating on a 5-point Likert scale, ranging 1-5, where higher values are associated with higher uncertainty about the expected rating. As in our theoretical framework, we assume that, upon consumption, consumers perfectly learn the match value of the good, $v_{i,x,t}^b = v_{i,x}$, and there is no uncertainty, $\sigma_{i,x,t}^b = 0$. We then take the reported ratings as part of our belief data.

The survey interface matches closely the platform’s interface for ratings, except in omitting the predicted rating and not displaying information other than the name and movie poster.¹⁴ Unlike on the homepage, it is not possible for the consumer to hover over or click into the details page to acquire additional information on the good.

We assume belief data were truthfully reported. While the survey was not incentivised, consumers do not have an incentive to misreport: the platform is noncommercial and helping the platform improve its recommendation system is in the consumers’ own interest. The analysis of the belief data provides assurance that the belief data are internally consistent and consistent with the consumers’ behaviour on the platform. In [Appendix A](#), we show that (i) consumers’ initial estimate of expected match value (i.e., the first elicitation before any experimental intervention) on average equals the realised match value rating they provide after consumption, and (ii) the distance between consumers’ expected and realised match value is increasing in our uncertainty measure. We also show that uncertainty about expected match value is decreasing (a) in consumers’ experience, measured by the number of movies rated at the outset of the intervention

¹⁴See [Online Appendix C](#) for interface screenshots.

period, (b) in the movie’s popularity, proxied by the number of community ratings, and (c) on whether the movie is a sequel or part of a franchise; see [Table 4](#).

Good Consumption. The platform does not offer consumption opportunities and therefore there is no direct observation of consumption. The most natural proxy for consumption is to rely on ratings that consumers input on the platform to determine what is consumed and when. Since we are concerned with treatment effects on consumption during our intervention, we exclude ratings referring to consumption that occurred prior to the start of the intervention, as per the report of approximate consumption date in the survey; our results are nevertheless robust to also including consumption occurring prior to the intervention.

3.4. Recruitment and Study Implementation

In our intervention, we targeted a random sample from a subset of the platform’s users.¹⁵ We sought to mitigate the heterogeneity of treatment effects across consumers arising from differences in the quality of the recommendations. As such, we recruited users with a minimum level of engagement, so that the recommender system algorithm used by the platform is able to produce valuable recommendations, while avoiding power users who would constitute extreme outliers. The specific criteria, laid out in [Appendix B](#), were chosen with the platform’s experts.

The intervention lasted from 29 March 2021 to 31 October 2021. It had a phased rollout, including a randomly selected portion of consumers from 29 March 2021 until 15 April 2021, followed by a full rollout to all eligible consumers. The length of the study period was selected based on power calculations and considering the possibly slow rate of consumption of movies over time. The intervention targeted 4,572 eligible individuals, of which 1,452 decided to enroll in the study; we conduct our analysis over these participants. During the intervention, on average consumers in our study visited the platform on 11.67 days, watched 24.02 movies, and completed 2.84 belief surveys.

4. The Impact of Recommendations on Consumption

In this section, we test [Hypothesis 1](#): whether consideration induces additional consumption and whether recommendation further influences consumption beyond consideration. Our baseline specification is as follows:

$$c_{i,x} = \beta_0 + \beta_1 e_{i,x} + \beta_2 r_{i,x} + \epsilon_{i,x} \quad (1)$$

¹⁵The use of or access to the platform is prohibited to individuals under the age of 18, as per the platform’s terms of service.

where $c_{i,x} = 1$ if, during the intervention, consumer i reported having consumed good x (and zero if otherwise), $e_{i,x} = 1$ if good x is in consumer i 's consideration-only or recommendation sets and zero if it is in the control set, and $r_{i,x} = 1$ if x is in i 's recommendation set.

Recall that consideration occurs for consumer i through belief elicitation of good x only if x is in the consideration-only or recommendation sets, and platform recommendation of good x occurs only if x is in the recommendation set. Naturally, the consumer may have other recommendation sources and, even through the platform, may be exposed to goods in either of these sets. Since we stratified our randomisation by consumer-specific tastes, exposure and recommendation to goods via other channels should be orthogonal to treatment assignment.

While this specification enables a clear and straightforward causal estimate of the impact of consideration and recommendation on consumption, it is potentially too conservative. Specifically, it does not take into account the fact that some goods in the recommendation set are never explicitly recommended and some goods in the consideration-only set are never selected for belief elicitation. In order to obtain better estimates on the treatment effects, we consider two additional strategies.

First, we consider the same specification (1), but we restrict the sample to goods in exposed randomisation blocks. While we would like to compare the effect of actual consideration and actual recommendation to the control group, realisations of consideration and recommendation are not independent of platform-predicted match value.¹⁶ In order to resolve this issue, we take not only goods that the consumers were actually exposed to through belief elicitation or recommendation but also goods in their randomisation blocks. Recall that, at the outset of the intervention, goods with the n -th, $(n+1)$ -th, or $(n+2)$ -th highest platform-predicted match value for consumer i are bundled into the same block and block-randomised into the control, consideration-only, and recommendation sets. Then, if good x was recommended ($r_{i,x} = 1$) to consumer i or exposed to it through our experimental intervention ($e_{i,x} = 1$), we include x in the sample, as well as the other goods in the same block used for block-randomisation. We then estimate the same specification (1), but we obtain more precise estimates on the average treatment effect of consideration and recommendation on consumption.

Second, we estimate a variation of the same specification (1), but we restrict the sample to goods that participants were actually exposed to and where they explicitly told us in the elicitation survey that they had not consumed them before. As such, we only keep goods about which beliefs were elicited, that is, those in the consideration-only and recommendation sets. This, together

¹⁶A design restriction, since recommendations need to remain useful and meaningful to consumers.

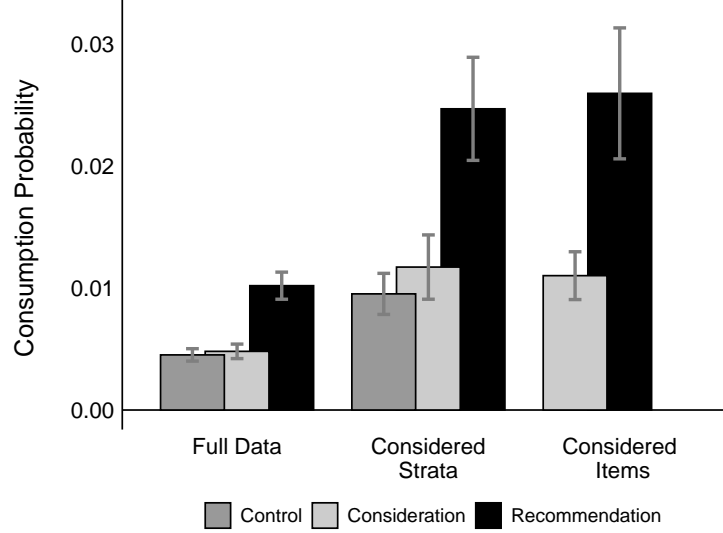


Figure 1: The Impact of Recommendation on Consumption

Notes: This figure tests whether consideration and recommendation impact consumption probability. The estimates correspond to those specified in Equation 1. Each component displays the baseline control and the estimated average treatment effect of consideration and recommendation on consumption for the different sample specifications. Each observation corresponds to a pair (consumer i , good x). “Full Data” includes all consumers i and all goods x in the consumer-specific control, consideration-only, and recommendation sets. “Considered Blocks” includes the goods to which a consumer was exposed to through the belief elicitation survey, and all the goods in the same consumer-specific randomisation block. “Considered Items” includes only goods to which a consumer was exposed through the belief elicitation survey; it necessarily excludes goods in the control set.

with our block randomisation, guarantees that actual recommendation $r_{i,x}$ is orthogonal to the good’s characteristics.

We summarise the results across these different specifications in Figure 1, where full data, considered (randomisation) blocks, and considered items correspond to the three estimation specifications discussed above.¹⁷ Under our preferred specification – restricting to considered blocks – the baseline consumption probability of items in the control group was 1% of the goods in the considered blocks. The effect of consideration alone results in a statistically significant 0.2 percentage point increase, which is a 20% increase relative to the control, providing support for the first part of Hypothesis 1. Furthermore, the effect of recommendation leads to an *additional* 1.3 percentage point increase and results in an aggregate 150% increase relative to the control group, providing support for the second prediction of Hypothesis 1. Overall, even the magnitude of the impact of recommendation on consumption probability is fairly consistent across the different estimation strategies.

¹⁷The regression tables for the different specifications are presented in Table 7 in Online Appendix A.

4.1. Persuasion Rates

Recommendations crucially depend on their credibility, on their ability to alter beliefs and, consequently, behaviour. One natural way of assessing the effect of recommendations on behaviour is by estimating their persuasion rates. Following [DellaVigna and Gentzkow \(2010\)](#), we compute the persuasion rate of consideration and recommendations, respectively, $p_{\text{consideration}}$ and $p_{\text{recommendation}}$, as

$$p_T := \frac{y_T - y_{\text{control}}}{e_T - e_{\text{control}}} \frac{1}{1 - y_{\text{control}}}, \quad (2)$$

where $y_T \in [0, 1]$ denotes the share of goods in group $T \in \{\text{consideration}, \text{recommendation}\}$ which were ultimately consumed and y_{control} that for the control group, while $e_{\text{consideration}}$ denotes the share of goods in the consideration treatment group that were forced consideration, and $e_{\text{recommendation}}$ the share of those in the recommendation group that were actually recommended.

