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 recommender 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 RecSys: Evidence from a Field Experiment on MovieLens." We thank Heski Bar-Isaac, Yeon-Koo Che, Mark Dean, Amelia Fletcher, Brett Gordon, Ilya Morozov, Peter Ormosi, Imran Rasul, Andrey Simonov, Greg Taylor, Hal Varian, Jack Willis, 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*: 4 January 2026.

1. Introduction

Recommender systems (RecSys) are nearly ubiquitous in the digital economy. They 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), to curation feeds on social media platforms (Aridor et al., 2024b), to cultural goods on media streaming platforms (Holtz et al., 2020), and to articles 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, determining exactly which recommendations are most useful to consumers has spawned a large debate, primarily about *how* they influence consumption decisions (McNee et al., 2006).

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-subject design that allows us to separately measure the effect of these mechanisms on consumption. We find that 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. This 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. Our experimental design exploits a unique aspect of our data to isolate the role of recommendation as we observe the intermediate outputs of the RecSys 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 consumer-specific quality according to the RecSys estimates.

The second challenge is that consumer beliefs are typically unobserved and that recommendations simultaneously make consumers consider goods and provide information about their 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 by-product of this elicitation is that it makes consumers consider the good, without providing the informational content of recommendation. We exploit the variation induced by this survey to include a third *forced consideration* group of goods for each consumer, which appears only 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.[1,2]

Our experiment is conducted on a movie-recommendation platform, MovieLens, that is non-commercial 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 to produce *consumer-specific* predicted ratings for unrated movies (Harper and Konstan, 2015). These predictions are displayed to users in the first row of the platform homepage and are used to tailor the platform's *consumer-specific* recommendations ("top picks"). As such, the recommendations are personalised and considered high-quality. We maintain the recommendations quality during our intervention by including only the top 750 goods for each consumer as determined by the RecSys estimates.[4]

Our first main finding is that recommendation induces a significant increase in consumption. Our Average Treatment Effect on the Treated (ATT) estimates imply that recommendation induces a substantial 341% increase in consumption beyond the effect of forced consideration alone. In contrast, we find a modest effect of forced consideration on consumption.

Our second set of findings indicates that the larger increase in consumption from recommenda-

[1] We are agnostic whether the informational gains come from direct inferences due to the recommendation or indirect information accrued due to reduced information acquisition costs for recommended goods.

[2] MovieLens does not directly offer consumption opportunities, so our consumption measure relies on a self-reported movie diary or an extended measure that includes platform ratings. We discuss these in Section 4.

[3] 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.

[4] The 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.

tions 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. 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 8.4 and 10.1 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. We conduct additional analysis that shows that the informational role is driven both by reduced information acquisition costs and the (idiosyncratic) quality signals provided by the recommendation. Finally, we find that the informational role of recommendations is larger for less experienced consumers – those with shorter consumption histories – as they are more uncertain in their assessments and experience a larger causal increase in consumption from recommendations.

Understanding the economic mechanisms through which recommendations influence choice is crucial for guiding the design and evaluation of RecSys. 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. 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.[5]

1.1. Related Literature

This paper contributes to a burgeoning literature on the economics of recommender systems.

The Impact of Recommendation on Consumption. Recent literature has examined whether RecSys impact consumption patterns. Senecal and Nantel (2004), Das et al. (2007), Freyne et al. (2009), Zhou et al. (2010), Claussen et al. (2023), and Holtz et al. (2020) show that personalised

[5]In a companion paper, we provide a scalable procedure for collecting belief data (Aridor et al., 2024a).

RecSys, compared to non-personalised benchmarks, meaningfully increase consumption.

Our paper contributes by identifying the effect of *recommendations* produced by a recommender system on consumption via good-level randomisation. Existing work compares consumption between personalised and non-personalised recommendations. This experimental design is natural for assessing the impact of the recommendation algorithm on behaviour, but *not* for isolating the role of the recommendation itself or discerning the mechanisms through which it operates.[6] Our experimental design instead focuses on isolating the role of the recommendation itself by having a control and a forced consideration groups of goods *that would have been recommended*, which allows us to identify the causal effect of recommendation and disentangle its mechanisms.

Two closely related papers 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 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 using a field experiment that changes the quality of recommendations. In contrast, our work directly measures the informational gains from recommendations and separates these gains from their effects on consideration.

In addition, there is a growing literature in marketing focused on quantifying the effect of personalised product rankings on consumption (Ursu, 2018; Greminger, 2022; Korganbekova and Zuber, 2023; Donnelly et al., 2024; Compiani et al., 2024). Our paper differs by taking an experimental approach, decomposing mechanisms, and focusing on product recommendations which are more common in markets with experience goods, whereas rankings are more prevalent in markets with search goods.

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 have been proposed to assess recommendation quality by combining predicted consumer ratings with consumers' past consumption history. These metrics can be categorised into two groups: those that define good recommendations as ones that expand consumers' consideration

[6] A well-implemented recommender system should result in a treatment group with systematically higher idiosyncratic quality relative to the control group, which biases measurement of the effect of recommendation itself. The econometric challenge is similar to that of measuring the effect of targeted advertising, since the econometrician needs to account for selection into targeting (Gordon et al., 2019).

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.

2. Experimental Hypotheses

This section describes our primary experimental hypotheses. The goal of our experimental intervention is to unpack the economic mechanisms through which recommendations affect choices. Throughout, we focus on experience goods since this is a common setting in which RecSys are deployed and a good match for our experimental environment.

We suppose that consumers have imperfect knowledge about the set of possible goods that exist in the market and that, even for the goods they actively consider, they hold beliefs about how much they will like the good conditional on consumption. We hypothesise that recommendations can therefore impact choices in two possible ways: they can make consumers consider goods (*consideration role*) and, conditional on consideration, provide (directly and indirectly) information about how much they will like the good upon consumption (*informational role*).

Following Honka et al. (2017), we hypothesise that the *consideration* channel includes both an awareness and a consideration component. Awareness means that the consumer is aware of the existence of a good but has never considered it (e.g., they've heard of *A Serious Man* but have never seriously considered watching it). Consideration means that, when deciding which goods to consume, the consumer actively evaluates whether to consume it. In practice, recommendations feature prominently on the homepage of MovieLens (and other platforms) which, by design, makes consumers aware of the goods. Whether or not this awareness translates to consideration depends on whether the consumer attends to the recommendation. In this paper, we consider the joint role of awareness and consideration, without decomposing the two, and design our intervention to force consideration. This motivates our first hypothesis:

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

Implicit in Hypothesis 1 is that recommendations act beyond just exposing consumers to goods. We consider the explanation that recommendations provide information that leads to changes in consumers' attitudes about the goods. Underlying this is an assumption that consumers make choices based on their underlying beliefs about how much they will like a particular good. Therefore, to test that recommendations influence choices through the informational channel, we first want to test the intermediate hypothesis that these beliefs matter for choices. We operationalise consumer beliefs by measuring consumers' expected quality for a specific good and the uncertainty about that assessment.[7,8] This provides us with the following hypothesis:

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

We expect Hypothesis 2 to be true based on any reasonable model of consumer behaviour in this context. However, in our experimental intervention, we will actively elicit these beliefs from consumers on MovieLens and so by testing Hypothesis 2 we are also implicitly testing that we are able to elicit meaningful beliefs from consumers. If Hypothesis 2 is indeed true, then one channel through which recommendations could impact choices is by shifting beliefs through information provision. This provides us with our final hypothesis:

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

Hypothesis 3(i) is a natural implication of standard models of information acquisition as information should primarily reduce uncertainty. We posit Hypothesis 3(ii) since, in many contexts where RecSys are deployed (including the platform we deploy the intervention on), the platform provides both the recommended good and the consumer-specific predicted match value. Thus, if consumers internalise this information, then their beliefs should shift toward the platform's predictions. We remain agnostic about the source of the information, which can operate through two channels: the recommendation itself may directly signal quality, and it may reduce information acquisition costs due to the prominent positioning on the platform.

3. Experimental Design

In order to study whether 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 match values prior to consumption to

[7]The hypothesis that uncertainty matters implicitly tests whether consumers are uncertainty-averse: all else being equal, consumers preferring goods with lower uncertainty to those with higher uncertainty.

[8]Since we are in a setting with experience goods, we assume that the true match value for the good is only revealed after consumption and that, upon consumption, they have no uncertainty about it.

examine the informational mechanisms through which recommendations act. 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.[9] Consumers use the platform to find and obtain information about movies, receive personalised recommendations based on their ratings, and rate movies after watching. It does not provide consumption opportunities, as it does not host movies to stream nor does it direct users to other platforms.[10]

The platform's home page displays movies organised by rows, with the first row showing eight "top picks", the platform's top recommended movies for the user. Movies are set in a grid, with their poster, title, and the platform-predicted rating for the user. The platform-predicted ratings are personalised to each user and obtained through a sophisticated collaborative filtering algorithm that combines the user's and others' past ratings, reviews, and other metadata (Harper and Konstan, 2015). The properties, effectiveness, and potential biases of this algorithm's outputs – in particular in the MovieLens context – have been widely studied by previous literature (e.g., see Herlocker et al. (1999); Sarwar et al. (2001); Ekstrand et al. (2011) for several examples). Our goal is to assess how consumers interpret the output of this algorithm, which has since been deployed across many domains outside of movies, and the extent to which it impacts their beliefs and choices.

3.2. Experimental Intervention and Measurements

In order to test the hypotheses laid out in Section 2, we conduct a six-month experimental intervention on MovieLens. Since recommendations are personalised, we implement a *within-subjects* design to control for consumer's idiosyncratic tastes, relying on the RecSys prediction of the rating for each consumer-good pair. For each consumer, we select the 750 previously unrated goods with the highest *personalised* predicted ratings and apply consumer-specific stratified block randomization. Within each consumer, goods are grouped into triplets based on that consumer's predicted ratings and randomly assigned within each triplet to one of three experimental conditions: *control*, *forced consideration*, and *recommendation*. This design ensures that, for a given consumer, comparisons across conditions are made among goods with similar baseline predicted quality for each consumer, rather than being driven by idiosyncratic quality differences.

[9]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 (Harper and Konstan, 2015). In Online Appendix C we provide additional discussion of the platform as well as screenshots of the interface.

[10]We define our proxies for consumption in Section 4.

