The Collective Dynamics of Sequential Search in Markets for Cultural Products

Pantelis P. Analytis
Hrvoje Stojic
Mehdi Moussaïd

SFI WORKING PAPER: 2015-06-023

The collective dynamics of sequential search in markets for cultural products

Pantelis P. Analytis[1], Hrvoje Stojic[2], and Mehdi Moussaïd[1]

[1]Max Planck Institute for Human Development
[2]Universitat Pompeu Fabra

June 23, 2015

Abstract

A few “hit” cultural products tend to dominate consumer attention. Yet, it is notoriously hard to predict ex ante which products will become “hits”. What are the decision-making processes that lead to these patterns of collective behavior? We advance a novel process model in which agents with diverse yet correlated preferences search the alternatives in order of popularity and choose the first alternative with utility higher than a certain satisficing threshold. The model goes beyond existing accounts of the popularity dynamics in that (i) it suggests a cognitive process through which social influence plays out in these markets, (ii) it allows us to study how inequality and unpredictability in the market change as a function of the diversity of preferences in the consumer population and the satisficing threshold, (iii) it is amenable to welfare analysis, and (iv) it facilitates comparisons with scenarios without social influence. In agent-based simulations we found that social influence led to an increase in inequality and unpredictability in the market, especially when agents employed a low satisficing threshold. In addition, we found that social influence led to a larger increase in the average consumer welfare when there was at least some diversity of preferences among consumers.

1 Introduction

In markets for cultural products such as books, music recordings, movies, or scientific articles, a great number of alternatives compete for consumer attention. People tend to consider only a small subset of the products available in the market and eventually buy only a few of the alternatives from those they carefully considered. There are two well-documented empirical observations about these markets. On the one hand, they are characterized by highly skewed popularity distributions a few "hit" products conquer a large market share, while many more attract only a few consumers (Merton, 1968; Mitzenmacher, 2004; Newman, 2005). On the other, it is notoriously hard to predict which products will become successful regardless of the tools at the forecasters disposal or a priori estimates of product quality (De Vany, 2004; Goel, Reeves, Watts, & Pennock, 2010). In this paper we report our search for answers to the following questions. First, what are the cognitive and social influence processes that lead to these patterns of collective behavior? Second, how do these processes operate on markets with different structural characteristics and lead to the aforementioned empirical patterns? Third, what is the impact of social influence on the average consumer welfare in the market?[1]

In the past, a number of different methodologies have been used to capture the empirical regularities of markets for cultural products. Rosen (1981) has pointed out that even in perfect knowledge markets small differences in quality can lead to superstar cultural products that capture most of the market returns. Yet, accounts based on quality differences alone cannot explain the observed unpredictability. Some of the first attempts to model both inequality and unpredictability were based on stochastic processes models (Yule, 1925; Simon, 1955b). These approaches capture these patterns quite well but do not take into account quality differences or postulate any decision-making processes (Price, 1976; Chung & Cox, 1994; Barabási & Albert, 1999). A second class of models suggests that the inequality and unpredictability in the market are driven by network externalities and socially acquired differences in the utility of the products or their cost of production. A host of models have illustrated that when consumers (producers) derive utility (reduce costs) from coordinating on the same consumption (production) choices as other consumers (producers), as expressed by the concept of network externalities, unequal and unpredictable consumption distributions might arise. Consumers may end up selecting an alternative that would be objectively inferior in the absence of social influence, simply because of their past observations or forward-looking beliefs about other peoples choices (Leibenstein, 1950; Adler, 1985; Elster, 1989; Brock & Durlauf, 2001).

A series of recent large-scale Internet experiments conducted by Salganik, Dodds, and Watts (2006) and Salganik and Watts (2008, 2009) provided compelling evidence that social information

[1]We would like to thank Mirta Galesic, Russel Golman, Amit Kothiyal, Shenghua Luan, Gäel Le Mens, Rasmus Pank Roulund, Tom Schmitz, Thomas Stork for their comments on previous versions of this work and Anita Todd for language editing. Mehdi Moussaid was supported by a grant from the German Research Foundation (DFG) as part of the priority program on New Frameworks of Rationality (SPP 1516) to Ralph Hertwig and Thorsten Pachur (HE 2768/7-2). This work was inspired by the Graduate Workshop in Computational Social Sciences, Modeling and Complexity organized at the Santa Fe Institute, which the first author had the chance to attend in June 2011.
Correspondence should be addressed to Pantelis P. Analytis (pantelispa@gmail.com).