We find that the persuasion rate of consideration is 10.34%, while that of recommendations is 33.34%. Furthermore, if we consider the persuasion rate of consideration absent recommendation, that is, focusing on consideration only, the persuasion rate drastically drops to 0.73%, indicating that the persuasion power is overwhelmingly through recommendations, and that consideration is not a significant force in driving consumption once one factors out actively recommended items.¹⁸ We also note that persuasion rates are relatively large compared to those found in other settings related to the persuasion of consumers (see [DellaVigna and Gentzkow, 2010](#)), in line with the predominance of the recommendation platform’s non-strategic informational role.

5. Recommendations, Beliefs, and Consumption

The results from [Section 4](#) support [Hypothesis 1](#), indicating that, while consideration increases consumption, recommendation has a significantly larger effect. In this section, we examine whether the large effect of recommendation on consumption can be rationalised by an informational mechanism. Namely, we examine the causal impact of beliefs on consumption, and then that of recommendations on beliefs.

5.1. Beliefs Explain Consumption

We start by evaluating the extent to which the belief data explains consumption behaviour. We test [Hypothesis 2](#), which – in line with our theoretical framework – suggests the likelihood of

¹⁸Recall, from our earlier discussion, that we expect this to be an upper bound on the effect of consideration since our intervention forces consideration of each good.

consumption is increasing in the expected match value and decreasing in reported uncertainty.

| | Consumption | | |
|------------------------------------|----------------------|----------------------|---------------------|
| | (1) | (2) | (3) |
| Uncertainty | −0.139*** (0.054) | | −0.126** (0.054) |
| Exp. Match Value | | 0.061*** (0.019) | 0.082*** (0.029) |
| Constant | 0.486*** (0.181) | −0.168*** (0.056) | 0.195 (0.172) |
| Weak Instruments: Uncertainty | 10.07 [0.002] | | 4.08 [0.007] |
| Weak Instruments: Exp. Match Value | | 55.40 [0.000] | 20.20 [0.000] |
| Wu-Hausman | 15.71 [0.000] | 8.39 [0.004] | 11.29 [0.000] |
| Sargan | | | 0.18 [0.672] |
| Observations | 20,895 | 20,895 | 20,895 |

Clustered standard errors at the subject level in parentheses.

p-values in squared brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Beliefs Explain Consumption (**Hypothesis 2**)

Notes: This table tests **Hypothesis 2** by estimating the causal effect of a good’s expected match value and uncertainty on whether it is consumed, as per **Equation 3**. It considers only the last time beliefs about a specific good are elicited from a given consumer. The instruments are as described in **Section 5.1**; Weak Instruments, Wu-Hausman, and Sargan, correspond to tests regarding weak instruments, endogeneity, and validity of overidentifying restrictions.

We evaluate the relationship through the following regression:

$$c_{i,x} = \beta_1 v_{i,x}^b + \beta_2 \sigma_{i,x}^b + \epsilon_{i,x} \quad (3)$$

where $v_{i,x}^b$ and $\sigma_{i,x}^b$ denote the last elicitation of consumer i ’s expected match value and uncertainty associated with good x .

In order to enable a causal interpretation of the relationship between beliefs and consumption, we rely on an instrumental variables approach. We instrument expected match value and uncertainty with two variables: (1) whether a belief elicitation occurred before or after a recommendation (à

la Haaland et al. (2023)), exploiting the randomisation of which movies are recommended; and (2) the user’s activity level, proxied by their study opt-in date, leveraging the phased rollout of the experiment and the orthogonality of platform usage to preferences. We include the interaction of both instruments when analysing both endogenous regressors.

Our results are displayed in Table 1. Columns (1) and (2) present the results of using the information provision and enrollment instrumental variables separately for uncertainty and expected match value, respectively. Column (3) presents the results of estimating specification (3) using both of the instrumental variables. The results are consistent across all specifications. Specifically, the estimates suggest that a one-point increase in expected match value and decrease in uncertainty lead to, respectively, a 8.2 and a 12.6 percentage point increase in consumption probability, lending support for Hypothesis 2. These estimates are economically meaningful as they suggest that the large effect sizes observed in Section 4 can be rationalised if recommendation induces changes to consumer beliefs.

We perform standard diagnostic tests to assess the validity of the instruments (Wooldridge, 2010). The F-statistics confirm that both IVs are strong, individually and jointly: 10.07 for uncertainty (Column 1), 55.40 for expected match value (Column 2), and 4.08 and 20.20, respectively, when combined with interactions (Column 3). A Wu-Hausman test reveals significant correlations ($p < 0.01$) between the endogenous regressors (uncertainty, expected match value) and the error term in all specifications, indicating that OLS estimates are biased and IV estimation is necessary. Supporting this, Table 9 in the Online Appendix A shows that OLS estimates significantly underestimate the effects of uncertainty and expected match value on consumption. Finally, the Sargan test in Column 3 yields a nonsignificant result ($p = 0.672$), supporting the validity of the exclusion restriction.

5.2. The Impact of Recommendations on Beliefs

Now that we have established that beliefs causally determine consumption, we explore the extent to which recommendation impacts these beliefs as a possible explanation for the increase in consumption resulting from recommendation. The information conveyed by recommendations about potentially high match value goods can lead to belief updating and subsequently influence consumption decisions. This particular informational channel of recommendations constitutes the crux of Hypothesis 3.

We examine two different aspects. We first test the first-order effect of whether recommendations provide information, that is, whether they decrease uncertainty. Second, we test if recommen-

| | Δ Uncertainty | Δ Consumer-Platform Exp. Match Value Gap |
|----------------|----------------------|--|
| | (1) | (2) |
| Recommendation | −0.063*** (0.013) | −0.015** (0.007) |
| Constant | −0.055*** (0.008) | 0.004 (0.004) |
| Observations | 21,283 | 21,283 |
| R ² | 0.001 | 0.0002 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: The Impact of Recommendation on Beliefs (**Hypothesis 3**)

Notes: Columns (1) and (2) test **Hypothesis 3(i)** and **3(ii)**, respectively, by estimating the effect of recommendations on the change in uncertainty $\Delta\sigma_{i,x}^b$ and on the change of the absolute difference between the consumer’s expected value and the platform’s predicted value, $\Delta|v_{i,x}^b - v_{i,x}^p|$, as per **Equation 4**. The change is taken to be over the course of the experiment, considering the first and last beliefs reported for each consumer i and good x .

dations drive consumers’ expected match value assessments closer to the platform’s predicted match value, that is, whether recommendations decrease the consumer-platform expected match value gap. We estimate the following specification:

$$y_{i,x} = \beta_0 + \beta_1 r_{i,x} + \epsilon_{i,x} \quad (4)$$

where $r_{i,x}$ is an indicator that equals 1 if good x was recommended to consumer i and is otherwise 0, and $y_{i,x} = \Delta\sigma_{i,x}^b$ or $\Delta|v_{i,x}^p - v_{i,x}^b|$ which denote, respectively, the change in consumer i ’s uncertainty about good x ’s value and the change in the consumer-platform match value gap, i.e., the difference between consumer i ’s expectation of good x ’s match value, $v_{i,x}^b$, and the platform’s predicted match value of good x for consumer i , $v_{i,x}^p$. The change is taken to be over the course of the experiment, considering the first and last beliefs reported for each consumer i and good x .

Since goods selected for belief elicitation are of the same expected (high) match value, regardless of whether they are recommended or not, we are able to identify a causal effect of recommendations on consumer beliefs. Columns (1) and (2) of **Table 2** present our estimates for equation (4) with $y_{i,x} = \Delta\sigma_{i,x}^b$ and $y_{i,x} = \Delta|v_{i,x}^p - v_{i,x}^b|$, respectively. We find support for **Hypothesis 3**, with recommendations decreasing uncertainty for the recommended good and decreasing the match

value difference between consumers' expectations and the platform's predictions. Recommending an item significantly decreases uncertainty by .06 points, and closes the match value gap by .015 points. Put together with the effect sizes estimated in [Section 5.1](#), this shows that the increase in consumption due to recommendations can be rationalised through its effect on beliefs.

6. Heterogeneous Effects and Robustness

In this section, we explore heterogeneous effects of recommendation and conduct different robustness exercises.

6.1. Heterogeneous Effects

We first examine how the treatment effect depends on prior beliefs and previous experience. If their informational role is a main mechanism through which recommendations operate on consumption, then recommendation should have a greater impact when consumers are less certain about their valuation of a good and when they are less experienced with the product space.

Prior Beliefs. In [Figure 2](#), we consider how consumption probability for goods in the consideration and recommendation treatment groups depends on the prior level of uncertainty (panel (a)) and expected match value (panel (b)). The figure highlights that, no matter which prior beliefs, recommendation always increases consumption probability beyond mere consideration. Furthermore, it indicates that lower uncertainty and higher expected match value are associated with higher consumption probability. Finally, [Figure 2](#) provides suggestive evidence that the effect of expected match value on consumption is virtually unchanged by recommendation (a level change), while recommendations strengthen the relationship between uncertainty and consumption (a steeper relationship).

In order to better understand the heterogeneity based on prior beliefs, we assess if and to what extent consumers differentially updated their beliefs based on their baseline uncertainty levels. To do so, we reestimate [Equation 4](#):

$$y_{i,x,t} = \beta_0 + \beta_1 r_{i,x} + \epsilon_{i,x} \quad (4)$$

again, where we consider $y_{i,x,t} = \Delta\sigma_{i,x}^b$ and $y_{i,x,t} = \Delta|v_{i,x}^p - v_{i,x}^b|$. Differently from before, we estimate this specification *conditional on prior beliefs*, that is, for each level of prior uncertainty.