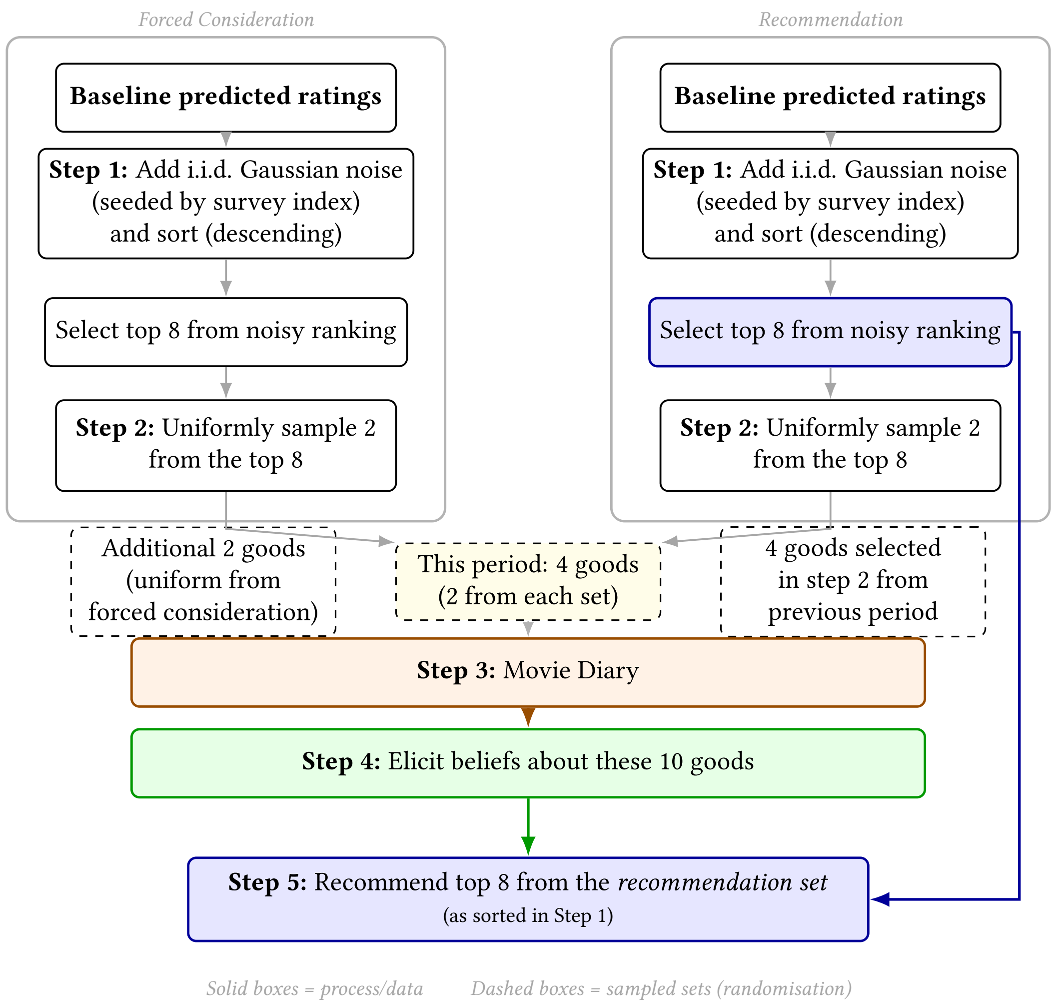


Figure 1: Belief Elicitation and Recommendation Procedure

Throughout the six-month period, we enforce that, for a given consumer, only goods from the *recommendation* set enter into the 'top picks' recommendations on the platform's homepage, whereas goods from the *control* set are held out from the recommendations entirely during this period. However, in order to test Hypothesis 1, we need a way to disentangle the effect of forced consideration from the informational role of recommendations.

We overcome this by introducing a survey that appears when consumers enter into the platform during the experimental period. The survey begins with a movie diary – where consumers self-report what movies they have watched since their last platform visit (see Figure 7). Then, for 10 goods from the *recommendation* and *forced consideration* sets, we ask whether they have watched the movie (see Figure 8). If they had watched it, we ask for their rating and approximate watch date. If they had not watched it, we elicit their expected rating using the platform's typical rating scale, a 10-point Likert scale ranging from 0.5 (worst) to 5 (best), and their level of certainty on a 5-point Likert Scale, ranging from 1 (most certain) to 5 (least certain).[11] We assume that, upon marking a movie as watched and rating it, consumers learn the match value and there is no remaining uncertainty. We then take the reported ratings as part of our belief data. Furthermore, throughout we will assume that belief data were truthfully reported and validate the internal and external validity of the belief data in Appendix A.[12]

We only include goods from the *forced consideration* and *recommendation* groups in the survey. Since responding entails actively thinking and expressing an opinion about the good, we assume that goods in *forced consideration* are actively added to the consumer's consideration set, enabling us to measure the effect of consideration on choices by comparing them with the control group. Importantly, the survey only provides the movie title and poster, stripping away the informational content of the recommendations[13] and ensuring this elicitation only serves to inform about the existence of the good. Thus, by comparing the *recommendation* and *forced consideration* goods, we are able to measure the further impact of recommendation beyond consideration as it additionally contains the information typically provided in recommendations.

The next component of the experimental design is to determine which goods appear in the recommendations and belief elicitations, since they comprise only a small fraction of the goods in the consumer-specific experimental groups. We want to induce exogenous variation in the recommendations, elicit beliefs about goods before and after they are recommended, and ensure that the predicted idiosyncratic quality for elicited goods is, on average, similar between the *recommendation* and *forced consideration* groups.[14] We implement this using the procedure for selection at a given time period documented in Figure 1.

Once a consumer completes the belief survey, it does not appear again until at least 24 hours and, importantly, a subset of goods from the current survey persists to the following survey. This persistence is important as it ensures that, for goods in the *recommendation* group, we elicit beliefs before and after the recommendation, whereas for goods in the *forced consideration* group, beliefs are elicited twice without the informational content from recommendations. We additionally note that (i) we keep predicted ratings fixed at their pre-experiment values and (ii) consumers can skip the survey, but the random seed remains fixed until they complete the current-period

[11] Throughout the rest of the paper, in line with our discussion in Section 2 we interpret the rating as the match value. Thus, we refer to the platform's predicted rating as the predicted match value, the consumer's pre-consumption expected rating as expected match value, the consumer's post-consumption rating as realised match value.

[12] While our survey is not incentivised, the platform is non-commercial and routinely uses consumer inputs, including ratings, to help improve a consumer's recommendations.

[13] This includes the platform-predicted ratings, tags, genre, and the ability to visit the details page.

[14] Additionally, we want to ensure that we retain the high idiosyncratic quality of the product recommendations.

survey, ensuring that recommendations and belief elicitations remain the same until completion.

3.3. Recruitment and Study Implementation

In our intervention, we target a random sample of the platform's users.[15] We restrict our sample to users who satisfy the following conditions: *(i)* have rated more than 100 movies in total; *(ii)* have rated fewer than 3,000 movies in total; and *(iii)* over the previous $m = 1, 2, 3, 4$ months, have rated a minimum of $\lceil 1.5m \rceil$ movies. The first condition is a minimum data requirement so that the recommender system is able to provide valuable recommendations. This is especially important given that, throughout the duration of the intervention, the assignment of movies to treatments is held fixed and therefore so is the set of movies that can be recommended. The second condition excludes power users. The last condition ensures that targeted users are 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: stable users who are familiar with the platform's recommender system. By targeting a minimum level of engagement, we are measuring the effects of recommendations on a set of consumers who have had repeated interactions with the recommender system and formed their expectations about the value of the information coming from it.[16]

The intervention lasted from 29 March 2021 to 31 October 2021 with a phased rollout – a randomly selected portion of consumers from 29 March 2021 until 15 April 2021, followed by a full rollout to all eligible consumers.[17] The length of the study period was selected based on power calculations and the potentially slow rate of movie consumption over time. The intervention targeted 4,572 eligible users, of which 1,452 decided to enrol in the study; we conduct our analysis on data from these participants. The summary statistics for belief survey completions, good consumption, platform visits, and beliefs are provided in Table 5.

4. Causal Effect of Recommendations on Consumption

In this section, we test Hypothesis 1 by asking whether consideration and recommendation causally induce additional consumption. First, we describe our consumption measures and then discuss our empirical specification for estimating treatment effects.

[15] The use of or access to the platform is prohibited to individuals under the age of 18.

[16] This can be interpreted as evaluating the value of recommendation once it has converged from a dynamic feedback process in which consumers arrive with priors about the usefulness of the recommendations, provide data based on consumption, receive recommendations, and subsequently update their beliefs based on the effectiveness of these recommendations. Better understanding this dynamic process and how platforms can build trust among their users is an important aspect of recommender system design, but is beyond the scope of this paper.

[17] The recruitment dialogue box is displayed in Figure 6.

Consumption Measures. Since the platform does not offer consumption opportunities, we rely on self-reported proxies for consumption.

Our first measure, denoted *movie diary*, comes from the belief-elicitation survey, where consumers are asked to indicate and rate any movies they have watched since they last completed the survey (see Figure 7). Our second measure, denoted *robust ratings*, combines on-platform user ratings with the ratings and viewing times reported in the belief-elicitation survey. Both have advantages and disadvantages. On the one hand, ratings may be selectively reported based on salience on the platform or enjoyment, even if this concern is less severe on MovieLens than on other review platforms (Schoenmueller et al., 2020; Rossi and Schleef, 2024). On the other hand, internal interview studies by MovieLens show that users treat ratings as a form of consumption diary itself, which implies that the movie diary is particularly prone to under-reporting precisely the most salient movies on the platform.[18] Although robust ratings are our preferred measure of consumption, to avoid over-estimating the treatment effects of exposure and recommendation on consumption, we primarily rely on the movie diary and the 804 consumers who used it to estimate the treatment effects on consumption discussed in this section. This yields conservative estimates, with attenuation in magnitudes and a bias against finding effects of forced consideration and recommendation.

Empirical Specification. We consider the following specification:

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

where $c_{i,x} = 1$ if, during the intervention, consumer i reported consuming good x (and 0 otherwise), $e_{i,x} = 1$ if good x is in consumer i's forced consideration or recommendation sets and 0 if it is in the control set, and $r_{i,x} = 1$ if x is in i's recommendation set.[19]

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

Estimated over the full sample, this specification provides Intent-to-Treat (ITT) estimates, since our randomisation assigns 250 goods to each group, but not every good is actually forced-considered

[18] Indeed, several movies were first marked as unseen in a belief-elicitation survey and then as seen in a later one, without ever featuring in the movie diary.

[19] Throughout the paper, our primary empirical specifications do not include consumer fixed effects. However, in Online Appendix A.5 we show that the results do not quantitatively change if we include them.

or recommended during the experimental period.[20]

	Dependent variable:	
	Consumption (Movie Diary)	
	ITT	ATT
	(1)	(2)
Consideration	−0.0001	
	(0.0001)	
Recommendation	0.0003***	0.0029***
	(0.0001)	(0.0008)
Constant	0.0008***	0.0009***
	(0.0001)	(0.0003)
Consumer FE	No	No
Observations	603,000	19,013
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 1: Causal Effect of Recommendation on Consumption (Movie Diary)

NOTES: This table tests Hypothesis 1 (the causal effect of recommendation and consideration on consumption) by estimating Equation 1. Each observation corresponds to a pair (consumer i, good x), and consumption is defined according to the movie diary measure. Column (1) reports Intent-to-Treat (ITT) estimates and Column (2) reports Average Treatment Effect on the Treated (ATT) estimates for the effect of recommendation on consumption. We estimate this over consumers with at least one entry in their movie diary during the experiment. We remove any goods that were marked as consumed before the consumer enrolled in the experiment. Standard errors are clustered at the consumer level.

This motivates estimating Equation 1 while restricting the sample to goods that consumers explicitly reported in the belief-elicitation survey that they had not consumed before. This comparison can only be done between the recommendation and forced consideration treatment groups since, by construction, we do not observe beliefs for the control group. However, restricting to elicited goods ensures that consumers (i) did not consume these goods before the intervention, (ii) actively thought about them (thus are both aware of and consider them), and (iii) were exposed to them via the intervention (forced consideration for both treatment groups via the elicitation and additionally recommendation for the recommendation group). As such, it provides an estimate of the Average Treatment Effect on the Treated (ATT).

[20]This results in an underestimate of the effect of recommendation, relative to forced consideration, since goods that appear in recommendations but not in the belief-elicitation survey may not be actively considered.