displayed as the sum of other peoples choices leads to an increase in inequality and unpredictability in the markets. They designed an online artificial music market called Music Lab with thousands of participants. Participants in the market decided sequentially to first listen to and then download any number of 48 songs from relatively unknown bands in conditions with and without social influence. Salganik et al.'s work demonstrated that previous modeling accounts assuming no quality differences or a priori knowledge of the utility of the products are at best incomplete and illustrated that attention has to be paid not only to the sequence of consumer decisions, but also to the individual sampling and stopping processes. In addition, their experiments hinted at the rich welfare dynamics of these markets and showed that the large consumption inequalities observed in the markets may reflect actual differences in quality, but only partly so. Clearly, hit alternatives that are eventually consumed by a large part of the population have a great impact on aggregate welfare. Consider scientific articles or whole research programs that accumulate citations and are followed by numerous graduate students, while work of higher quality perhaps remains undiscovered.

The Salganik et al. experiments (Salganik et al., 2006; Salganik & Watts, 2008, 2009) have spurred a novel interest in modeling the collective dynamics of markets for cultural products. Borghesi and Bouchaud (2007) responded first to this new challenge. They constructed a model inspired by the Ising model of ferromagnetism and employed it to replicate the end of simulation results of the Music Lab experiment. In their model, social influence can be seen as a network externality parameter, which is added to the available public and private information when the agents choose whether to buy (download) a product. Krumme, Cebrian, Pickard, and Pentland (2012) went a step further and described the decision to buy (download) a cultural product as a two-step process. In their model, the agents first choose whether to sample an alternative and then whether to buy (download) it. Social influence enters as an availability parameter that controls for the probability that an item will be sampled given its position on the screen. Finally, Hendricks, Sorensen, and Wiseman (2012), leveraging the insights from informational cascade models (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992), developed a model in which agents with heterogeneous preferences follow a simple search model to decide whether to sample and then buy each alternative. By breaking down the process into two steps — a sampling and a buying step — the last two modeling accounts have made significant progress. Yet they are characterized by a crucial shortcoming. Each consumer is assumed to decide whether to sample and then select an alternative independently of the alternatives that he or she has sampled so far. In markets with a very large number of products this assumption is highly implausible and may lead to false generalizations.

The contribution of this paper is fourfold. First, we advance a novel cognitive process model popularity ordered search that reproduces the patterns of collective behavior in the markets for cultural products and provides new insights about the rich-get-richer dynamics. We then show how the inequality and unpredictability in the market on the one hand and the aggregate consumer welfare on the other vary as a function of the interaction between the cognitive process and crucial characteristics of the environment, such as the diversity of preferences in the consumer population. Last but not least, as in the Salganik et al. (2006) experiments, markets with social interaction can be directly compared to scenarios without it, which enables us to measure and then possibly manage the impact of social influence.

1.1 Our modeling approach

We sought to develop a simple model of an individual decision-making process with assumptions that reflect human cognitive capacities and the most salient characteristics of the markets for cultural products. We started with some key observations about the markets: People can learn the true utility of a product fairly quickly. For example, before a purchase, people can listen to a song, watch a movie trailer, skim through a book, or read an abstract of a scientific article to assess its utility. However, to learn the utility people examine the products one by one and pay a nonnegligible cost in terms of time or effort. For instance, people can read only one abstract at a time and usually pay a cost in time spent reading and not doing something else. Further, people are constrained by their time or budget to consume only a limited number of these products. Finally, simply because there are a great number of products in the environment, people do not have the time resources to examine each one before making a choice. Thus, consumers necessarily consider only a subset of the total number of alternatives available in the environment.

These elements of the decision-making process can be neatly and formally captured by a sequential search model (e.g., Simon, 1955a; DeGroot, 1970; McCall, 1970; Telser, 1973). In our model each agent makes decisions following this process. Intuitively, such an agent stops searching after encountering a good-enough alternative, although better alternatives may remain undiscovered in the market. This can be expressed in the form of a stopping threshold: The agent continues to sample alternatives until he or she encounters one that satisfies this threshold. When certain assumptions are satisfied and when agents make decisions in isolation, optimal solutions can be computed. There are experimental results showing that individual behavior in such problems can be approximated by threshold decision strategies (Rapoport & Tversky, 1970; Hey, 1987).