The results are presented in [Figure 3](#), in panel (a), for the change in uncertainty ($\Delta\sigma_{i,x}^b$), and, in panel (b), for the change in the expected match value gap ($\Delta|v_{i,x}^p - v_{i,x}^b|$). [Figure 3\(a\)](#) shows that

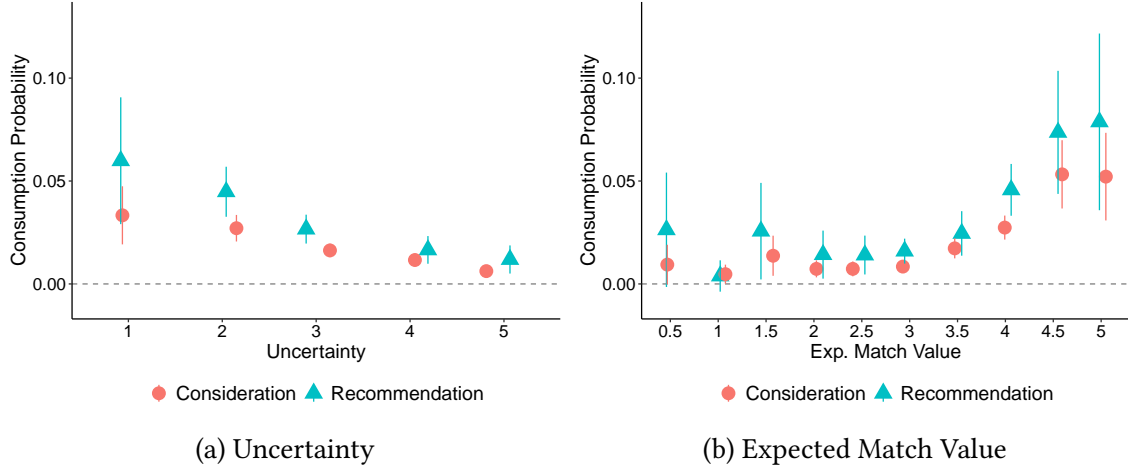


Figure 2: Beliefs Explain Consumption: Heterogeneous Effect by Treatment

Notes: The figure exhibits the estimates of consumption probability by treatment – consideration-only set (red circle) and recommendation set (blue triangle) – and conditional on the reported prior uncertainty (Panel (a)) and the prior expected match value (Panel (b)). Since these estimates are conditional on prior beliefs, only goods about which beliefs were elicited were included. Lines represent 95% confidence intervals with clustered standard errors at the consumer level.

the information provided by recommendation is largely driven by initially high prior uncertainty and that there is little change in uncertainty for elicitations with initially low uncertainty levels. This is consistent with our theoretical framework and intuitively plausible as there is more scope for recommendations to provide informational value the larger the initial uncertainty levels are. In line with this observation, [Figure 3\(b\)](#) indicates that the expected match value gap is brought closer to the platform’s predicted value the higher prior uncertainty is, but overall the estimates are less precise.

Overall, combined with the results of [Figure 2](#), the data suggest that the effects of recommendation are moderated by the degree of prior uncertainty, and thus provide further evidence of the informational role played by recommendations.

Consumer Experience. A related dimension of potential heterogeneity in the effects of recommendations is that of consumers’ past experience. Our underlying working hypothesis is that consumers who have explored a significant portion of the product space are likely to hold more precise beliefs about the match values of goods. Indeed, as mentioned earlier, we find a negative association between prior uncertainty and previous consumer experience (see column (3) of [Table 4](#)). Consequently, the informational gain from recommendations may be lower for such more experienced consumers. We proxy consumer experience using the log of past consumption, specifically the number of movies rated prior to the experimental intervention.

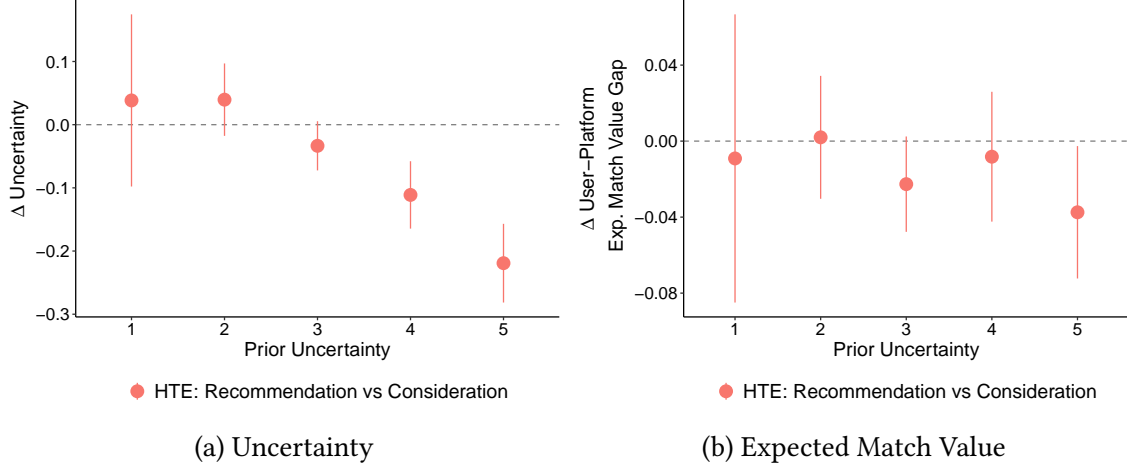


Figure 3: Impact of Recommendations on Beliefs: Heterogeneous Effect by Treatment

Notes: The figure estimates the average treatment effect of recommendations on beliefs compared to that of consideration (Equation 4), conditional on prior uncertainty. Panel (a) exhibits the estimated treatment effect of recommendations (compared to consideration) on uncertainty, whilst Panel (b) displays the treatment effect on the distance between the consumer’s expected value and the platform’s predicted value, both conditional on the reported prior uncertainty. Lines represent 95% confidence intervals with clustered standard errors at the consumer level. The change is taken to be over the course of the experiment, considering the first and last beliefs reported for each consumer i and good x .

To assess whether the extent of past consumption experience mediates how recommendations impact consumption, we interact the recommendation term in specification Equation 1 with our measure of consumer experience. The results, presented in Table 3, use the same specifications considered in Table 1. The estimates indicate a negative heterogeneous effect: recommendations have less of an effect on consumption for more experienced consumers, with recommendations having approximately 20% less impact on consumption for consumers who had watched 500 movies at the outset of the experiment compared to those who had watched only 100.

This conclusion is further supported by considering how persuasion rates change when focusing on consumers with above-median consumption experience at the outset of the intervention. While the persuasion rate for the consideration-only treatment remains similar for more experienced consumers (0.64% compared to 0.73%), the persuasion rate for recommendations drops significantly, from 33.34% in the overall sample to 22.35% for this subset.

In short, recommendations have less sway over consumption for consumers who are more experienced, and thus better informed. These findings are consistent with a decreasing marginal value of information and support the predominance of an informational channel in the effect of recommendations on consumption in our setting.

| | Consumption | | |
|---|----------------------|----------------------|--------------------|
| | (1) | (2) | (3) |
| Consideration | 0.0004** (0.0002) | 0.002** (0.001) | |
| Recommendation | 0.011*** (0.003) | 0.053*** (0.012) | 0.045** (0.019) |
| Recommendation \times Log(Past Consumption) | -0.001* (0.0005) | -0.006*** (0.002) | -0.005* (0.003) |
| Log(Past Consumption) | -0.0001 (0.0003) | -0.001* (0.001) | -0.0004 (0.001) |
| Constant | 0.005** (0.002) | 0.019*** (0.005) | 0.014* (0.007) |
| Observations | 1,026,342 | 56,040 | 20,715 |
| R ² | 0.001 | 0.004 | 0.003 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Heterogeneous Effects of Recommendation by Consumption Experience

Notes: This table tests whether recommendations impact consumption differently depending on consumers' past experience, as proxied by the log of past consumption, that is, the number of movies rated at the outset of the experimental intervention. The estimates correspond to those specified in Equation 1 expanded to include the log of past consumption and its interaction with whether a good was in the recommendation treatment for a particular consumer. Each component displays the baseline control and the estimated average treatment effect of consideration and recommendation on consumption for the different sample specifications.

6.2. Assessing Robustness

We now turn to a discussion of several robustness checks to further validate our findings.

Consumer Fixed Effects. The first possible concern is that there is a large amount of consumer heterogeneity – both in their propensity to consume goods, but also in their frequency of visiting the platform and providing beliefs. As such, it is possible that some of the results that we find are due to idiosyncratic differences across consumers and that the effects are driven by a few power consumers. In Online Appendix B, we rule out this explanation by redoing all of the analyses with consumer fixed effects. We find that this leads to little difference in the economic magnitude and statistical significance of all the analyses conducted in the main portion of the paper.

Normalised Beliefs. The second possible concern is that, along the lines of consumer heterogeneity, a five-star rating for one consumer has a different interpretation than for another consumer.¹⁹ We explore if this influences our conclusions regarding the link between recommendations and beliefs. To this effect, we produce consumer-specific, normalised measures of uncertainty and expected match values: For each consumer, we subtract from the elicited specific measure (uncertainty or expected match value) the consumer-specific mean at the first elicitation and divide by the consumer-specific standard deviation. We then re-estimate [Equations 4](#) using the standardised beliefs. We report the results in [Table 11](#) and in [Table 16](#) including consumer fixed effects – in [Online Appendices B](#) and [C](#), respectively. This leads to no change in our conclusions regarding the magnitude of the effects or their economic/statistical significance.