Results. We estimate these specifications and display the results using the movie diary measure in Table 1. The effect of consideration alone is negligible, refuting the first part of Hypothesis 1. Recommendation leads to a statistically and economically significant increase in the probability of consumption – 30% (ITT) and 341% (ATT) increase, providing support for the second part of Hypothesis 1. The ATT estimates indicate that recommendation shifts choice toward the recommended good – raising its consumption probability by 341% relative to consideration alone. Because goods across treatment groups within a consumer have similar ex ante choice probabilities by construction, we interpret this effect primarily as reallocation across available goods induced by recommendation. In Online Appendix A.1 we replicate this using the robust ratings measure, which yields qualitatively similar results, except that we find a statistically significant (but economically small) 7% increase (ITT) due to forced consideration.

5. Recommendations, Beliefs, and Consumption

The results from Section 4 support Hypothesis 1(ii), indicating that recommendation has an economically large effect beyond consideration alone. In this section, we examine whether this 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 Cause Consumption

We start by evaluating whether beliefs cause consumption by testing Hypothesis 2, which — in line with our theoretical framework — predicts that the likelihood of consumption is increasing in expected match value and decreasing in reported uncertainty. We evaluate this relationship through the following regression:

$$c_{i,x} = \beta_1 v^b_{i,x} + \beta_2 \sigma^b_{i,x} + \epsilon_{i,x} \tag{2}$$

where $v^b_{i,x}$ and $\sigma^b_{i,x}$ denote the last elicitation of consumer i's expected match value and the uncertainty associated with good x.[21] We rely on the robust ratings consumption measure for this analysis since it already conditions on goods consumers have reported not having watched before the intervention and under-reporting is less of a concern with this measure.[22]

In order to enable a causal interpretation of the relationship between beliefs and consumption, we use an instrumental-variables approach. We instrument expected match value and uncertainty

[21] In Table 8, we show that our conclusions on the causal impact of beliefs on consumption remain robust to considering the first elicitation, instead of the last.

[22] In Online Appendix A.3 we nonetheless show that the same qualitative patterns arise if we replicate this analysis using the movie diary consumption measure.

	Dependent variable:		
	Consumption (Robust Rating)		
	(1)	(2)	(3)
Uncertainty	-0.113^{**}		-0.101^{**}
	(0.046)		(0.047)
Exp. Match Value		0.068^{***}	0.084^{***}
		(0.020)	(0.027)
Constant	0.400^{**}	-0.188^{***}	0.107
	(0.155)	(0.060)	(0.155)
Weak Instruments (Uncertainty)	11.22	-	4.40
(p-value; df.)	(0.001; (1, 21233))	-	(0.004; (3, 21231))
Weak Instruments (Exp. Match Value)	-	54.86	19.71
(p-value; df.)	-	(0.000; (1, 21233))	(0.000; (3, 21231))
Wu-Hausman	10.05	9.79	9.55
(p-value; df.)	(0.002; (1, 21232))	(0.002; (1, 21232))	(0.000; (2, 21230))
Sargan	-	-	0.16
(p-value; df.)	-	-	(0.689; 1)
Observations	21,235	21,235	21,235
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2: Beliefs Cause Consumption (Hypothesis 2)

NOTES: This table tests Hypothesis 2 — the causal effect of a good's expected match value and uncertainty on whether it is consumed (according to the robust rating measure) — by estimating Equation 2. The covariates are uncertainty and expected match value, using the last elicitation for a (consumer i, good x). We use the instruments described in Section 5.1 — whether the elicitation occurred before or after a recommendation and the opt-in date of the consumer to the intervention. Weak Instruments, Wu-Hausman, and Sargan correspond to tests regarding weak instruments, endogeneity, and validity of over-identifying restrictions. OLS estimates are provided in Table 10. Standard errors are clustered at the consumer level.

with two variables: (1) whether a belief elicitation occurred before or after a recommendation, exploiting the randomisation of which movies are recommended; and (2) the consumer's activity level, proxied by their study opt-in date, leveraging the phased rollout. We include the interaction of both instruments when analysing both endogenous regressors.

The intuition behind these instrumental variables is as follows. For (1), we interpret our experiment as a typical information provision experiment in economics (Haaland et al., 2023) that provides randomised variation in beliefs, so long as recommendation provides information that shifts beliefs (which we explicitly test in Hypothesis 3). For (2), we exploit two sources of exogenous variation: consumers' activity level on the platform, which proxies for how much information they have about movies, and when they begin receiving the information intervention, which varies exogenously across consumers due to the phased rollout. Consumers who opt in earlier are differentially affected over time by the information intervention. Because platform activity and opt-in timing influence consumption only through their impact on updated expectations, and have no other direct effect on how much they consume, they together satisfy relevance (shift beliefs) and exogeneity (affect consumption via beliefs).

Our results are displayed in Table 2 and are consistent across all specifications.[23] Columns (1) and (2) present the results of using the information provision and enrolment instrumental variables separately for uncertainty and expected match value, respectively. Column (3) presents the results of estimating Equation 2 using both instrumental variables. The estimates suggest that a one-point increase in expected match value and a one-point decrease in uncertainty lead to, respectively, an 8.4 and a 10.1 percentage point increase in consumption probability, lending support for Hypothesis 2. These estimates are economically meaningful, suggesting that the effect sizes observed in Section 4 can be rationalised if recommendation changes consumer beliefs.

5.2. Causal Effect of Recommendations on Beliefs

Now that we have established that beliefs causally determine consumption, we explore whether recommendation affects these beliefs. The information conveyed by recommendations about potentially high-match-value goods can lead to belief updating and subsequently influence choices. This informational channel constitutes the crux of Hypothesis 3.

We examine two aspects of this. First, we test whether recommendations provide information, that is, whether they decrease uncertainty. Second, we test whether recommendations drive consumers' expected match value assessments closer to the platform's predicted match value, that is, whether recommendations reduce the consumer–platform expected match value gap. We estimate the following specification:

$$y_{i,x} = \beta_0 + \beta_1 r_{i,x} + \epsilon_{i,x} \tag{3}$$

where $r_{i,x}$ is an indicator that equals 1 if good x was recommended to consumer i and is otherwise 0, and $y_{i,x} = \Delta\sigma^b_{i,x}$ or $\Delta|v^p_{i,x} - v^b_{i,x}|$, 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, $v^b_{i,x}$, and the platform's predicted match value of good x for consumer i, $v^p_{i,x}$. The change is taken over the course of

[23] 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. 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. Finally, the Sargan test in Column 3 yields a non-significant result (p = 0.672), supporting the validity of the exclusion restriction.

	Dependent variable:	
	Δ Uncertainty	Δ User-Platform Exp. Match Value Gap
	(1)	(2)
Recommendation	-0.063^{***}	-0.015^{**}
	(0.013)	(0.007)
Constant	-0.055^{***}	0.004
	(0.008)	(0.004)
Consumer FE	No	No
Observations	21,283	21,283
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3: The Causal Effect of Recommendation on Beliefs (Hypothesis 3)

NOTES: Columns (1) and (2) test, respectively, Hypothesis 3(i) — the causal effect of recommendation on uncertainty — and Hypothesis 3(ii) — the causal effect of recommendation on the shift in expected match value toward that provided by the platform's predictions. The dependent variables are uncertainty $y_{i,x} = \Delta\sigma^b_{i,x}$ for Column (1) and the change in the absolute difference between the consumer's expected value and the platform's predicted value, $y_{i,x} = \Delta|v^b_{i,x} - v^p_{i,x}|$ for Column (2). In order to measure the dependent variable, we compute the difference between the first and last beliefs reported for each consumer i and good x. We estimate Equation 3 for each dependent variable separately. Standard errors are clustered at the consumer level.

the experiment, comparing the first and last beliefs reported for each consumer i and good x, as specified in Section 3.

Since goods selected for belief elicitation are of the same expected (high) match value, regardless of whether they are recommended, we are able to identify a causal effect of recommendations on consumer beliefs. Columns (1) and (2) of Table 3 present our estimates for Equation 3 with $y_{i,x} = \Delta\sigma^b_{i,x}$ and $y_{i,x} = \Delta|v^p_{i,x} - v^b_{i,x}|$, respectively. We find support for Hypothesis 3, with recommendations decreasing uncertainty for the recommended good and reducing the match value difference between consumers' expectations and the platform's predictions.[24] Recommending a good significantly decreases uncertainty by 0.063 and closes the match value gap by 0.015. Together with the effect sizes estimated in Section 5.1, these effect sizes are economically large.

These results are consistent with the increase in consumption from recommendations coming largely from an informational mechanism. In Online Appendix B we show that both informational channels play a role: consumers acquire more information about recommended goods, which drives belief updating and, once we control for information acquisition, recommendation

[24] A possible concern is that a five-star rating has different meaning across consumers. In Online Appendix A.4 we show that our results are robust to normalising beliefs within consumer to address this concern.

still has a positive informational role, indicating that it provides a quality signal to consumers.

To further assess whether the informational role is the primary mechanism driving the consumption increase due to recommendations, we assess whether recommendation has a greater impact when consumers are less certain about their valuation of a good and when they are less experienced with the product space. In Appendix B we provide empirical evidence for this: the reduction in uncertainty is stronger for goods with higher prior uncertainty and, using the number of baseline ratings as a proxy for experience, we show that the effect of recommendations on consumption is larger for less experienced consumers.

6. Discussion

In this section, we discuss the implications of our results and highlight caveats related to what our experimental intervention cannot capture.

External Validity. One potential limitation is the generalisability of our findings to other domains. The intervention was conducted on MovieLens, a platform with two notable features: it focuses exclusively on movies and lacks profit-driven motives in its recommender system. By contrast, recommender systems are deployed in domains where consumption processes may differ from those for movies, and many online platforms are profit-oriented.

Nevertheless, we believe our findings offer insights that extend beyond this setting. MovieLens data have been instrumental in advancing research on recommendation algorithms, serving as a benchmark for evaluating new algorithms since the 1990s and influencing the development of algorithms employed by commercial platforms across domains, including news, social media, and online marketplaces (Harper and Konstan, 2015). 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.[25,26] For platforms where recommendations incorporate profit motives, the same design would still allow mechanism decomposition; in the movie

[25] A notable difference between MovieLens and some contemporary platforms is its display of the predicted rating for each consumer. This is still common for typically non-personalised recommender systems (e.g., Tripadvisor, Yelp, etc.), on many movie platforms (e.g., Netflix until recently), and other smaller platforms (e.g., Beli). Nonetheless, we do not believe that our results would be qualitatively different if there were no personalised rating provided to the consumer.

[26] An additional concern is that our sample comprises consumers willing to respond to surveys and participate in an experiment. Despite this, in Aridor et al. (2024a), where belief elicitation is embedded on the homepage, we find similar descriptive results about the relationship between expected and realised match value. Furthermore, as we decompose the joint effect of awareness and consideration from information conditional on consideration, less attention paid to recommendations would not affect the value of information conditional on attention.

domain, to the extent that monetisation reduces perceived recommendation quality, we would expect our estimates to be an upper bound on the informational role.

In short, our work enables 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. Moreover, the approach developed here is portable and can be applied to other contexts to deepen our understanding of how recommendations influence choices under varying objectives and consumption processes.

Implications for Recommender System Design and Evaluation. Our results have implications for recommender system design and evaluation. A key distinction in this literature lies in whether the primary goal is to expand consumers' consideration sets or to provide better information about goods they might already consider. This literature largely explores how to convert predicted ratings/utilities into a slate of recommendations and proposes different objectives for doing so. Some systems prioritise surfacing goods to broaden consumers' options, while others focus on enhancing the informational value of recommendations.