Further, in markets for cultural products social influence shapes the interaction among consumers — in our highly interconnected world consumers often possess some information about the popularity of the alternatives, or their information on available alternatives is influenced by popularity. This is reflected in the popularity of "best of" lists for movies, books, songs, albums, and so on. Popularity affects the probability that a person will hear a song on the radio (Hendricks & Sorensen, 2009); people are more likely to talk about popular products with their friends; news websites are more likely to report on such products; and in online stores items are often by default ordered according to their popularity, or consumers can easily do so. At the cognitive level, popularity is also reflected by the fluency with which people process the alternatives in memory before they more carefully consider them (Hertwig, Herzog, Schooler, & Reimer, 2008). We assumed that the agents would act as they do with a bestseller list and would start by examining the alternative that has been selected by most agents and move down to less popular alternatives.

Conceptually, our model incorporates some of the notions of existing social influence models such as sequentiality among agents, threshold rules, and heterogeneity in preferences (Granovetter

& Soong, 1986; Arthur, 1989; Banerjee, 1992; Hendricks et al., 2012) and illustrates how these can be operationalized in multi-alternative choice environments. However, it goes beyond these models as it illustrates that a simple cognitive mechanism — limited sequential search — captures the patterns of collective behavior, such as skewed popularity distributions, outcome unpredictability, and the imperfect relation between perceived quality and popularity, better than previous theoretical accounts.

2 The model

2.1 The environment

There are N alternatives (or goods) $X_1, ..., X_N$ in the market, populated by M agents $A_1, ..., A_M$. The alternatives have an objective utility component u_o, which is identical for all the agents, and a subjective component u_s, which is agent specific. The overall utility of an alternative u_n for an agent is then a simple sum of these two components $u_n = u_{no} + u_{ns}$. The objective component u_o of each alternative is a draw from an independent and identically distributed (iid) random variable, normally distributed with mean μ and variance σ_o^2. The subjective component u_s of each alternative is iid normal with the same mean μ, but with a different variance σ_s^2. The overall utility is then a sum of two draws from iid normal variables, which itself is an iid normal variable with variance $\sigma^2 = \sigma_o^2 + \sigma_s^2$.

The agents encounter alternatives sequentially and can learn the utility of an alternative u_n only by sampling it and paying a cost c. We refer to c throughout the text as the search cost. The agents can sample as many alternatives as desired, but they can choose only one of the alternatives they have sampled. An example of sampling behavior could be listening to a couple of songs or skimming through several books and finally buying one of them. We assume that the search cost is constant and that there is no post sampling uncertainty — by sampling the alternative the agents learn its true utility. The basic framework can be easily extended to cases where the agents can choose more than one alternative or cases where some uncertainty remains even after sampling.

An agent's return function can be summarized as,

$$R_n(u_1, u_2, ..., u_n) = max(u_1, u_2, ...u_n) - n \times c \tag{1}$$

Note that the returns depend on the utility of the best alternative discovered so far and the search cost. The returns from sampling one more alternative can be formulated as

$$\begin{aligned} R(u_n, c) \quad = \quad & Pr(u_n > max(u_1, u_2, ...u_{n-1})) \times \\ & E(u_n - max(u_1, u_2, ...u_{n-1}) | u_n > max(u_1, u_2, ...u_{n-1})) - c \end{aligned} \tag{2}$$

When viewed on the level of a single consumer, this is a classical optimal stopping problem—examined extensively in statistics and economics (e.g DeGroot, 1970)—in which after examining a new alternative the agents decide whether to sample further or to stop search. Optimal stopping problems with a payoff function defined this way are called stopping problems *with recall*.