Timing of Belief Elicitation. The final point we address regards the timing of the belief elicitation considered when analysing the causal relationship between beliefs and consumption. Specifically, in [Table 1](#) we consider only the last elicitation of a given (consumer, good) pair, as it would be the one that is significant for consumption. However, in [Table 8](#) ([Online Appendix A](#)), we show that our conclusions on the causal impact of beliefs on consumption remain robust to considering the first elicitation, instead of the last.

7. Discussion

In this section, we provide a discussion of the implications for the results as well as relevant caveats given the aspects that we cannot capture with our experimental intervention.

Welfare Effects. Our results suggest that recommendations provide consumers with information that decreases uncertainty about the goods and shifts consumption. This does not directly imply that such recommendations are welfare-improving. In particular, by decreasing uncertainty, recommendations could be steering consumers toward goods they like less, generating an informational trap either due to risk aversion or by preempting consumers from seeking further information on alternative products.²⁰ MovieLens, however, does not provide consumption opportunities and deploys recommendation algorithms that are deemed to guide consumers toward ex-post optimal goods. In short, as recommendations are personalised to consumers and target goods with high match value, their effects on shifting consumption suggest an increase ex-post, realised utility, while the reduction of uncertainty provides an immediate increase in expected payoffs for risk-averse consumers.

¹⁹This is a canonical issue for interpreting ratings data since before recommender systems used these data extensively (e.g., see ([Greenleaf, 1992](#))).

²⁰For a recent paper that makes this point clearly in social learning environments, see [Hoezelmann et al. \(2024\)](#).

External Validity. One potential limitation is whether the findings from this analysis generalise to other domains. The intervention was conducted on MovieLens, a platform with two notable features: it lacks profit-driven motives in its recommender system, and it focuses exclusively on a single product type – movies. By contrast, many online platforms are profit-oriented, and recommender systems are deployed across diverse domains, such as news and social media, where consumption processes may differ from those for movies.

Nevertheless, we believe our findings offer meaningful insights for online platforms. First, MovieLens data have been instrumental in advancing research on recommendation algorithms, serving as a benchmark for evaluating new algorithms since the 1990s ([Harper and Konstan, 2015](#)), and inevitably influencing the development of algorithms employed by commercial platforms. Second, the collaborative filtering algorithms used by MovieLens are widely applied across domains, including news, social media, and online marketplaces. While these contexts involve varying consumption behaviours, the conceptual framework and experimental design in this paper provide tools to disentangle underlying mechanisms and estimate their relative importance in environments driven by collaborative filtering algorithms ([Ekstrand et al., 2011](#)).

In short, our work allows for an evaluation of the mechanisms through which recommendations influence choices on a platform foundational to algorithmic development. It serves as a benchmark for assessing the relative magnitudes of economic mechanisms driving choices under a transparent, canonical recommender system, absent profit motives. Furthermore, the approach developed here is portable and can be applied to other contexts to deepen our understanding of how collaborative filtering algorithms influence choices under varying objectives, consumption behaviours, and platform constraints.

Implications for Recommendation System Evaluation. A key distinction in the literature on recommendation system evaluation lies in whether the primary goal is to expand consumers’ consideration sets or to provide better information about goods they might already consider. As mentioned in our discussion of the related literature, both approaches leverage collaborative filtering algorithms to estimate preferences, but they apply different heuristics to generate recommendations. Some systems prioritise surfacing items to broaden consumers’ options, while others focus on enhancing the informational value of recommendations to provide users with more useful insights.

Our findings shed light on the relative importance of these channels by showing that recommendations primarily operate through the informational channel rather than by expanding consideration sets. This suggests that systems designed to inform consumers – by helping them make

better choices from within their existing options – are more effective at driving consumption than those focused on consideration set expansion.

This shift in focus has important implications for recommendation system design. As discussed in the introduction, several metrics are often used as proxies for recommendation informativeness, but they conflate broadening options with providing consumers with information, without directly capturing how recommendations achieve the latter. In our experimental intervention, we elicited belief data to explicitly measure the informational value of recommendations, demonstrating its potential for optimising recommendation systems. In a companion paper (Aridor et al., 2024), we adapt our methodology and introduce a procedure for collecting belief data, suitable for large-scale implementation, along with an open-source dataset from MovieLens that incorporates this information. Together, these contributions help us better understand the economic forces behind consumption choices in recommendation systems and can be used to guide their design and evaluate their impact.

8. References

- Abaluck, J., and A. Adams-Prassl.** 2021. “What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses.” *Quarterly Journal of Economics* 136 (3): 1611–1663.
- Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar.** 2022. “Learning From Reviews: The Selection Effect and the Speed of Learning.” *Econometrica* 90 (6): 2857–2899.
- Ackerberg, D. A.** 2003. “Advertising, learning, and consumer choice in experience good markets: an empirical examination.” *International Economic Review* 44 (3): 1007–1040.
- Aridor, G., D. Goncalves, R. Kong, D. Kluver, and J. Konstan.** 2024. “The MovieLens Beliefs Dataset: Collecting Pre-Choice Data for Online Recommender Systems.” In *Proceedings of the 18th ACM Conference on Recommender Systems*.
- Armstrong, M., J. Vickers, and J. Zhou.** 2009. “Prominence and consumer search.” *The RAND Journal of Economics* 40 (2): 209–233.
- Bennett, J., and S. Lanning.** 2007. “The Netflix Prize.” In *Proceedings of KDD Cup and Workshop*, Volume 2007. 35, New York.
- Bondi, T., M. Rossi, and R. Stevens.** 2022. “The Good, the Bad and the Picky: Reference Dependence and the Reversal of Movie Ratings.” *Working Paper* 1–35.
- Brynjolfsson, E., Y. Hu, and D. Simester.** 2011. “Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales.” *Management Science* 57 (8): 1373–1386.
- Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello.** 2022. “Artificial Intelligence, product recommendations and concentration.” *Working Paper*.

- Caplin, A., M. Dean, and J. Leahy.** 2019. "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice." *Review of Economic Studies* 86 (3): 1061–1094.
- Castells, P., N. J. Hurley, and S. Vargas.** 2015. "Novelty and diversity in recommender systems." In *Recommender Systems Handbook*, 881–918, Springer.
- Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov.** 2020. "What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses." *Journal of Political Economy* 128 (7): 2473–2837.
- Chen, G., T. Chan, D. Zhang, S. Liu, and Y. Wu.** 2023. "The Effects of Diversity in Algorithmic Recommendations on Digital Content Consumption: A Field Experiment." *Working Paper*.
- Chen, Y., F. M. Harper, J. Konstan, and S. X. Li.** 2010. "Social comparisons and contributions to online communities: A field experiment on movielens." *American Economic Review* 100 (4): 1358–1398.
- Chiou, L., and C. Tucker.** 2017. "Content aggregation by platforms: The case of the news media." *Journal of Economics & Management Strategy* 26 (4): 782–805.
- Claussen, J., C. Peukert, and A. Sen.** 2023. "The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News." *Management Science* 70 (3): 1697–1715.
- Das, A. S., M. Datar, A. Garg, and S. Rajaram.** 2007. "Google news personalization: scalable online collaborative filtering." In *Proceedings of the 16th international conference on World Wide Web*, 271–280.
- DellaVigna, S., and M. Gentzkow.** 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics* 2 643–69.
- Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan.** 2018. "Consumer Price Search and Platform Design in Internet Commerce." *American Economic Review* 108 (7): 1820–1859.
- Donnelly, R., A. Kanodia, and I. Morozov.** 2024. "Welfare effects of personalized rankings." *Marketing Science* 43 (1): 92–113.
- Ekstrand, M. D., M. Ludwig, J. Kolb, and J. T. Riedl.** 2011. "LensKit: a modular recommender framework." In *Proceedings of the fifth ACM conference on Recommender systems*, 349–350.
- Fleder, D., and K. Hosanagar.** 2009. "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity." *Management Science* 55 (5): 697–712.
- Freyne, J., M. Jacovi, I. Guy, and W. Geyer.** 2009. "Increasing engagement through early recommender intervention." In *Proceedings of the third ACM conference on Recommender systems*, 85–92.
- Goeree, M. S.** 2008. "Limited Information and Advertising in the U.S. Personal Computer Industry." *Econometrica* 76 (5): 1017–1074.
- Greenleaf, E. A.** 1992. "Improving rating scale measures by detecting and correcting bias components in some response styles." *Journal of Marketing Research* 29 (2): 176–188.
- Grossman, G. M., and C. Shapiro.** 1984. "Informative advertising with differentiated products." *The Review of Economic Studies* 51 (1): 63–81.
- Haaland, I., C. Roth, and J. Wohlfart.** 2023. "Designing information provision experiments."

Journal of Economic Literature 61 (1): 3–40.

Hagiu, A., and B. Jullien. 2011. “Why do intermediaries divert search?” *The RAND Journal of Economics* 42 (2): 337–362.

Harper, F. M., and J. A. Konstan. 2015. “The MovieLens datasets: History and context.” *ACM Transactions on Interactive Intelligent Systems (TIIS)* 5 (4): 1–19.