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 more informed choices among options they already consider—are more effective at driving consumption than those focused on consideration set expansion.

This shift in focus has implications for recommender system design. As discussed in the introduction, several metrics are often used as proxies for recommendation informativeness, but they conflate broadening options with providing 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 RecSys design. In a companion paper (Aridor et al., 2024a), we adapt our methodology and introduce a procedure for collecting belief data suitable for large-scale implementation, along with an open-source MovieLens dataset that incorporates this information. These contributions help clarify the economic forces behind consumption choices due to recommender systems and can guide their design and evaluation.

Competing Interests: All authors certify that they have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. This work is supported by a grant from the

National Science Foundation.

7. References

Aridor, G., D. Goncalves, R. Kong, D. Kluver, and J. Konstan. 2024a. "The MovieLens beliefs dataset: Collecting pre-choice data for online recommender systems." In *Proceedings of the 18th ACM Conference on Recommender Systems.*

Aridor, G., R. Jiménez-Durán, R. Levy, and L. Song. 2024b. "The economics of social media." *Journal of Economic Literature* 62 (2): 601–653.

Bennett, J., and S. Lanning. 2007. "The Netflix prize." In *Proceedings of KDD Cup and Workshop*, Volume 2007. 35, New York.

Castells, P., N. J. Hurley, and S. Vargas. 2015. "Novelty and diversity in recommender systems." In *Recommender Systems Handbook*, 881–918, Springer.

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."

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.

Compiani, G., G. Lewis, S. Peng, and P. Wang. 2024. "Online search and optimal product rankings: An empirical framework." *Marketing Science* 43 (3): 615–636.

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.

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.

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.

Gordon, B. R., F. Zettelmeyer, N. Bhargava, and D. Chapsky. 2019. "A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook." *Marketing Science* 38 (2): 193–225.

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.

Greminger, R. P. 2022. "Heterogeneous position effects and the power of rankings."

Haaland, I., C. Roth, and J. Wohlfart. 2023. “Designing information provision experiments.” *Journal of Economic Literature* 61 (1): 3–40.
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.
Herlocker, J. L., J. A. Konstan, A. Borchers, and J. Riedl. 1999. “An algorithmic framework for performing collaborative filtering.” In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 230–237.
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.”
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.
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.
Korganbekova, M., and C. Zuber. 2023. “Balancing user privacy and personalization.”
Kotkov, D., S. Wang, and J. Veijalainen. 2016. “A survey of serendipity in recommender systems.” *Knowledge-Based Systems* 111 180–192.
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.
Resnick, P., and H. R. Varian. 1997. “Recommender systems.” *Communications of the ACM* 40 (3): 56–58.
Rossi, M., and F. Schleef. 2024. “Quality disclosures and disappointment: Evidence from the Academy nominations.”
Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. “Item-based collaborative filtering recommendation algorithms.” In *Proceedings of the 10th International Conference on World Wide Web*, 285–295.
Schoenmueller, V., O. Netzer, and F. Stahl. 2020. “The polarity of online reviews: Prevalence, drivers and implications.” *Journal of Marketing Research* 57 (5): 853–877.
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.
Ursu, R. M. 2018. “The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions.” *Marketing Science* 37 (4): 530–552.
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.
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 show that the belief data we collect exhibit reasonable patterns and are informative about realised match value. In other words, we provide evidence that consumers have well-formed beliefs about movies and that survey-based measures can accurately capture them.

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:

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

where $v_{i,x}$ denotes the realised match value of good x for consumer i, and, to capture consumers' initial beliefs before any intervention, $v^b_{i,x}$ denotes consumer i's first belief elicitation about good x. The results in Column (1) of Table 4 show that the estimated coefficient β_1 is precisely estimated

	Dependent variable:		
	Realised Match value	Realised vs. Exp. Match Value Gap	Prior Uncertainty
	(1)	(2)	(3)
Exp. Match Value	1.017^{***}		
	(0.010)		
Uncertainty		0.201^{***}	
		(0.075)	
Log(User Ratings)			-0.465^{***}
			(0.060)
Log(Movie Ratings)			-0.056^{*}
			(0.033)
Movie is Sequel			-0.061^{***}
			(0.010)
Constant			4.167^{***}
			(0.227)
Consumer FE	No	Yes	No
Observations	408	408	20,788
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Properties of Belief Data

Notes: This table demonstrates sensible patterns in the belief data. Column (1) estimates the correlation between realised match value (rating after consumption) and expected match value (expected rating from the belief survey). Column (2) estimates the relationship between the distance between expected and realised match value and prior consumer uncertainty about expected match value. Column (3) displays the relationship between consumer uncertainty and consumers' log past consumption, as given by the number of movies rated at the outset of the experimental period ("User Ratings"), movies' log popularity, as given by the number of ratings at the outset of the experimental period on MovieLens ("Movie Ratings"), and whether the movie is a sequel. Prior beliefs (uncertainty and expected match value) refer to the first belief elicitation for a given consumer and movie.

and close to 1: consumers' prior beliefs 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 reported uncertainty. Specifically, we estimate:

$$\left|v_{i,x} - v^b_{i,x}\right| = \beta_0 + \beta_1 \sigma^b_{i,x} + \kappa_i + \epsilon_{i,x}.$$

Column (2) of Table 4 reports $\beta_1 > 0$, indicating a positive relationship between uncertainty and the resulting difference, and validating that greater uncertainty is associated with less aligned expected and realised match values.

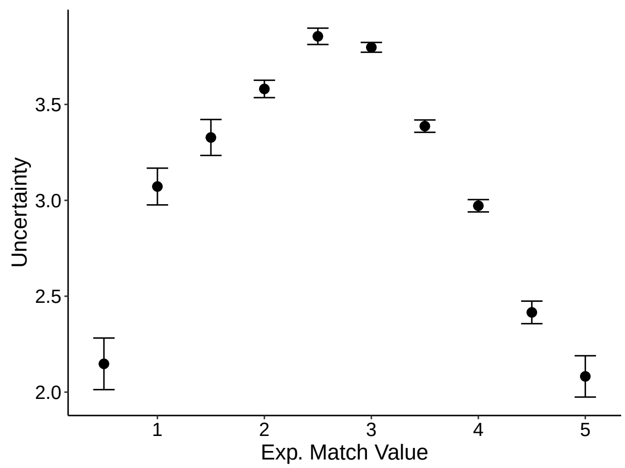


Figure 2: Uncertainty by Expected Quality

NOTES: This figure shows 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. We estimate this on the full set of belief elicitations.

We then explore the relationship between expected match value and uncertainty and show in Figure 2 that, as one would expect, consumers are more certain about extreme value assessments (i.e., close to 0 or 5 stars) than about more moderate ones (i.e., 3 stars).

The final validity check explores how uncertainty relates to the popularity of a good (measured by the number of community ratings on MovieLens), consumers' past experience (measured by their number of pre-experiment ratings), and whether the movie is a sequel (measured by joining to IMDb data). We therefore run the following regression:

$$\sigma^b_{i,x} = \beta_1 \log(\text{Past Consumption}_i) + \beta_2 \log(\text{Popularity}_x) + \beta_3 \text{Sequel}_x + \epsilon_{i,x} \tag{4}$$

where the notation is similar to the previous specifications. To isolate any role of the experimental intervention in modifying beliefs, we restrict attention 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. It also shows that uncertainty is decreasing in past consumption, indicating that less experienced consumers are, on average, more uncertain about expected match value.

	Mean	Std. Dev.	Median	25th Percent.	75th Percent.	Min	Max
Goods Consumed per consumer (Overall, Movie Diary)	2.11	2.12	1.00	1.00	2.00	1.00	14.00
Goods Consumed per consumer (Intervention, Movie Diary)	0.60	1.43	0.00	0.00	1.00	0.00	12.00
Goods Consumed per consumer (Overall, Robust Rating)	32.87	49.86	19.00	8.00	38.00	1.00	576.00
Goods Consumed per consumer (Intervention, Robust Rating)	4.94	10.22	1.00	0.00	6.00	0.00	177.00
Belief Surveys per consumer	3.98	5.88	2.00	1.00	4.00	1.00	72.00
Unique Visit Days per consumer	16.61	19.31	10.00	5.00	20.00	1.00	157.00
Realised Match Values	3.71	0.88	4.00	3.50	4.00	0.50	5.00
Avg. Realised Match Values per consumer	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
Avg. Exp. Match Values per consumer	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
Avg. Uncertainty per consumer	3.38	0.71	3.33	2.90	3.88	1.00	5.00
Platform Predicted Match Values	3.98	0.50	4.02	3.71	4.31	0.00	5.00
Platform Predicted Match Values per consumer	3.98	0.43	4.01	3.71	4.27	1.59	5.00
Platform Predicted Match Values for Elicited	4.36	0.42	4.37	4.09	4.69	2.64	5.00
Platform Predicted Match Values for Elicited per consumer	4.39	0.37	4.42	4.13	4.68	2.81	5.00

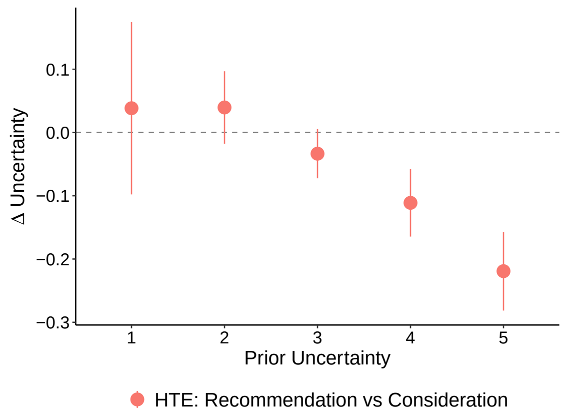
Table 5: Summary Statistics on Consumption and Beliefs

NOTES: This table provides summary statistics during the experimental intervention. The first four rows present summary statistics on good consumption. The first two rows display consumption using the movie diary: the first row presents consumption of any goods, whereas the second row presents consumption of goods targeted for the experimental intervention that occurred during the intervention. For this measure, we restrict to consumers with at least one entry in the movie diary. The next two rows display the same summary statistics, except using the robust ratings measure instead of the movie diary. The fifth row displays the number of completed belief elicitation surveys per consumer for consumers who completed at least one survey. The sixth row displays the number of unique platform visit days per consumer. The seventh and eighth rows display realised match value (rating) across all of the data (row 7) and across consumers (row 8). The ninth and tenth rows display expected match value across all of the data (row 9) and across consumers (row 10). The eleventh and twelfth rows display the reported uncertainty regarding the match values for all goods in the experimental intervention across all of the data (row 11) and across consumers (row 12). The thirteenth and fourteenth rows display the platform's predicted match values for all goods in the experimental intervention across all of the data (row 13) and across consumers (row 14). The fifteenth and sixteenth rows display the platform's predicted match values for the goods that appeared in the belief elicitation survey across all of the data (row 15) and across consumers (row 16).