In studying the effects of social influence on people's behavior, it is important to have baselines against which the effects of search behavior and social influence on the structure of the market and the welfare outcomes can be measured. In network theory, for example, the random graph model continues to serve the role of baseline against which other network structures are compared (e.g. Watts & Strogatz, 1998). For this reason, we examine two types of search scenarios — one in which agents do not have access to popularity information and one in which they do. We call the first market a random search market (random search, for short) and the second a popularity heuristic market (popularity heuristic, for short). When the cost of search is 0, these two processes converge to the perfect knowledge assumption of neoclassical models (for a similar experimental comparison see Reutskaja, Nagel, Camerer, & Rangel, 2011).

2.2 Random search

The agents do not have access to any social information and they search the alternatives in random order. This scenario has been studied extensively and complete theoretical treatments can be found in Chow, Robbins, and Siegmund (1971). An optimal threshold T can be found for which the return from sampling one more alternative is zero, $R(u_n, c) = 0$. A standard result is that as the cost of sampling increases, the optimal threshold T decreases. Intuitively, for a high cost, an agent is willing to sample more only if the highest u discovered thus far came from the lower part of the distribution. Moreover, the return is always positive as long as the utility of the best sampled alternative is lower than the optimal threshold.

When sampling from a distribution with known variance and mean, the optimal threshold depends only on the cost of sampling, $T(c)$. This result holds irrespective of the number of alternatives that remain to be sampled. The optimal sampling policy can be formally summarized as:

$$\begin{aligned} &\text{if} \quad u_n > T_{opt}(c), \quad \text{stop sampling} \\ &\text{if} \quad u_n \leq T_{opt}(c), \quad \text{continue sampling} \end{aligned} \tag{3}$$

The agents $A_1, ..., A_M$ make their choices sequentially. An agent is chosen randomly without replacement. The agent then searches through alternatives and chooses one of them.[2] This process is repeated until all M agents have made their choice and the market outcome is known. Since agents do not make use of social information we use random search as a reference frame to evaluate social influence.

2.3 Social Interaction

The choices made by searchers generate social information in the market. This information can then be leveraged by agents following the popularity heuristic. After the first agent, a second or m^{th}

[2] In our model we do not allow our agents to strategically delay their decisions.

agent is asked to make its choice, again at random without replacement. Now the agent can observe the sum of choices made by previous searchers in the form of a popularity vector $P = \{P_1, ..., P_N\}$ that records the choices of the corresponding alternatives $X_1, ..., X_N$ and is updated whenever an individual makes a choice. Examples of such popularity information would be product sales, number of song downloads, or number of article citations. Thus, the only point of social interaction is via the publicly available information on popularity P.

2.4 Popularity heuristic

The agents sample the alternatives $X_1, X_2...X_n$ in decreasing order of popularity $P_1 > P_2 > ... > P_n$, where popularity is defined as the number of times that an alternative has been selected in the past. When there is a tie, the agents choose which one to sample next at random. Thus popularity heuristic can be seen as a heuristic, where popularity is the only cue considered by the agents to order the alternatives (for single attribute heuristics see Bagwell & Ramey, 1994; Hogarth & Karelaia, 2005). We call the order in which the alternatives are searched the *search path*. The agents stop search when they encounter an alternative with utility higher than a threshold T_{pop}. The returns from examining one more alternative can be expressed as in (2). The sampling policy is as follows:

$$\begin{aligned} &\text{if} \quad u_n > T_{pop}, \quad \text{stop sampling} \\ &\text{if} \quad u_n \leq T_{pop}, \quad \text{continue sampling} \end{aligned} \tag{4}$$

Although the popularity heuristic is behaviorally intuitive it is not closed-form optimal. A closed-form optimal solution would require that the agents infer the expected utility and the distributions of utilities for the alternatives available in the market on the basis of the current popularity information. In addition, a fully rational policy should be based on the assumption that other individuals are acting in a closed-form optimal manner. Such an approach is computationally intractable for both artificial and human agents. The popularity heuristic is boundedly rational. In both random search and the popularity heuristic, the threshold employed by the agents can be seen as a parameter in the model. In random search one can always tune the parameter to correspond to the optimal behavior. In the popularity heuristic, by contrast, no threshold level exists for which optimality is guaranteed.

In both search rules, when all the alternatives in the market have been searched and no alternative has utility higher than T, the agent simply selects the alternative with the highest utility. Further, we assume that the agents can refrain from sampling any alternative when the cost of search is very high. In that case the agents are satisfied with the status quo outcome.