Hoezelmann, J., G. Manso, A. Nagaraj, and M. Tranchero. 2024. “The Streetlight Effect in Data-Driven Exploration.” *NBER Working Paper* 1–33.

Holtz, D., B. Carterette, P. Chandar, Z. Nazari, H. Cramer, and S. Aral. 2020. “The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify.” *Working Paper* 1–57.

Honka, E., A. Hortaçsu, and M. A. Vitorino. 2017. “Advertising, consumer awareness, and choice: Evidence from the US banking industry.” *The RAND Journal of Economics* 48 (3): 611–646.

Horton, J. J. 2017. “The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment.” *Journal of Labor Economics* 35 (2): 345–385.

Hosanagar, K., D. Fleder, D. Lee, and A. Buja. 2013. “Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation.” *Management Science* 60 (4): 805–823.

Kaminskas, M., and D. Bridge. 2016. “Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems.” *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7 (1): 1–42.

Kawaguchi, K., K. Uetake, and Y. Watanabe. 2021. “Designing context-based marketing: Product recommendations under time pressure.” *Management Science* 67 (9): 5642–5659.

Kleinberg, J., S. Mullainathan, and M. Raghavan. 2023. “The challenge of understanding what users want: Inconsistent preferences and engagement optimization.” *Management Science* 69 (11): 7353–7375.

Korganbekova, M., and C. Zuber. 2023. “Balancing user privacy and personalization.” *Working Paper*.

Kotkov, D., S. Wang, and J. Veijalainen. 2016. “A survey of serendipity in recommender systems.” *Knowledge-Based Systems* 111 180–192.

Lee, D., and K. Hosanagar. 2019. “How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment.” *Information Systems Research* 30 (1): 239–259.

Levy, R. 2021. “Social Media, News Consumption, and Polarization: Evidence from a Field Experiment.” *American Economic Review* 111 (3): 831–870.

Lewis, R. A., and J. M. Rao. 2015. “The Unfavorable Economics of Measuring the Returns to Advertising.” *Quarterly Journal of Economics* 130 (4): 1941–1973.

Manzini, P., and M. Mariotti. 2014. “Stochastic Choice and Consideration Sets.” *Econometrica* 82 (3): 1153–1176.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay. 2012. “Revealed Attention.” *American Economic*

Review 102 (5): 2183–2205.

McNee, S. M., J. Riedl, and J. A. Konstan. 2006. “Being accurate is not enough: how accuracy metrics have hurt recommender systems.” In *CHI’06 extended abstracts on Human factors in computing systems*, 1097–1101, ACM.

Meurer, M., and D. O. S. II. 1994. “Informative advertising and product match.” *International Journal of Industrial Organization* 12 (1): 1–19.

Nguyen, T. T., P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan. 2014. “Exploring the filter bubble: the effect of using recommender systems on content diversity.” In *Proceedings of the 23rd International Conference on World Wide Web*, 677–686, ACM.

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 (6): 1944–1971.

Resnick, P., and H. R. Varian. 1997. “Recommender systems.” *Communications of the ACM* 40 (3): 56–58.

Rossi, M. 2021. “Quality Disclosures and Disappointment: Evidence from the Academy Awards.” In *Proceedings of the 22nd ACM Conference on Economics and Computation*, 790–791.

Sahni, N. S., and H. S. Nair. 2020. “Does advertising serve as a signal? Evidence from a field experiment in mobile search.” *The Review of Economic Studies* 87 (3): 1529–1564.

Senecal, S., and J. Nantel. 2004. “The influence of online product recommendations on consumers’ online choices.” *Journal of Retailing* 80 (2): 159–169.

Steck, H. 2018. “Calibrated recommendations.” In *Proceedings of the 12th ACM Conference on Recommender Systems*, 154–162.

Tadelis, S. 2016. “Reputation and feedback systems in online platform markets.” *Annual Review of Economics* 8 321–340.

Tsai, Y.-L., and E. Honka. 2021. “Informational and Noninformational Advertising Content.” *Marketing Science* 40 (6): 1030–1058.

Ursu, R., A. Simonov, and E. An. 2023. “Online Advertising as Passive Search.” *Marketing Science* 42 (3): 377–401.

Vig, J., S. Sen, and J. Riedl. 2012. “The tag genome: Encoding community knowledge to support novel interaction.” *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2 (3): 1–44.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Zhao, Q., S. Chang, F. M. Harper, and J. A. Konstan. 2016. “Gaze prediction for recommender systems.” In *Proceedings of the 10th ACM Conference on Recommender Systems*, 131–138.

Zhou, R., S. Khemmarat, and L. Gao. 2010. “The impact of YouTube recommendation system on video views.” In *Proceedings of the 10th ACM SIGCOMM conference on Internet Measurement*, 404–410.

Appendix.

Appendix A. Belief Data Validation Exercises

In this appendix, we demonstrate that the belief data we collect not only exhibit reasonable patterns, but are also informative about the resulting match value. In other words, we provide evidence that consumers have well-formed beliefs about films and that survey-based measures can accurately capture them.

| | Realised Match Value | Prior Consumer-Platform Exp. Match Value Gap | Prior Uncertainty |
|-----------------------|----------------------|---|----------------------|
| | (1) | (2) | (3) |
| Exp. Match Value | 1.017*** (0.010) | | |
| Uncertainty | | 0.073*** (0.023) | |
| Log(Past Consumption) | | | -0.465*** (0.060) |
| Log(Popularity) | | | -0.056* (0.033) |
| Film is Sequel | | | -0.061*** (0.010) |
| Constant | | 1.128*** (0.089) | 4.167*** (0.227) |
| Observations | 408 | 21,283 | 20,788 |
| R ² | 0.460 | 0.008 | 0.020 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Properties of Belief Data

Notes: This table demonstrates the sensible patterns in the belief data. Column (1) estimates the correlation between realised match value and expected match value. Column (2) estimates the relationship between the distance between expected and realised match value on prior consumer uncertainty. Column (3) displays the relationship between consumer uncertainty and consumers' log of past consumption as given by the number of films rated at the outset of the experimental period, films' log popularity as given by the number of ratings at the outset of the experimental period on the MovieLens platform, and whether the film is a sequel or not. Prior beliefs (uncertainty and expected match value) refer to the first belief elicitation about a particular film from a particular consumer.

First, we show that consumers' beliefs are an *unbiased* statistic for their value assessments after consumption, arguably settling any question about the validity of the expected match value measure. We estimate:

$$v_{i,x} = \beta_1 v_{i,x}^b + \epsilon_{i,x}$$

where $v_{i,x}$ denotes the realised match value of good x for consumer i , and, in order to capture the initial beliefs of consumers before any intervention, $v_{i,x}^b$ denotes the consumers' first belief elicitation about good x for consumer i . The results, in column (1) of [Table 4](#), show the estimated coefficient β_1 is a precisely estimated 1: prior beliefs of consumers are on average correct.

Second, we show that the (Euclidean) distance between the expected match value assessment and the realised match value is increasing in the reported uncertainty level. Specifically, we estimate:

$$|v_{i,x} - v_{i,x}^b| = \beta_0 + \beta_1 \sigma_{i,x}^b + \epsilon_{i,x}.$$

Column (2) of [Table 4](#) reports $\beta_1 > 0$, a positive relationship between uncertainty and the resulting difference, validating that larger uncertainty results in less aligned belief and actual match value.

We then explore the relationship between expected match value and uncertainty and show in [Figure 4](#) that, as one would expect, consumers are more certain of extreme value assessments (i.e., close to 0 or 5 stars) relative to more moderate ones (i.e., 3 stars).

The final validity check that we conduct explores how uncertainty relates to the popularity of a good (measured by the number of community ratings on MovieLens), the past consumption experience of a consumer (measured by their number of pre-experiment ratings), and whether the film is a sequel or not (measured by joining to IMDb data). We therefore run the following regression to assess the association between consumers' reported uncertainty and these measures:

$$\sigma_{i,x}^b = \beta_1 \log(\text{Past Consumption}_i) + \beta_2 \log(\text{Popularity}_x) + \beta_3 \text{Sequel}_x + \epsilon_{i,x} \quad (5)$$

where the notation is similar to the previous specifications. To isolate the possible role of the experimental intervention in modifying beliefs, we restrict focus to the first belief elicitation of a given good for each consumer.

Column (3) of [Table 4](#) displays the results, showing that greater popularity and being a sequel are associated with lower uncertainty. Interestingly, it also shows that uncertainty is decreasing in past consumption – indicating that less experienced consumers are, on average, more uncertain about their expected match value for goods.