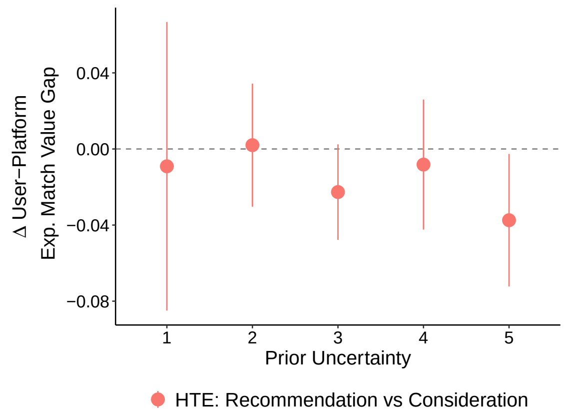
We conclude the validation section by presenting summary statistics about the belief data and the experiment in Table 5. Table 5 highlights differences between realised and expected match values, uncertainty measures, and the platform's predicted match values for the full set of goods in the experimental intervention versus the goods that appear in elicitations and recommendations.

Appendix B. Heterogeneous Effects

In this appendix, we explore heterogeneity in the effects of recommendations. We consider two dimensions of heterogeneity: differences in belief updating according to baseline beliefs and differences in consumption according to baseline consumption levels.



(a) Uncertainty



(b) Expected Match Value

Figure 3: Effect of Recommendations on Beliefs: Heterogeneous Effect by Uncertainty

NOTES: The figure estimates the average treatment effect of recommendations on beliefs relative to forced consideration, conditional on prior uncertainty, by estimating Equation 3. Panel (a) exhibits the estimated treatment effect of recommendations (relative to forced consideration) on uncertainty, whereas Panel (b) displays the treatment effect on the distance between the consumer's expected match value and the platform's predicted match value, both conditional on reported prior uncertainty. Lines represent 95% confidence intervals with clustered standard errors at the consumer level. The change is taken over the course of the experiment, comparing the first and last beliefs reported for each consumer i and good x.

Prior Beliefs. To better understand heterogeneity by prior beliefs, we assess whether consumers differentially update their beliefs based on baseline uncertainty levels. To do so, we re-estimate Equation 3:

$$y_{i,x,t} = \beta_0 + \beta_1 r_{i,x} + \epsilon_{i,x}$$

again, where we consider $y_{i,x,t} = \Delta\sigma^b_{i,x}$ and $y_{i,x,t} = \Delta|v^p_{i,x} - v^b_{i,x}|$. Different from before, we estimate this specification *conditional on prior beliefs*, that is, separately for each level of prior uncertainty.

The results are presented in Figure 3: Panel (a) shows the change in uncertainty ($\Delta\sigma^b_{i,x}$) and Panel (b) shows the change in the expected match value gap ($\Delta|v^p_{i,x} - v^b_{i,x}|$). Figure 3(a) shows that 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 when initial uncertainty is higher. In line with this, Figure 3(b) indicates that the expected match value gap is brought closer to the platform's predicted value when prior uncertainty is higher, although the estimates are less precise.

Consumer Experience. A related dimension of heterogeneity is consumers' prior experience. Our working hypothesis is that consumers who have explored a significant portion of the product space hold more precise beliefs about match values. Indeed, we find a negative association between prior uncertainty and consumer experience (see Column (3) of Table 4). Consequently, the informational gain from recommendations may be lower for more experienced consumers. We proxy consumer experience using log past consumption, specifically the number of movies rated prior to the experimental intervention.

	Dependent variable:	
	Consumption (Robust Rating)	
	(1)	(2)
Consideration	0.0004**	
	(0.0002)	
Recommendation	0.011***	0.048**
	(0.003)	(0.019)
Recommendation × Log(User Ratings)	−0.001**	−0.005*
	(0.0005)	(0.003)
Log(User Ratings)	−0.0002	−0.001
	(0.0003)	(0.001)
Constant	0.006**	0.018**
	(0.002)	(0.007)
Consumer FE	No	No
Observations	1,029,601	20,742
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6: Heterogeneous Effects of Recommendation by Consumption Experience

NOTES: This table tests whether recommendations affect consumption differently depending on consumers' past experience with movies, as proxied by the log of the number of movies rated at the outset of the experimental intervention. We estimate Equation 1 expanded to include log past consumption and its interaction with whether a good was in the recommendation treatment for a particular consumer. Each column displays the baseline control and the estimated treatment effect of consideration and recommendation on consumption for the ITT and ATT empirical specifications, as discussed in Section 4. Standard errors are clustered at the consumer level.

To assess whether past consumption experience mediates how recommendations affect consumption, we interact the recommendation term in specification Equation 1 with our measure of consumer experience. The results, presented in Table 6, indicate a negative heterogeneous effect:

recommendations have less effect on consumption for more experienced consumers, with recommendations having approximately 30% less impact for consumers who had watched 500 movies at the outset of the experiment compared to those who had watched only 100.

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.

Online Appendix.

Online Appendix A. Omitted Tables and Robustness Exercises

A.1. Causal Effect of Recommendation on Consumption (Robustness)

	Dependent variable:	
	Consumption (Robust Rating)	
	ITT	ATT
	(1)	(2)
Consideration	0.0003^{**}	
	(0.0002)	
Recommendation	0.005^{***}	0.015^{***}
	(0.0005)	(0.003)
Constant	0.005^{***}	0.012^{***}
	(0.0003)	(0.001)
Consumer FE	No	No
Observations	1,086,523	21,235
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7: The Causal Effect of Recommendation on Consumption (Robust Ratings)

NOTES: This table tests Hypothesis 1 (the causal effect of recommendation and consideration on consumption) by estimating Equation 1. Each observation corresponds to a pair (consumer i, good x), and consumption is defined according to the robust rating measure. Column (1) reports Intent-to-Treat (ITT) estimates and Column (2) reports Average Treatment Effect on the Treated (ATT) estimates for the effect of recommendation on consumption. We remove any goods that were marked as consumed before the consumer enrolled in the experiment. Standard errors are clustered at the consumer level.

A.2. Beliefs Cause Consumption (Robustness)

	Dependent variable:		
	Consumption (Robust Rating)		
	(1)	(2)	(3)
Uncertainty	-0.138^{**}		-0.137^{**}
	(0.063)		(0.069)
Exp. Match Value		0.070^{***}	0.094^{***}
		(0.021)	(0.035)
Constant	0.487^{**}	-0.193^{***}	0.200
	(0.213)	(0.062)	(0.213)
Weak Instruments (Uncertainty)	7.57	-	3.13
(p-value; df.)	(0.006; (1, 21233))	-	(0.025; (3, 21231))
Weak Instruments (Exp. Match Value)	-	53.18	18.63
(p-value; df.)	-	(0.000; (1, 21233))	(0.000; (3, 21231))
Wu-Hausman	10.35	10.21	10.14
(p-value; df.)	(0.001; (1, 21232))	(0.001; (1, 21232))	(0.000; (2, 21230))
Sargan	-	-	0.00
(p-value; df.)	-	-	(0.951; 1)
Observations	21,235	21,235	21,235
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 8: Beliefs Cause 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 2. It differs from Table 2 by considering the first instead of the last belief elicitation for a given consumer and good. 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 over-identifying restrictions. OLS estimates are provided in Table 9. Standard errors are clustered at the consumer level.

	Dependent variable:		
	Consumption (Robust Rating)		
	(1)	(2)	(3)
Uncertainty	-0.007^{***}		-0.006^{***}
	(0.001)		(0.001)
Exp. Match Value		0.010^{***}	0.008^{***}
		(0.001)	(0.001)
Constant	0.042^{***}	-0.012^{***}	0.013^{***}
	(0.004)	(0.004)	(0.004)
Consumer FE	No	No	No
Observations	21,235	21,235	21,235
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 9: Correlation Between Beliefs and Consumption (Robust Rating): 1st Elicitation

NOTES: This table estimates the correlation between a good's expected match value and uncertainty and its consumption probability, as per Equation 2. It differs from Table 10 by considering the first instead of the last belief elicitation for a given consumer and good. Unlike Table 8, the estimates do not have a causal interpretation. Standard errors are clustered at the consumer level.

	Dependent variable:		
	Consumption (Robust Rating)		
	(1)	(2)	(3)
Uncertainty	-0.008^{***}		-0.007^{***}
	(0.001)		(0.001)
Exp. Match Value		0.011^{***}	0.010^{***}
		(0.001)	(0.001)
Constant	0.043^{***}	-0.015^{***}	0.011^{**}
	(0.004)	(0.003)	(0.004)
Consumer FE	No	No	No
Observations	21,235	21,235	21,235
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 10: Correlation Between Beliefs and Consumption (Robust Rating)

NOTES: This table estimates the correlation between a good's expected match value and uncertainty and its consumption probability (using the robust rating measure), as per Equation 2. It considers only the last belief elicitation for a given consumer and good. Unlike Table 2, the estimates do not have a causal interpretation. Standard errors are clustered at the consumer level.

A.3. Beliefs Cause Consumption (Movie Diary Robustness)

	Dependent variable:		
	Consumption (Movie Diary)		
	(1)	(2)	(3)
Uncertainty	-0.026^{**}		-0.023^{*}
	(0.013)		(0.013)
Exp. Match Value		0.013^{**}	0.017^{**}
		(0.006)	(0.007)
Constant	0.089^{**}	-0.039^{**}	0.030
	(0.044)	(0.018)	(0.042)
Weak Instruments (Uncertainty)	10.91	-	4.25
(p-value; df.)	(0.001; (1, 21281))	-	(0.005; (3, 21279))
Weak Instruments (Exp. Match Value)	-	53.81	19.39
(p-value; df.)	-	(0.000; (1, 21281))	(0.000; (3, 21279))
Wu-Hausman	6.03	4.43	4.78
(p-value; df.)	(0.014; (1, 21280))	(0.035; (1, 21280))	(0.008; (2, 21278))
Sargan	-	-	0.20
(p-value; df.)	-	-	(0.658; 1)
Observations	21,283	21,283	21,283
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11: Beliefs Cause Consumption (Movie Diary)

NOTES: This table tests Hypothesis 2 – the causal effect of a good's expected match value and uncertainty on whether it is consumed (using the movie diary measure) – by estimating Equation 2. The covariates are uncertainty and expected match value, using the last belief elicitation for a (consumer i, good x). We use the instruments described in Section 5.1 – whether the elicitation occurred before or after a recommendation and the opt-in date of the consumer to the intervention. Weak Instruments, Wu-Hausman, and Sargan correspond to tests regarding weak instruments, endogeneity, and validity of over-identifying restrictions. It differs from Table 2 by using the movie diary measure of consumption instead of robust ratings. Standard errors are clustered at the consumer level.

	Dependent variable:		
	Consumption (Movie Diary)		
	(1)	(2)	(3)
Uncertainty	-0.031^{*}		-0.031^{*}
	(0.017)		(0.018)
Exp. Match Value		0.014^{**}	0.019^{**}
		(0.006)	(0.009)
Constant	0.109^{*}	-0.040^{**}	0.051
	(0.058)	(0.018)	(0.056)
Weak Instruments (Uncertainty)	7.34	-	3.02
(p-value; df.)	(0.007; (1, 21281))	-	(0.029; (3, 21279))
Weak Instruments (Exp. Match Value)	-	52.13	18.29
(p-value; df.)	-	(0.000; (1, 21281))	(0.000; (3, 21279))
Wu-Hausman	6.05	4.47	4.99
(p-value; df.)	(0.014; (1, 21280))	(0.034; (1, 21280))	(0.007; (2, 21278))
Sargan	-	-	0.03
(p-value; df.)	-	-	(0.865; 1)
Observations	21,283	21,283	21,283
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 12: Beliefs Cause Consumption (Movie Diary): 1st Elicitation

NOTES: This table tests Hypothesis 2 – the causal effect of a good's expected match value and uncertainty on whether it is consumed (using the movie diary measure) – by estimating Equation 2. The covariates are uncertainty and expected match value, using the first belief elicitation for a (consumer i, good x). We use the instruments described in Section 5.1 – whether the elicitation occurred before or after a recommendation and the opt-in date of the consumer to the intervention. Weak Instruments, Wu-Hausman, and Sargan correspond to tests regarding weak instruments, endogeneity, and validity of over-identifying restrictions. It differs from Table 2 by using the movie diary measure of consumption instead of robust ratings and from Table 11 by considering the first, instead of the last, belief elicitation. Standard errors are clustered at the consumer level.