2.5 Assumptions

Here we make transparent and justify the modeling choices made in the previous section. First, we have assumed that the agents employ a fixed threshold, which in random search corresponds to the optimal stopping rule. This is a simplifying assumption, as in real life, people learn stopping rules gradually from experience in previous realizations of the market. Conlisk (2003) has shown how agents in random search problems can gradually converge to the optimal threshold by following a simple learning process.

Second, we have assumed that the agents can always recall alternatives that they have sampled in the past. This assumption is well suited for the markets of cultural products as the consumers can in almost all cases find music albums, books, or papers that they have listened to, read, or heard of in the past. Internet marketplaces and search engines have contributed a great deal to making previously examined cultural products available at any time.

Third, we have assumed that agents use only the popularity information to order the alternatives. In some scenarios this assumption may reflect actual consumer behavior (e.g., consider most-read lists that are often widely used in online press), but in many cases agents have access to more information and they can use it to order the alternatives. Our model can be seen as a boundary case of an ordered search model in which several pieces of information are processed (e.g., Bagwell & Ramey, 1994; Armstrong, Vickers, & Zhou, 2009; Analytis, Kothiyal, & Katsikopoulos, 2014). This strong assumption allows us to focus on the interplay between sequentiality in choice and diversity of agent preferences in the environment.

Fourth, we assumed that the products do not have a price tag and that they can be consumed by an infinite number of agents. Indeed, in terms of production, cultural products are nonrivalrous information goods with a potentially unlimited supply. Clearly, after publishing a scientific article a great number of scientists can read it without increasing the marginal cost of production. If a universally best alternative existed, for example, an outstanding online course in statistics, everybody would be better off by consuming that alternative without restricting the supply and increasing the price for other individuals. Most cultural products are excludable goods; that is, a price can be assigned to them. However, for many cultural products, for example, movies, music albums, and books, there tends to be a small price dispersion and in some cases there is even a predefined price (as in cinema tickets) for an entire product category. Yet sometimes, they are supplied freely by producer choice, as, for example, in the case of YouTube videos, massive online courses, or designer products distributed under a creative commons license. In the latter case cultural products can be categorized as pure public goods.

This brings us to the last major assumptions of the model. We have assumed that consumers have diverse preferences. The research on recommender systems demonstrates that this is clearly the case in all markets for cultural products (for an example see Analytis, Barkoczi, & Herzog, 2015). In fact, the degree of diversity in preferences varies in different cultural markets. We have opted to describe the subjective and objective utility components with a normal distribution for its neat aggregation properties and its plausibility, which allow us to vary the diversity of preferences while keeping overall variability in quality stable.

3 Results

We simulated markets consisting of 1,000 agents that decided sequentially which of 100 alternatives to choose. For each market we drew the objective utilities u_o of the alternatives from a normal distribution with mean equal to 0 and variance equal to σ_o^2. For each individual agent in the market we drew the subjective utilities u_s of the alternatives from a different normal distribution with mean equal to 0 and variance equal to σ_s^2. For all the analyses reported in the results section we assumed that agents following random search employ the optimal stopping threshold. Then, we fixed the threshold parameter of the popularity heuristic to correspond to the optimal threshold for random search. We chose that implementation for two reasons. First, even if we were to test slightly different threshold rules, the collective behavior patterns would be very similar. It will soon become clear that the collective behavioral patterns are primarily determined by the ordering of the alternatives, while the stopping threshold moderates their intensity. Second, in this way we can directly compare the welfare outcomes of the two models at the individual and aggregate level.

We systematically varied three parameters in our simulations. First, we varied the contribution of the objective utility u_o and the subjective utility u_s in the overall utility of the alternatives available in the market. This was done by setting the subjective variance σ_s^2 equal to d, where $d = \{0, 0.1, 0.2, ..., 1\}$ is a diversity vector, and the variance of the objective utility for each individual agent σ_o^2 equal to $1-d$. Second, we varied the cost of search c that agents have to pay when sampling an alternative. We tested five possible costs, $c = \{1/2^3, 1/2^4, 1/2^5, 1/2^6, 1/2^7\}$, which correspond to the thresholds $T = \{0.78, 1.15, 1.47, 1.76, 2.03\}$. In this way we covered a large space of plausible costs encountered in actual markets for cultural products. When $c = 1/2^3$, the cost of searching one more alternative is a significant proportion of the average utility eventually enjoyed by the agents, whereas when $c = 1/2^7$, the cost of search is negligible. We opted to use the cost consistently on the x axis rather than the threshold although they can be used interchangeably. Note that in the first part of the Result section only the thresholds need to be known, whereas for the welfare analyses both the thresholds and the costs of search need to be known. Finally, we varied the decision-making process—random search or the popularity heuristic. Overall, this resulted in $11x5$ possible market environments for random search and the popularity heuristic, respectively. To account for randomness in the whole process, we ran each condition 100 times. In all the analyses conducted, except when we focused on unpredictability, we generated all the alternatives in the environment anew.