We conclude the validation section by presenting summary statistics about the belief data and the experiment, presented in [Table 5](#). [Table 5](#) highlights the differences between the realised and expected match values, the uncertainty measures, and the platform's predicted match values for the full set of goods in the experimental intervention versus the goods that show up in recom-

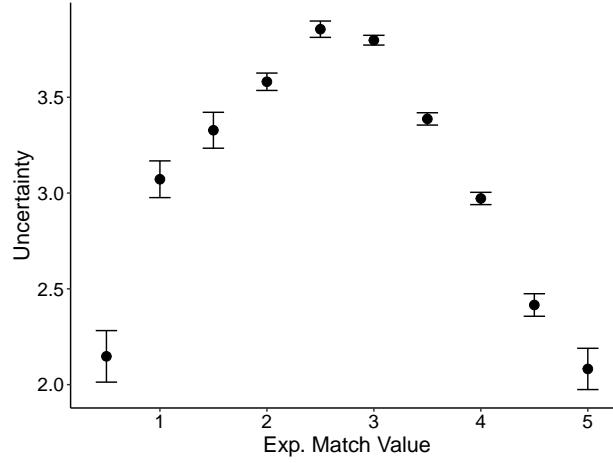


Figure 4: Uncertainty by Expected Quality

Notes: This figure shows the expected match value on the x-axis and the associated conditional average uncertainty score on the y-axis, as well as the associated 95% confidence interval.

mendations. Overall, we collect 39,608 different belief elicitations for 6,674 distinct films and 1,031 distinct goods. During the experimental period, 8,902 films were rated, with 6,240 qualifying as consumption under our robust measure. [Table 6](#) additionally provides summary statistics of the number of completed belief surveys and the number of films consumed per consumer.

| | Mean | Std. Dev. | Percentile | | | | |
|---|------|-----------|------------|------|------|------|------|
| | | | Median | 25th | 75th | Min | Max |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Realised Match Values | 3.71 | 0.88 | 4.00 | 3.50 | 4.00 | 0.50 | 5.00 |
| <i>Avg. per user</i> | 3.77 | 0.59 | 3.83 | 3.50 | 4.16 | 0.50 | 5.00 |
| Exp. Match Values | 3.02 | 0.94 | 3.00 | 2.50 | 3.50 | 0.50 | 5.00 |
| <i>Avg. per user</i> | 3.02 | 0.57 | 3.05 | 2.70 | 3.40 | 0.50 | 4.53 |
| Uncertainty | 3.40 | 1.15 | 3.00 | 3.00 | 4.00 | 1.00 | 5.00 |
| <i>Avg. per user</i> | 3.38 | 0.71 | 3.33 | 2.90 | 3.88 | 1.00 | 5.00 |
| Platform Exp. Match Values | 3.98 | 0.50 | 4.02 | 3.71 | 4.31 | 0.00 | 5.00 |
| <i>Avg. per user</i> | 3.98 | 0.43 | 4.01 | 3.72 | 4.27 | 1.58 | 5.00 |
| Platform Exp. Match Values for Recommended | 4.36 | 0.42 | 4.37 | 4.09 | 4.69 | 2.64 | 5.00 |
| <i>Avg. per user</i> | 4.39 | 0.37 | 4.42 | 4.13 | 4.68 | 2.81 | 5.00 |

Table 5: Belief Data Summary Statistics

Notes: We present summary statistics on the belief data. We show the distributional statistics for the realised and expected match values, the uncertainty measures, the platform’s predicted match values for the films considered in the intervention, and the platform’s predicted match value for the set of goods that show up in recommendations. Within each field, the first row shows the distribution across the full dataset, whereas the second row shows the distribution across consumers (averaging to a single average value for each consumer).

| | Mean | Std. Dev. | Percentile | | | | |
|-------------------------|------|-----------|------------|------|------|------|--------|
| | | | Median | 25th | 75th | Min | Max |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Films Consumed per user | 4.30 | 9.30 | 1.00 | 0.00 | 5.00 | 0.00 | 157.00 |
| Belief Surveys per user | 3.98 | 5.88 | 2.00 | 1.00 | 4.00 | 1.00 | 72.00 |

Table 6: Summary Statistics on Consumption and Beliefs

Notes: This table provides summary statistics on the number of completed belief surveys (each including 10 goods) and the number of films consumed per consumer.

Appendix B. Recruitment and Study Implementation: Additional Details

In our intervention, we target a random sample from a subset of the platform’s users.²¹ In order to mitigate the heterogeneity of treatment effects across consumers arising from differences in the match value of the recommendations, we restrict our sample to users who satisfy the following conditions: (i) having rated more than 100 films in total; (ii) having rated fewer than 3,000 films in total; and (iii) over the previous $m = 1, 2, 3, 4$ months, having rated a minimum of $\lceil 1.5m \rceil$ films. The first condition is a minimum data requirement so that the recommender system algorithm utilised by the platform is able to provide valuable recommendations. This is especially important given that, throughout the duration of the intervention, the assignment of films to treatments is held fixed and therefore so is the set of films that can be recommended. The second excludes power users. The last condition seeks to guarantee that the targeted user is minimally active on the platform over the recent past. These criteria were chosen in consultation with the platform’s experts in order to ensure that the data are representative of the overall platform population, with stable users who are familiar with the platform’s recommender system.

Appendix C. Additional Related Work Discussion

We provide some additional discussion of related work for similar, yet distinct, marketing mechanisms such as advertising and online reviews.

Advertising. There is a vast literature on the economics of advertising that similarly attempts to decompose whether and why advertisements influence consumption. For example, some (e.g. [Honka et al., 2017](#); [Tsai and Honka, 2021](#); [Ursu et al., 2023](#)) argue that advertising in several contexts acts through the consideration channel, and others (e.g. [Grossman and Shapiro, 1984](#); [Meurer and II, 1994](#); [Akerberg, 2003](#); [Sahni and Nair, 2020](#)) focus on its informational effects. The nuances in measuring the effectiveness and the mechanisms behind the impact of advertising, as discussed in [Lewis and Rao \(2015\)](#), motivate our empirical design for measuring the impact of recommendations. While there are similarities in the underlying mechanisms, advertising and recommendation systems target distinct aspects of the consumer choice process and are generated using different methods and by agents with differing incentives. Recommendation systems aggregate consumer data on a platform to provide predictions for multiple goods present on the platform, while advertisements are aimed at persuading consumers to purchase one good or goods from one brand. Thus, we view one of our contributions as, similarly to the literature on advertising, decomposing the role of the mechanisms that drive the role of this distinct and relatively new tool for online platforms.

Online Reviews. Another source of possible information that guides consumers in their choices comes from online reviews ([Tadelis, 2016](#)). The basis for most recommendation systems, and the one used in this context, is the set of consumer-provided ratings and reviews that they leave

²¹The use of or access to the platform is prohibited to individuals under the age of 18, as per the platform’s terms of service.

on the platform. The primary difference between the two channels is that while reviews provide non-personalised information, a recommendation system provides *personalised* recommendations tailored to the tastes of the consumer. Furthermore, the economic mechanisms behind reviews are primarily informational and lack a consideration component since consumers have to actively seek out the good in order to view its reviews. Consistent with this, the literature on consumer reviews has primarily focused on issues such as differentiating between the effectiveness of online versus professional reviews (Reimers and Waldfogel, 2021) and the feedback loops between consumption and rating (Acemoglu et al., 2022; Bondi et al., 2022). In our setting, there are no reviews with consumer text – only ratings. Recommendations, in contrast, are not only distinct from reviews but are also compelling to study in their own right, as they involve complex personalisation algorithms and economic mechanisms that influence consumer choice in ways fundamentally different from reviews. While reviews are typically complementary to recommendations, they have distinctly different roles in the consumer choice process.

Online Appendix.

Online Appendix A. Additional Tables

| | Consumption | | |
|----------------|----------------------|---------------------|---------------------|
| | (1) | (2) | (3) |
| Consideration | 0.0003* (0.0002) | 0.002** (0.001) | |
| Recommendation | 0.005*** (0.0005) | 0.013*** (0.002) | 0.015*** (0.003) |
| Constant | 0.005*** (0.0003) | 0.010*** (0.001) | 0.011*** (0.001) |
| Observations | 1,083,123 | 57,396 | 21,208 |
| R ² | 0.001 | 0.003 | 0.003 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: The Impact of Recommendation on Consumption (**Hypothesis 1**)

Notes: This table tests whether consideration and recommendation impact consumption probability. The estimates correspond to those specified in **Equation 1** providing a direct counterpart to **Figure 1**. Each component displays the baseline control and the estimated average treatment effect of consideration and recommendation on consumption for the different sample specifications. Each observation corresponds to a pair (consumer i , good x). Column (1) includes the “full data”, that is, it includes all consumers i and all goods x in the consumer-specific control, exposure-only, and recommendation sets. Column (2) includes only “considered strata”, i.e., it includes the goods to which a consumer was exposed through the belief elicitation survey, and all the goods in the same consumer-specific stratum. Column (3) further restricts the sample to only “considered items”, corresponding to goods to which a consumer was exposed through the belief elicitation survey; it necessarily excludes goods in the control set.

| | Consumption | | |
|-------------------------------------|---------------------|----------------------|---------------------|
| | (1) | (2) | (3) |
| Uncertainty | −0.159** (0.067) | | −0.162** (0.075) |
| Exp. Match Value | | 0.062*** (0.019) | 0.091** (0.037) |
| Constant | 0.558** (0.229) | −0.171*** (0.057) | 0.293 (0.226) |
| Weak Instruments (Uncertainty) | 7.69 [0.006] | | 3.23 [0.021] |
| Weak Instruments (Exp. Match Value) | | 55.19 [0.000] | 19.82 [0.000] |
| Wu-Hausman | 15.71 [0.000] | 8.76 [0.003] | 11.92 [0.000] |
| Sargan | | | 0.03 [0.858] |
| Observations | 20,895 | 20,895 | 20,895 |

Clustered standard errors at the subject level in parentheses.

p-values in squared brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Beliefs Explain Consumption (**Hypothesis 2**): 1st Elicitation

Notes: This table tests **Hypothesis 2** by estimating the causal effect of a good's expected match value and uncertainty on whether it is consumed, as per **Equation 3**. It differs from **Table 1** by considering the first instead of the last time that beliefs about a specific good are elicited from a given consumer. The instruments are as described in **Section 5.1**; Weak Instruments, Wu-Hausman, and Sargan, correspond to tests regarding weak instruments, endogeneity, and validity of overidentifying restrictions.