A.4. Robustness using Standardised Beliefs

A possible concern is that a five-star rating has a different interpretation across consumers. This is a canonical issue for interpreting ratings data that predates recommender systems' extensive use of such data (e.g., see Greenleaf, 1992). We examine whether this affects our conclusions regarding the link between recommendations and beliefs. To this end, we construct consumer-specific, normalised measures of uncertainty and expected match value. For each consumer, we subtract the consumer-specific mean at the first elicitation from the elicited measure (uncertainty or expected match value) and divide by the consumer-specific standard deviation. We then re-estimate Equation 3 using standardised beliefs. We report the results in Table 13 and Table 14 (which includes consumer fixed effects). This yields no change in our conclusions regarding magnitudes or statistical and economic significance.

	Dependent variable:	
	Δ Uncertainty	Δ User-Platform Exp. Match Value Gap
	(1)	(2)
Recommendation	-0.063^{***}	-0.014^{**}
	(0.013)	(0.007)
Constant	-0.056^{***}	0.003
	(0.008)	(0.004)
Consumer FE	No	No
Observations	20,704	20,704
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 13: Causal Effect of Recommendation on Beliefs (Hypothesis 3) – Standardised Beliefs

Notes: Columns (1) and (2) test Hypothesis 3(i) – the causal effect of recommendation on uncertainty – and Hypothesis 3(ii) – the causal effect of recommendation on the shift in expected match value toward that provided by the platform's predictions, respectively. The dependent variables are uncertainty $y_{i,x} = \Delta\sigma^b_{i,x}$ for Column 1 and the change in the absolute difference between the consumer's expected value and the platform's predicted value, $y_{i,x} = \Delta|v^b_{i,x} - v^p_{i,x}|$ for Column 2. In order to measure the dependent variable, we compute the difference between the first and last beliefs reported for each consumer i and good x. We estimate Equation 3 for each dependent variable separately. It differs from Table 3 by using standardised beliefs, subtracting from the elicited measure (uncertainty or expected match value) the consumer-specific mean at the first elicitation and dividing by the consumer-specific standard deviation. Standard errors are clustered at the consumer level.

	Dependent variable:	
	Δ Uncertainty	Δ User-Platform Exp. Match Value Gap
	(1)	(2)
Recommendation	-0.058^{***}	-0.025^{**}
	(0.015)	(0.011)
Consumer FE	Yes	Yes
Observations	20,704	20,704
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 14: Causal Effect of Recommendation on Beliefs – Standardised Beliefs; Consumer FE

Notes: Columns (1) and (2) test Hypothesis 3(i) – the causal effect of recommendation on uncertainty – and Hypothesis 3(ii) – the causal effect of recommendation on the shift in expected match value toward that provided by the platform's predictions, respectively. The dependent variables are uncertainty $y_{i,x} = \Delta\sigma^b_{i,x}$ for Column 1 and the change in the absolute difference between the consumer's expected value and the platform's predicted value, $y_{i,x} = \Delta|v^b_{i,x} - v^p_{i,x}|$ for Column 2. In order to measure the dependent variable, we compute the difference between the first and last beliefs reported for each consumer i and good x. We estimate Equation 3 for each dependent variable separately. It differs from Table 3 by using standardised beliefs, subtracting from the elicited measure (uncertainty or expected match value) the consumer-specific mean at the first elicitation and dividing by the consumer-specific standard deviation, and from Table 13 only in that it includes consumer fixed effects. Standard errors are clustered at the consumer level.

A.5. Tables with Consumer Fixed Effects

	Dependent variable:	
	Consumption (Movie Diary)	
	ITT	ATT
	(1)	(2)
Consideration	−0.0003	
	(0.0003)	
Recommendation	0.0010***	0.0054***
	(0.0003)	(0.0017)
Consumer FE	Yes	Yes
Observations	177,750	8,139
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 15: Causal Effect of Recommendation on Consumption (Hypothesis 1): Consumer Fixed Effects, Movie Diary

NOTES: This table tests Hypothesis 1 (the causal effect of recommendation and consideration on consumption) by estimating Equation 1. Each observation corresponds to a pair (consumer i, good x), and consumption is defined according to the movie diary measure. Column (1) reports Intent-to-Treat (ITT) estimates and Column (2) reports Average Treatment Effect on the Treated (ATT) estimates for the effect of recommendation on consumption. It differs from Table 1 only in that it includes consumer fixed effects. Standard errors are clustered at the consumer level.

	Dependent variable:	
	Consumption (Robust Rating)	
	ITT	ATT
	(1)	(2)
Consideration	0.0003**	
	(0.0002)	
Recommendation	0.005***	0.014***
	(0.0005)	(0.003)
Consumer FE	Yes	Yes
Observations	1,086,523	21,235
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 16: Causal Effect of Recommendation on Consumption (Hypothesis 1): Consumer Fixed Effects, Robust Rating

NOTES: This table tests Hypothesis 1 (the causal effect of recommendation and consideration on consumption) by estimating Equation 1. Each observation corresponds to a pair (consumer i, good x), and consumption is defined according to the robust rating measure. Column (1) reports Intent-to-Treat (ITT) estimates and Column (2) reports Average Treatment Effect on the Treated (ATT) estimates for the effect of recommendation on consumption. It differs from Table 7 only in that it includes consumer fixed effects. Standard errors are clustered at the consumer level.

	Dependent variable:	
	Δ Uncertainty	Δ User-Platform Exp. Match Value Gap
	(1)	(2)
Recommendation	-0.055^{***}	-0.015^{**}
	(0.013)	(0.007)
Consumer FE	Yes	Yes
Observations	21,283	21,283
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 17: Causal Effect 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 $\Delta\sigma^b_{i,x}$ and on the change in the absolute difference between the consumer's expected value and the platform's predicted value, $\Delta|v^b_{i,x} - v^p_{i,x}|$, as per Equation 3. The change is taken over the course of the experiment, comparing the first and last beliefs reported for each consumer i and good x. It differs from Table 3 only in that it includes consumer fixed effects. Standard errors are clustered at the consumer level.

	Dependent variable:	
	Consumption (Robust Rating)	
	(1)	(2)
Consideration	0.0004^{**}	
	(0.0002)	
Recommendation	0.011^{***}	0.045^{**}
	(0.003)	(0.018)
Recommendation × Log(User Ratings)	-0.001^{**}	-0.005^{*}
	(0.0005)	(0.003)
Consumer FE	Yes	Yes
Observations	1,029,601	20,742
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 18: Heterogeneous Effects of Recommendation by Consumption Experience: Consumer Fixed Effects

Notes: This table tests whether recommendations affect consumption differently depending on consumers' past experience with movies, as proxied by the log of the number of movies rated at the outset of the experimental intervention. We estimate Equation 1 expanded to include log past consumption and its interaction with whether a good was in the recommendation treatment for a particular consumer. Each column displays the baseline control and the estimated treatment effect of consideration and recommendation on consumption for the ITT and ATT empirical specifications, as discussed in Section 4. It differs from Table 6 only in that it includes consumer fixed effects. Standard errors are clustered at the consumer level.

Online Appendix B. Recommendation and Information Acquisition

In this section, we provide additional analysis to better understand the informational impact of recommendations. We consider two channels through which recommendations provide information: signalling consumer-specific quality (via the fact that a movie is recommended and the platform's predicted rating) and reducing information acquisition costs. We provide additional analysis that further unpacks the relative roles of these two channels.

	Dependent variable:			
	Information Acquisition			
	ITT	ITT	ATT	ATT
	(1)	(2)	(3)	(4)
Consideration	-0.0001	-0.0001		
	(0.0002)	(0.0002)		
Recommendation	0.008^{***}	0.008^{***}	0.066^{***}	0.063^{***}
	(0.001)	(0.001)	(0.007)	(0.007)
Constant	0.005^{***}		0.015^{***}	
	(0.001)		(0.003)	
Consumer FE	No	Yes	No	Yes
Observations	1,088,250	1,088,250	21,283	21,283
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 19: Causal Effect on Information Acquisition

Notes: This table tests whether recommendation causes information acquisition by estimating Equation 1. Each observation corresponds to a pair (consumer i, good x), and information acquisition is defined as whether consumer i visits the details page for good x at any point during the intervention. Columns (1) and (2) report Intent-to-Treat (ITT) estimates and Columns (3) and (4) report Average Treatment Effect on the Treated (ATT) estimates for the effect of recommendation on information acquisition. Columns (2) and (4) include consumer fixed effects, while Columns (1) and (3) do not. Standard errors are clustered at the consumer level.

First, we use click and search data on MovieLens to measure the causal effect of recommendation on the probability that a consumer acquires additional information about a good on the platform.[27] We define information acquisition as visiting the movie details page (see Figure 5), which

[27]Consumers can also provide ratings on the details page. To ensure that we capture information acquisition—rather than rating behaviour—we define a details page visit as information acquisition only if it occurs more than 30 minutes before a rating is provided.

consumers can access either by clicking the movie poster on the homepage or by searching for the good. We estimate the ITT and ATT specifications for the causal effect on consumption (Equation 1) using information acquisition as the dependent variable and present this analysis in Table 19. The results indicate that recommended goods experience a 0.8 percentage point increase in visits to the details page relative to control (ITT) and a 6.6 percentage point increase relative to forced consideration alone (ATT). Furthermore, we estimate a precise null effect of forced consideration on additional information acquisition, indicating that part of the informational effect operates through reduced information acquisition costs, which increases information acquisition primarily in the recommendation group.

	Δ Uncertainty		Δ User-Platform Exp. Match Value Gap	
	(1)	(2)	(3)	(4)
Info Acquisition	-0.451^{***}	-0.437^{***}	-0.027	-0.021
	(0.053)	(0.053)	(0.023)	(0.024)
Recommendation		-0.035^{***}		-0.013^{*}
		(0.013)		(0.007)
Constant	-0.058^{***}	-0.048^{***}	0.0003	0.004
	(0.008)	(0.007)	(0.004)	(0.004)
Consumer FE	No	Yes	No	Yes
Observations	21,283	21,283	21,283	21,283
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 20: Causal Effect of Recommendation on Beliefs (Hypothesis 3) – Controlling for Information Acquisition

NOTES: Columns (1)-(2) and (3)-(4) test Hypothesis 3(i) – the causal effect of recommendation on uncertainty – and Hypothesis 3(ii) – the causal effect of recommendation on the shift in expected match value toward that provided by the platform's predictions, respectively. The dependent variables are uncertainty $y_{i,x} = \Delta\sigma^b_{i,x}$ for Columns (1)-(2) and the change in the absolute difference between the consumer's expected value and the platform's predicted value, $y_{i,x} = \Delta|v^b_{i,x} - v^p_{i,x}|$ for Columns (3)-(4). To construct the dependent variable, we compute the difference between the first and last beliefs reported for each consumer i and good x. We estimate Equation 3 for each dependent variable separately. It differs from Table 3 by controlling for whether consumer i acquired information about good x. Standard errors are clustered at the consumer level.