In the following sections we first examine the interplay between search rules and the structural characteristics of the market, such as market share inequality and outcome unpredictability, and then we turn to the welfare analysis.[3]

[3]We wrote the code for running the simulation, analysing the results, and producing the figures in R language (Ihaka & Gentleman, 1996), and it is publicly available at the website: https://github.com/pantelispa/popularityHeuristic. The exact dynamics of the original simulation can be reproduced in https://hstojic.shinyapps.io/popularity/. We programmed the interactive environment with the Shiny application. Note that the application requires a strong computational system to run properly. In some systems it takes up to a minute per condition to load the original data of the simulation.

3.1 Utility, popularity and inequality

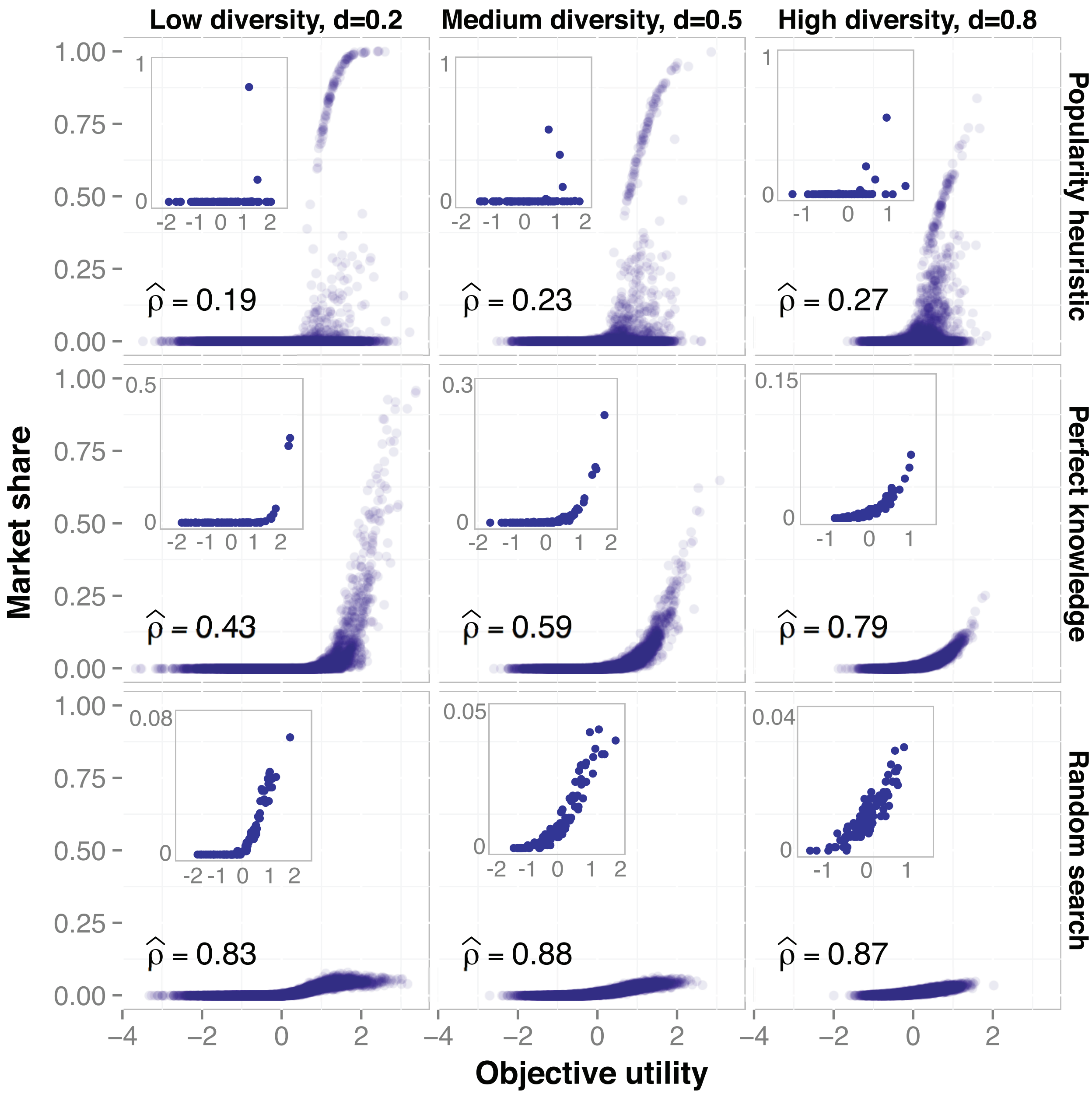

Figure 1: Top and bottom row: The market share of all 100 alternatives in all 100 repetitions of the popularity heuristic and random search rule for the conditions with cost of search $c = 1/2^3$ and diversity of preferences $d = \{0.2, 0.5.0.8\}$. Middle row: The perfect knowledge condition ($c = 0$). The inset plots depict the results from a randomly drawn single run of the simulation. In the popularity heuristic condition, the objective utility of the alternatives is only weakly related to the obtained market shares at the end of the simulation. In contrast, in random search the market share is a very accurate predictor of the utility of the alternatives. We report the Pearson correlation in all the conditions.

Let us first examine the patterns of aggregate behavior implied by each of the search rules. The popularity heuristic markets are characterized by evident rich-get-richer dynamics. The decisions of the first agents shape the *search path* that will be followed by agents deciding after them. Over time the *search path* tends to stabilize, which implies that the agents examine the alternatives in approximately the same order and tend to choose from the same subset. When the agents employ a lower threshold they examine a smaller subset of alternatives and swarm on fewer of them. In contrast, in random search markets for each agent each alternative that satisfies the threshold stands an equal chance of being selected ($1/k$, where k is the number of alternatives crossing the threshold). Thus, a lower threshold implies that more alternatives stand an equal chance of being selected. Inversely, a higher threshold means sampling more broadly but choosing more selectively.