| | Consumption | | |
|------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) |
| Uncertainty | −0.007*** (0.001) | | −0.006*** (0.001) |
| Exp. Match Value | | 0.010*** (0.001) | 0.010*** (0.001) |
| Constant | 0.038*** (0.004) | −0.016*** (0.003) | 0.006 (0.004) |
| Observations | 20,895 | 20,895 | 20,895 |
| R ² | 0.004 | 0.006 | 0.009 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Correlation Between Beliefs and Consumption

Notes: This table estimates the correlation between a good’s expected match value and uncertainty with its consumption probability as per Equation 3. It considers only the last time that beliefs about a specific good are elicited from a given consumer. Differently from Table 1, the estimates do not have a causal interpretation.

| | Consumption | | |
|------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) |
| Uncertainty | −0.006*** (0.001) | | −0.006*** (0.001) |
| Exp. Match Value | | 0.009*** (0.001) | 0.009*** (0.001) |
| Constant | 0.037*** (0.004) | −0.013*** (0.003) | 0.008** (0.004) |
| Observations | 20,895 | 20,895 | 20,895 |
| R ² | 0.004 | 0.005 | 0.008 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Correlation Between Beliefs and Consumption: 1st Elicitation

Notes: This table estimates the correlation between a good’s expected match value and uncertainty with its consumption probability, as per Equation 3. It differs from Table 9 by considering the first instead of the last time that beliefs about a specific good are elicited from a given consumer. Differently from Table 8, the estimates do not have a causal interpretation.

| | Δ Std. Uncertainty | Δ Std. Consumer-Platform Exp. Match Value Gap |
|----------------|---------------------------|---|
| | (1) | (2) |
| Recommendation | -0.067*** (0.015) | -0.026** (0.010) |
| Constant | -0.063*** (0.009) | 0.004 (0.006) |
| Observations | 20,704 | 20,704 |
| R ² | 0.001 | 0.0004 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: The Impact of Recommendation on Beliefs (**Hypothesis 3**) – Standardised Beliefs

Notes: Columns (1) and (2) test **Hypothesis 3(i)** and **3(ii)**, respectively, by estimating the effect of recommendations on the change in uncertainty $\Delta\sigma_{i,x}^b$ and on the change of the absolute difference between the consumer's expected value and the platform's predicted value, $\Delta|v_{i,x}^b - v_{i,x}^p|$, as per **Equation 4**. It differs from **Table 2** by considering standardised measures of beliefs, subtracting from the elicited specific measure (uncertainty or expected match value) the consumer-specific mean at the first elicitation and dividing by the consumer-specific standard deviation.

Online Appendix B. Tables with Consumer Fixed Effects

| | Consumption | | |
|----------------|----------------------|---------------------|---------------------|
| | (1) | (2) | (3) |
| Consideration | 0.0003* (0.0002) | 0.002** (0.001) | |
| Recommendation | 0.005*** (0.0005) | 0.013*** (0.002) | 0.014*** (0.002) |
| Consumer FEs | Yes | Yes | Yes |
| Observations | 1,083,123 | 57,396 | 21,208 |
| R ² | 0.029 | 0.044 | 0.067 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: The Impact of Recommendation on Consumption (**Hypothesis 1**): Consumer Fixed Effects

Notes: This table tests whether consideration and recommendation impact consumption probability. It differs from **Table 7** only in that it includes consumer fixed effects. Each component displays the baseline control and the estimated average treatment effect of consideration and recommendation on consumption for the different sample specifications. Each observation corresponds to a pair (consumer i , good x). Column (1) includes the “full data”, that is, it includes all consumers i and all goods x in the consumer-specific control, exposure-only, and recommendation sets. Column (2) includes only “considered strata”, i.e., it includes the goods to which a consumer was exposed through the belief elicitation survey, and all the goods in the same consumer-specific stratum. Column (3) further restricts the sample to only “considered items”, corresponding to goods to which a consumer was exposed through the belief elicitation survey; it necessarily excludes goods in the control set.

| | Consumption | | |
|------------------|----------------------|---------------------|----------------------|
| | (1) | (2) | (3) |
| Uncertainty | −0.009*** (0.001) | | −0.007*** (0.001) |
| Exp. Match Value | | 0.013*** (0.002) | 0.011*** (0.001) |
| Consumer FEs | Yes | Yes | Yes |
| Observations | 20,895 | 20,895 | 20,895 |
| R ² | 0.070 | 0.072 | 0.074 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Correlation Between Beliefs and Consumption: Consumer Fixed Effects

Notes: This table estimates the correlation between a good’s expected match value and uncertainty with its consumption probability, as per Equation 3. It considers only the first time beliefs about a specific good are elicited from a given consumer. It differs from Table 9 only in that it includes consumer fixed effects.

| | Consumption | | |
|------------------|----------------------|---------------------|----------------------|
| | (1) | (2) | (3) |
| Uncertainty | −0.009*** (0.001) | | −0.007*** (0.001) |
| Exp. Match Value | | 0.012*** (0.002) | 0.010*** (0.001) |
| Consumer FEs | Yes | Yes | Yes |
| Observations | 20,895 | 20,895 | 20,895 |
| R ² | 0.070 | 0.070 | 0.073 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Correlation Between Beliefs and Consumption: 1st Elicitation; Consumer Fixed Effects

Notes: This table estimates the correlation between a good’s expected match value and uncertainty with its consumption probability, as per Equation 3. It differs from Table 9 by considering the first instead of the last time that beliefs about a specific good are elicited from a given consumer. It differs from Table 10 only in that it includes consumer fixed effects.

| | Δ Uncertainty | Δ Consumer-Platform Exp. Match Value Gap |
|----------------|----------------------|--|
| | (1) | (2) |
| Recommendation | -0.055*** (0.013) | -0.015** (0.007) |
| Consumer FEs | Yes | Yes |
| Observations | 21,283 | 21,283 |
| R ² | 0.074 | 0.056 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: The Impact of Recommendation on Beliefs (**Hypothesis 3**): Consumer Fixed Effects

Notes: Columns (1) and (2) test **Hypothesis 3(i)** and **3(ii)**, respectively, by estimating the effect of recommendations on the change in uncertainty $\Delta\sigma_{i,x}^b$ and on the change of the absolute difference between the consumer's expected value and the platform's predicted value, $\Delta|v_{i,x}^b - v_{i,x}^p|$, as per **Equation 4**. The change is taken to be over the course of the experiment, considering the first and last beliefs reported for each consumer i and good x . It differs from **Table 2** only in that it includes consumer fixed effects.

| | Δ Std. Uncertainty | Δ Std. Consumer-Platform Exp. Match Value Gap |
|----------------|---------------------------|---|
| | (1) | (2) |
| Recommendation | -0.058*** (0.015) | -0.025** (0.011) |
| Consumer FEs | Yes | Yes |
| Observations | 20,704 | 20,704 |
| R ² | 0.074 | 0.054 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: The Impact of Recommendation on Beliefs – Standardised Beliefs; Consumer FE

Notes: This table tests **Hypothesis 2** by estimating the causal effect of a good's expected match value and uncertainty on whether it is consumed, as per **Equation 3**. Similar to **Table 11**, it employs standardised measures of beliefs, subtracting from the elicited specific measure (uncertainty or expected match value) the consumer-specific mean at the first elicitation and dividing by the consumer-specific standard deviation. It differs from **Table 11** only in that it includes consumer fixed effects.

| | Consumption | | |
|---|----------------------|----------------------|--------------------|
| | (1) | (2) | (3) |
| Consideration | 0.0004** (0.0002) | 0.002** (0.001) | |
| Recommendation | 0.011*** (0.003) | 0.053*** (0.012) | 0.041** (0.018) |
| Recommendation \times Log(Past Consumption) | -0.001* (0.0005) | -0.006*** (0.002) | -0.004 (0.003) |
| Consumer FEs | Yes | Yes | Yes |
| Observations | 1,026,342 | 56,040 | 20,715 |
| R ² | 0.028 | 0.044 | 0.065 |

Clustered standard errors at the subject level in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Heterogeneous Effects of Recommendation by Consumption Experience: Consumer Fixed Effects

Notes: This table tests whether recommendations impact consumption differently depending on consumers' past experience, as proxied by the log of the past consumption, that is, the number of films rated at the outset of the experimental intervention. It differs from [Table 3](#) only in that it includes consumer fixed effects. The estimates correspond to those specified in [Equation 1](#) expanded to also include the log of past consumption and its interaction with whether or not a good was in the recommendation treatment for a particular consumer. Each component displays the baseline control and the estimated average treatment effect of consideration and recommendation on consumption for the different sample specifications.