Are all informational gains driven by information acquisition, or are there gains from direct inference? While we cannot causally decompose the relative importance of the quality signal and the information gained from explicit information acquisition, we replicate our main specification for the causal effect of recommendation on beliefs (Equation 3) while additionally controlling for whether the consumer acquired information about the movie. The results are reported in Table 20. First, information acquisition induces an economically large and statistically significant reduction in uncertainty, but, importantly, there remains an economically and statistically significant residual effect of recommendation. Second, information acquisition has little effect on the distance between the platform's predicted match value and the consumer's expected match value, while the coefficient on recommendation remains marginally statistically significant. These results provide evidence for the presence of both informational channels: consumer-specific quality signals and reduced information acquisition costs due to recommendation.

Online Appendix C. Background on 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.[28]

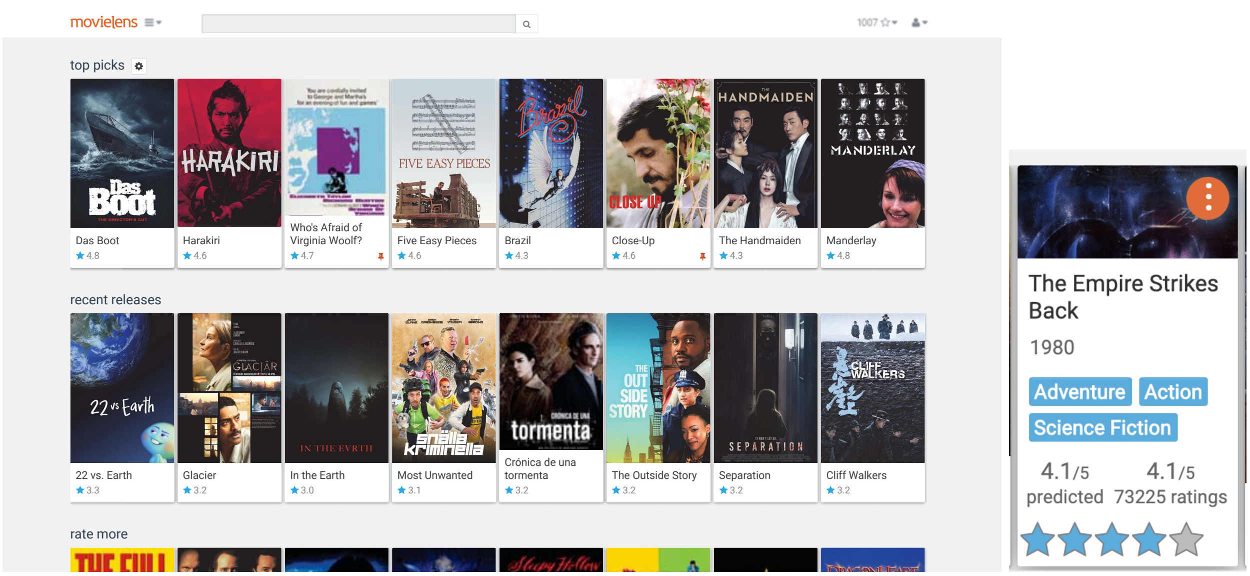
The platform's home page displays movies organised by categories in rows, with the very first one showing eight "top picks", the platform's top recommended movies for the user. Movies are set in a grid fashion, with their poster, title, and the platform-predicted rating for the user. The platform-predicted 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). Screenshots of the platform's interface are included at the end of this section. 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 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,

[28] For 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.

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 and makes it incentive-compatible for users to report their ratings.



(a) Platform Homepage

(b) Movie Card

Figure 4: MovieLens Home Page

NOTES: This presents the default interface on the platform. Panel (a) shows the homepage of the platform which has 8 movies per row. The first row is always the 'top picks' which are the personalized recommendations provided to consumers by the platform. The subsequent rows include thematic rows (e.g., recent releases) that are typically not considered product recommendations. On the homepage, consumers can hover over a given movie which presents the details presented in Panel (b) – genre, release date, and platform's personalized predicted rating.

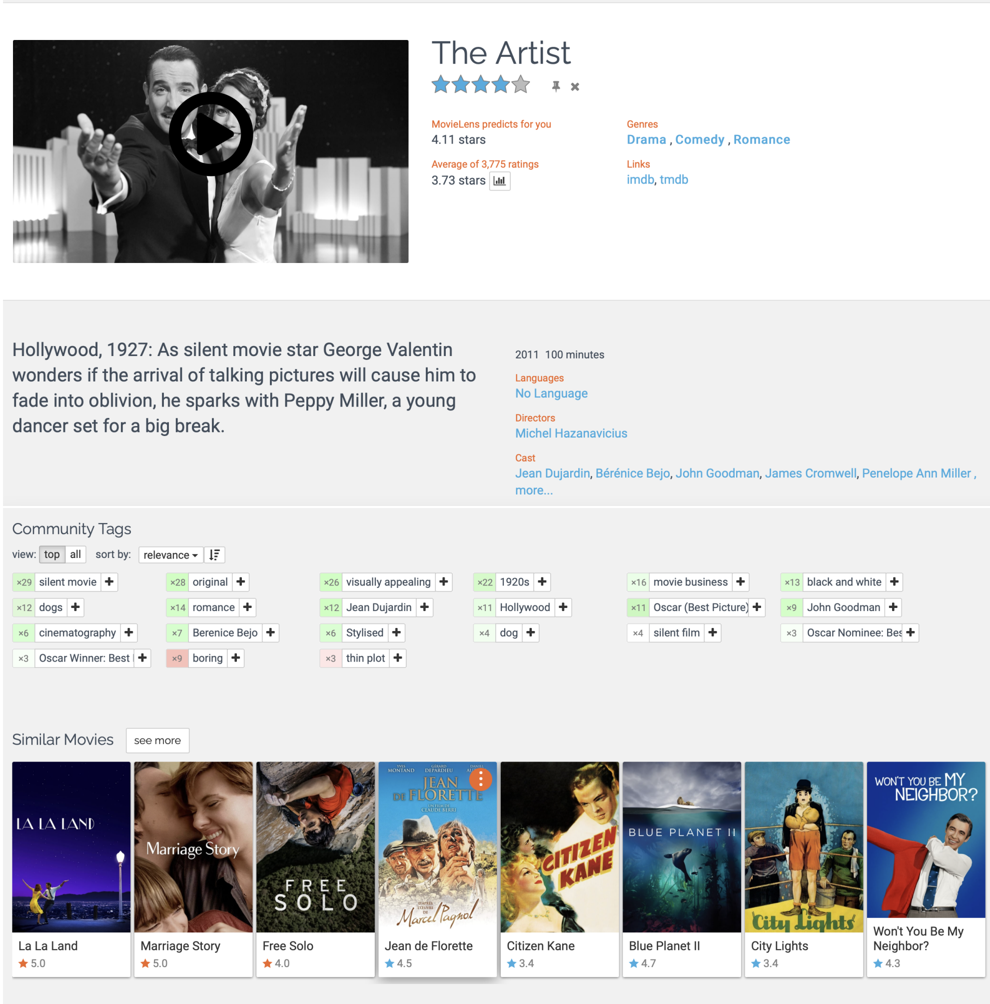


Figure 5: Movie Details Page

NOTES: This figure provides an example of a movie details page on the platform. Consumers can access it by explicitly clicking through the movie card on the homepage or by discovering it through the search bar. The details page includes the genre, synopsis, the ability to watch the trailer, a large set of community tags (Vig et al., 2012), similar movies, and other information about the movie.

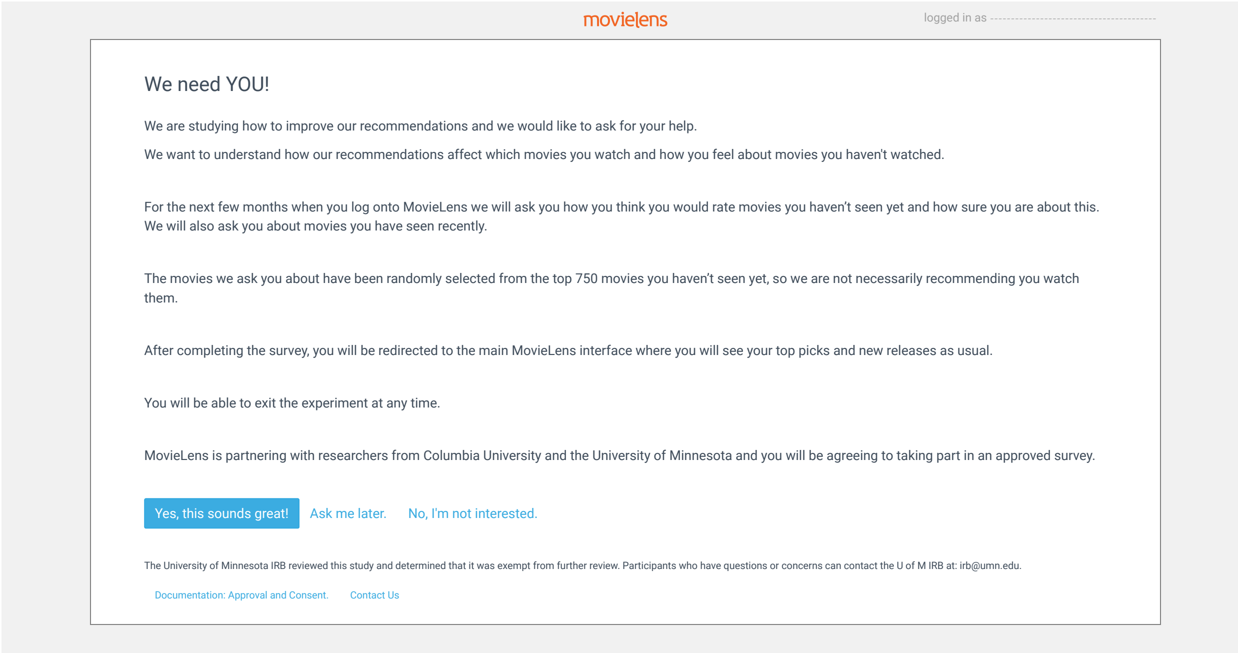


Figure 6: Recruitment

NOTES: This figure presents the recruitment dialog that eligible users of MovieLens saw at the beginning of the study.

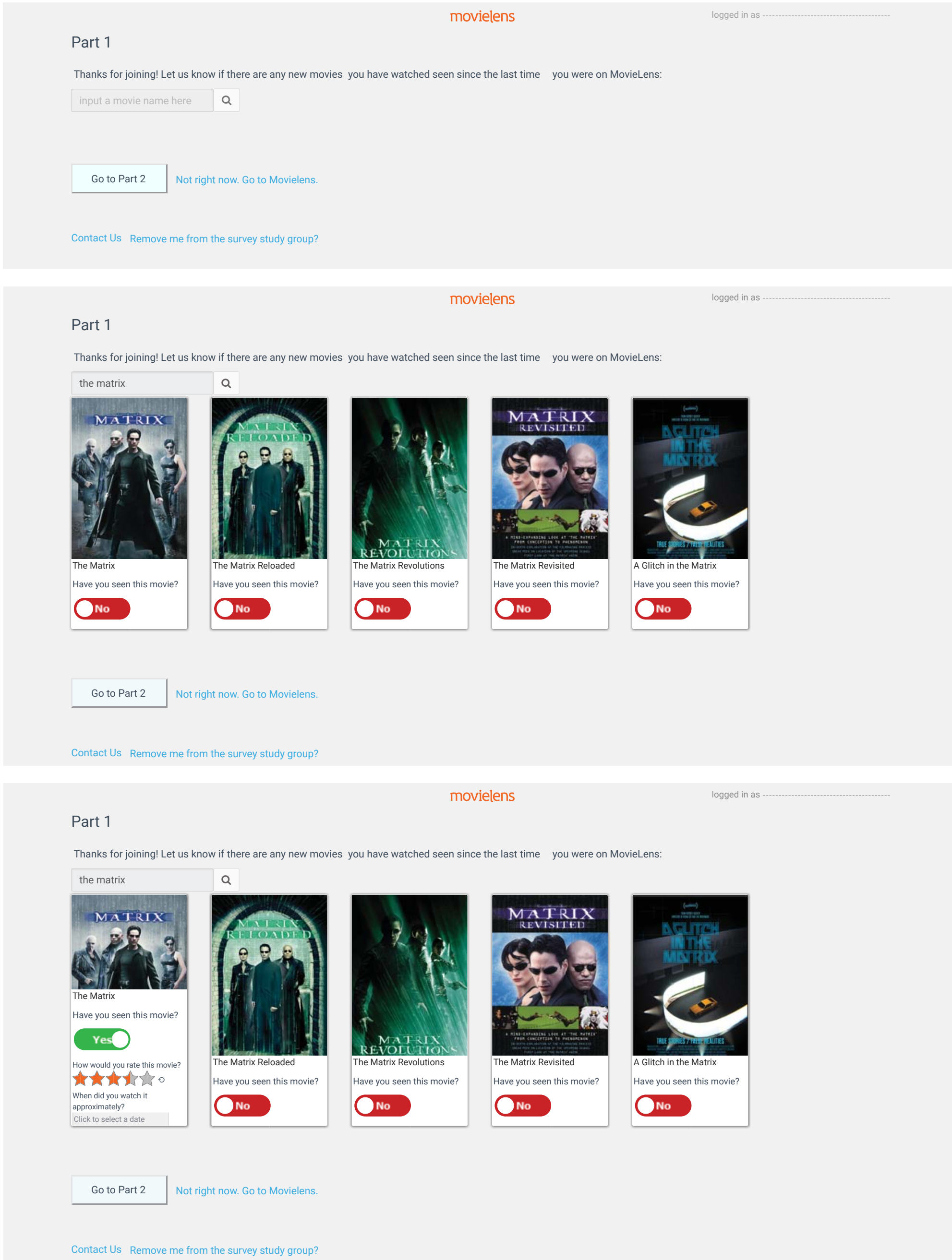


Figure 7: Movie Diary

NOTES: This figure provides the interface for the movie diary. This is the first part of the survey that consumers in the experiment complete when visiting the platform. The top figure shows the default interface when the survey loads where consumers are asked to manually input any movies they have seen since the last time they used MovieLens. The middle figure provides the interface for an example query (The Matrix) which functions similarly to the same relevance query as if the consumer inputted it on the homepage of the platform. The bottom figure provides the interface by which the consumer marks the movie as seen, indicating their rating and approximate watch date.

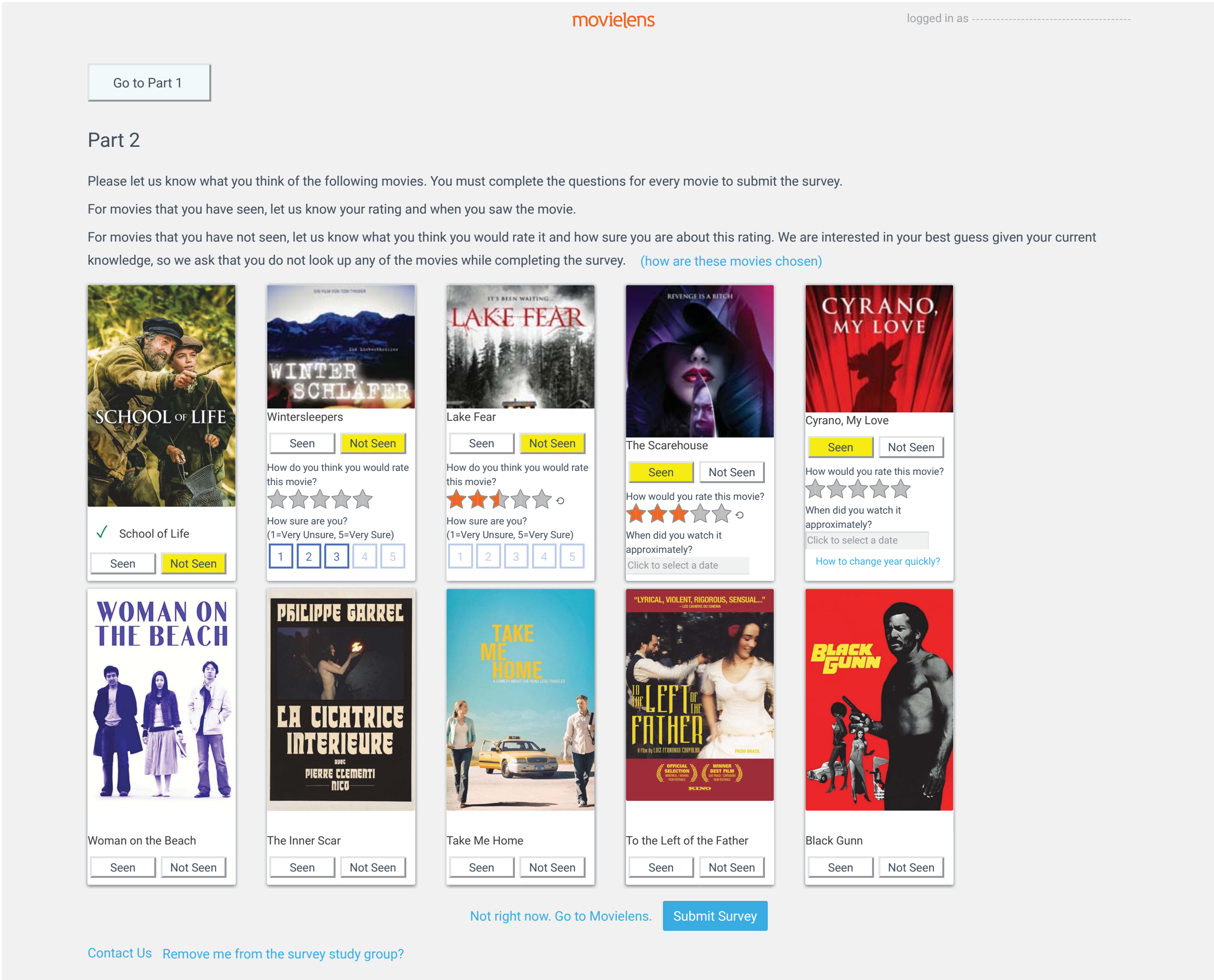


Figure 8: Belief Elicitation Survey

NOTES: This figure provides the interface for the belief elicitation survey. This is the second part of the survey that consumers in the experiment complete when visiting the platform. The figure presents the ten movies that consumers are asked about (chosen via the procedure shown in Figure 1). The first movie shows a completed elicitation (marked as unseen with the green checkmark). The second movie shows the interface that pops up when the consumer chooses not seen where the consumer is asked what they think they would rate the movie and their certainty. The fourth and fifth movie show the interface if the consumer chooses seen, where they are asked to rate the movie and provide their watch date.

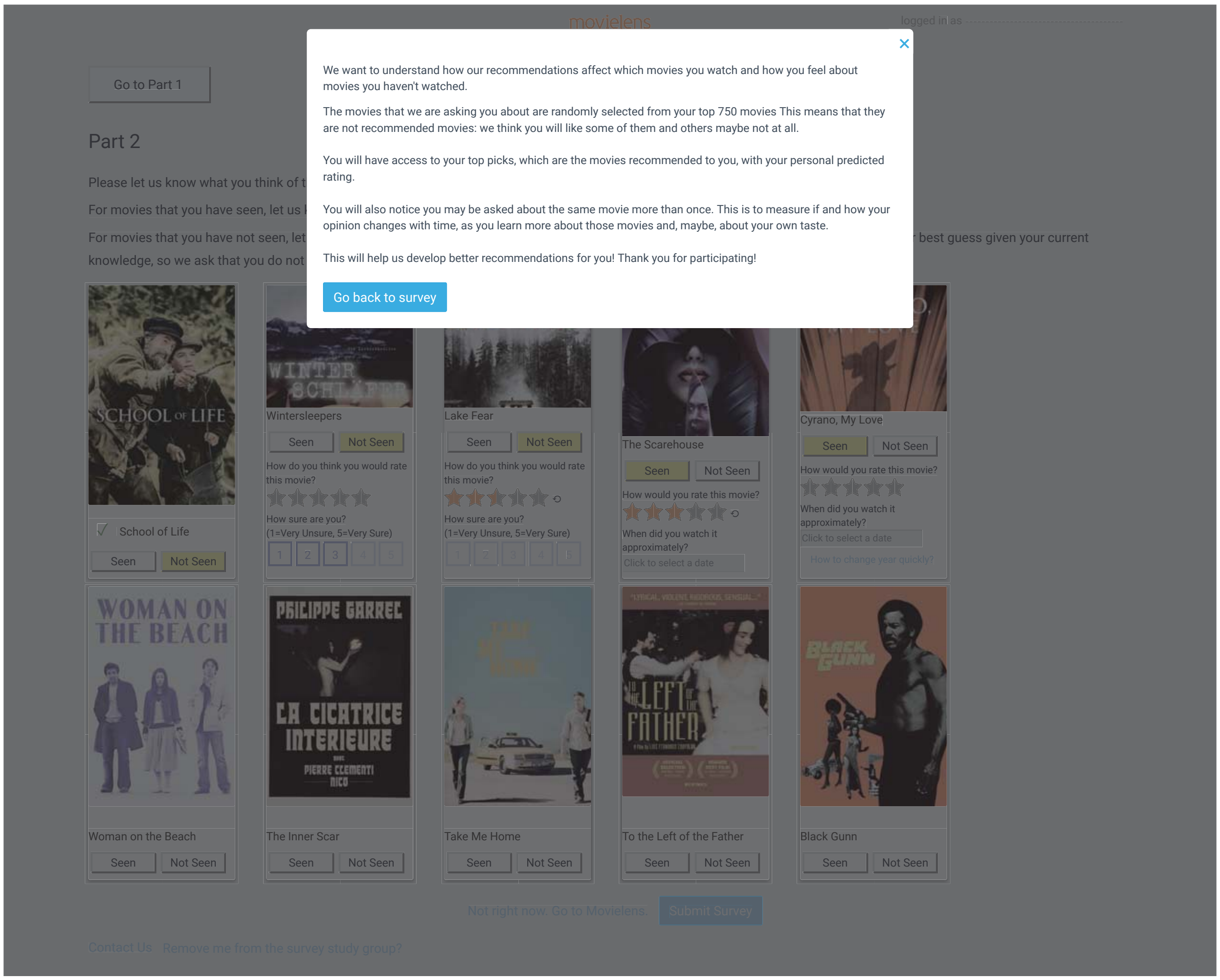


Figure 9: Belief Elicitation Explanation

Notes: This figure shows the informational dialog that pops up if they click on the “how are these movies shown" link in the instructions.