Online Appendix C. Screenshots

C.1. Platform Interface

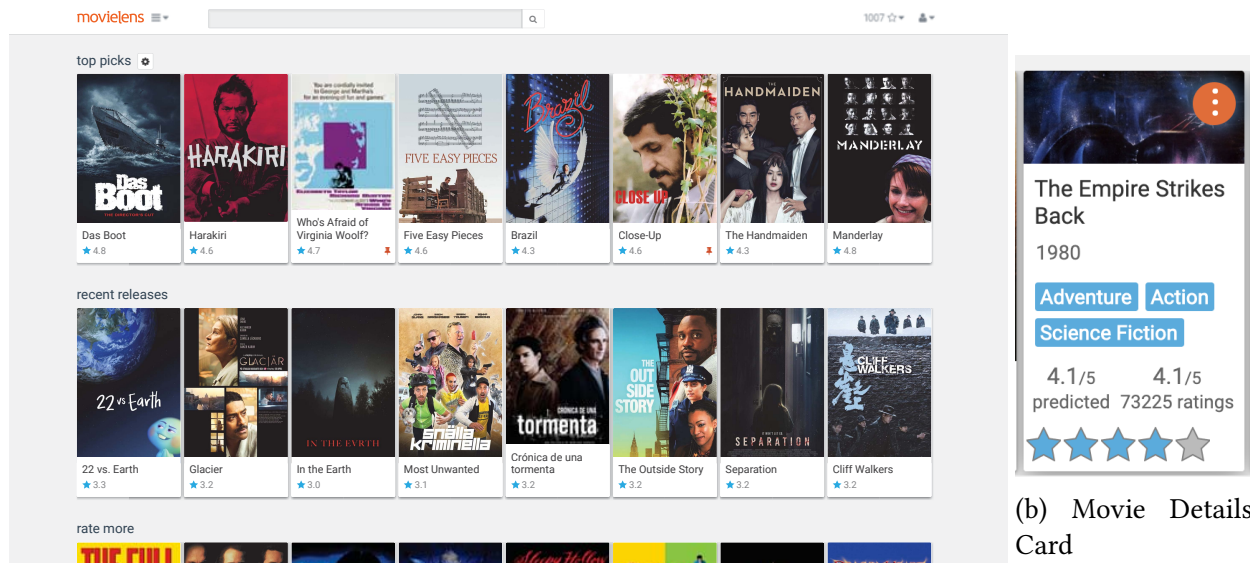




Figure 5: MovieLens Home Page



The Artist

★★★★★
🔍
✕

MovieLens predicts for you
 4.11 stars
 Average of 3,775 ratings
 3.73 stars
 


Genres
 Drama , Comedy , Romance

Links
[imdb](#), [tmdb](#)

Hollywood, 1927: As silent movie star George Valentin wonders if the arrival of talking pictures will cause him to fade into oblivion, he sparks with Peppy Miller, a young dancer set for a big break.

2011 100 minutes
Languages
 No Language
Directors
[Michel Hazanavicius](#)
Cast
[Jean Dujardin](#), [Bérénice Bejo](#), [John Goodman](#), [James Cromwell](#), [Penelope Ann Miller](#), [more...](#)

Community Tags

view: [top](#) [all](#) sort by: [relevance](#) 

×29 silent movie +

×12 dogs +

×6 cinematography +

×3 Oscar Winner: Best +

×28 original +

×14 romance +

×7 Bérénice Bejo +

×9 boring +

×26 visually appealing +

×12 Jean Dujardin +

×6 Stylised +

×3 thin plot +

×22 1920s +

×11 Hollywood +

×4 dog +

×16 movie business +

×11 Oscar (Best Picture) +

×4 silent film +

×13 black and white +

×9 John Goodman +

×3 Oscar Nominee: Best +

Similar Movies [see more](#)



La La Land

★ 5.0



Marriage Story

★ 5.0



Free Solo

★ 4.0



Jean de Florette

★ 4.5



Citizen Kane

★ 3.4



Blue Planet II

★ 4.7



City Lights

★ 3.4



Won't You Be My Neighbor?

★ 4.3

Figure 6: Movie Details Page

C.2. Survey

movielens

logged in as

Part 1

Thanks for joining! Let us know if there are any new movies you have watched seen since the last time you were on MovieLens:

[Not right now. Go to MovieLens.](#)


[Contact Us](#) [Remove me from the survey study group?](#)

movielens


logged in as

Part 1


Thanks for joining! Let us know if there are any new movies you have watched seen since the last time you were on MovieLens:




The Matrix
Have you seen this movie?
☐ No




The Matrix Reloaded
Have you seen this movie?
☐ No



The Matrix Revolutions
Have you seen this movie?
☐ No



The Matrix Revisted
Have you seen this movie?
☐ No



A Glitch in the Matrix
Have you seen this movie?
☐ No

[Not right now. Go to MovieLens.](#)

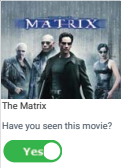
[Contact Us](#) [Remove me from the survey study group?](#)

movielens


logged in as

Part 1


Thanks for joining! Let us know if there are any new movies you have watched seen since the last time you were on MovieLens:




The Matrix
Have you seen this movie?
☒ Yes
How would you rate this movie?
★ ★ ★ ★ ☆ ○
When did you watch it approximately?




The Matrix Reloaded
Have you seen this movie?
☐ No



The Matrix Revolutions
Have you seen this movie?
☐ No



The Matrix Revisted
Have you seen this movie?
☐ No



A Glitch in the Matrix
Have you seen this movie?
☐ No

[Not right now. Go to MovieLens.](#)

[Contact Us](#) [Remove me from the survey study group?](#)

movielens

logged in as


Go to Part 1

Part 2

Please let us know what you think of the following movies. You must complete the questions for every movie to submit the survey.

For movies that you have seen, let us know your rating and when you saw the movie.

For movies that you have not seen, let us know what you think you would rate it and how sure you are about this rating. We are interested in your best guess given your current knowledge, so we ask that you do not look up any of the movies while completing the survey. [\(how are these movies chosen\)](#)



SCHOOL OF LIFE

✓ School of Life

SeenNot Seen

Wintersleepers

SeenNot Seen

How do you think you would rate this movie?

★★★★★

How sure are you?

(1=Very Unsure, 5=Very Sure)

12345

Lake Fear

SeenNot Seen

How do you think you would rate this movie?

★★★★★

How sure are you?

(1=Very Unsure, 5=Very Sure)

12345

The Scarehouse

SeenNot Seen

How would you rate this movie?

★★★★★

When did you watch it approximately?

Click to select a date

Click to select a date

Cyrano, My Love

SeenNot Seen

How would you rate this movie?

★★★★★

When did you watch it approximately?

Click to select a date

Click to select a date

Woman on the Beach

SeenNot Seen

Philippe Garrel

SeenNot Seen

How do you think you would rate this movie?

★★★★★

How sure are you?

(1=Very Unsure, 5=Very Sure)

12345

Take Me Home

SeenNot Seen

To the Left of the Father

SeenNot Seen

Black Gunn

SeenNot Seen

Not right now. Go to Movielens.

Submit Survey

Contact Us

Remove me from the survey study group?

movielens

logged in as


Go to Part 1

Part 2

Please let us know what you think of the following movies. You must complete the questions for every movie to submit the survey.

For movies that you have seen, let us know your rating and when you saw the movie.

For movies that you have not seen, let us know what you think you would rate it and how sure you are about this rating. We are interested in your best guess given your current knowledge, so we ask that you do not look up any of the movies while completing the survey. [\(how are these movies chosen\)](#)



SCHOOL OF LIFE

✓ School of Life

SeenNot Seen

Wintersleepers

SeenNot Seen

How do you think you would rate this movie?

★★★★★

How sure are you?

(1=Very Unsure, 5=Very Sure)

12345

Lake Fear

SeenNot Seen

How do you think you would rate this movie?

★★★★★

How sure are you?

(1=Very Unsure, 5=Very Sure)

12345

The Scarehouse

SeenNot Seen

How would you rate this movie?

★★★★★

When did you watch it approximately?

Click to select a date

Click to select a date

Cyrano, My Love

SeenNot Seen

How would you rate this movie?

★★★★★

When did you watch it approximately?

Click to select a date

Click to select a date

Woman on the Beach

SeenNot Seen

Philippe Garrel

SeenNot Seen

How do you think you would rate this movie?

★★★★★

How sure are you?

(1=Very Unsure, 5=Very Sure)

12345

Take Me Home

SeenNot Seen

To the Left of the Father

SeenNot Seen

Black Gunn

SeenNot Seen

Not right now. Go to Movielens.

Submit Survey

Contact Us

Remove me from the survey study group?

We want to understand how our recommendations affect which movies you watch and how you feel about movies you haven't watched.

The movies that we are asking you about are randomly selected from your top 750 movies. This means that they are not recommended movies: we think you will like some of them and others maybe not at all.

You will have access to your top picks, which are the movies recommended to you, with your personal predicted rating.

You will also notice you may be asked about the same movie more than once. This is to measure if and how your opinion changes with time, as you learn more about those movies and, maybe, about your own taste.

This will help us develop better recommendations for you! Thank you for participating!

Go back to survey

47

C.3. Recruitment

movielens

logged in as

We need YOU!

We are studying how to improve our recommendations and we would like to ask for your help.

We want to understand how our recommendations affect which movies you watch and how you feel about movies you haven't watched.

For the next few months when you log onto MovieLens we will ask you how you think you would rate movies you haven't seen yet and how sure you are about this. We will also ask you about movies you have seen recently.

The movies we ask you about have been randomly selected from the top 750 movies you haven't seen yet, so we are not necessarily recommending you watch them.

After completing the survey, you will be redirected to the main MovieLens interface where you will see your top picks and new releases as usual.

You will be able to exit the experiment at any time.

MovieLens is partnering with researchers from Columbia University and the University of Minnesota and you will be agreeing to taking part in an approved survey.

[Yes, this sounds great!](#) [Ask me later.](#) [No, I'm not interested.](#)

The University of Minnesota IRB reviewed this study and determined that it was exempt from further review. Participants who have questions or concerns can contact the U of M IRB at: irb@umn.edu.

[Documentation: Approval and Consent.](#) [Contact Us](#)