Note that for cost of search 0, both search processes converge to a perfect knowledge market (Figure 1 middle row), because the agents sample all the alternatives regardless of the search rule. Then, a convex relationship between objective utility and market share exists that is especially pronounced when the subjective utility component is small. Thus perfect knowledge markets fulfill the assumption put forward by Rosen (1981) in his famous superstar model. However, as the cost of search increases (stopping threshold decreases), the two processes diverge, leading to very distinct aggregate behavioral patterns.

At the end of the simulation under the popularity heuristic, many alternatives with high objective utility are never sampled or are sampled by very few individuals (see the inset graphs in Figure 1). Clearly, there is a great deal of luck involved at the first few steps of a market, and having a high objective quality is necessary but not sufficient for turning a product into a hit. Although it is possible to reliably predict that low objective quality alternatives will flop, it is impossible to foretell which of the high objective quality alternatives will become "hits". In contrast, the market share captured by the different alternatives increases gradually as a function of the objective utility component of the alternative in random search gradually converging to the steep convex relationship of the perfect knowledge condition. In the popularity heuristic market, the objective utility is only weakly correlated with the obtained market share, while in random search more alternatives are chosen at least once and the correlation is much more pronounced.

As in Salganik et al. (2006), we used the Gini coefficient to measure the inequality. The coefficient is defined as $G = \sum_{i=1}^{n} \sum_{j=1}^{n} \mid m_i - m_j \mid /2n \sum_{k=1}^{n} m_k$. The term m_i stands for the market share of an alternative that is defined as $m_i = d_i / \sum_{l=1}^{n} d_l$, where d_i is the number of times the alternative was selected so far and d_l is the sum of choices made until decision maker l. The results for all the 110 conditions are shown in Figure 2. Lower costs of search (higher stopping thresholds) imply a lower Gini coefficient for the popularity heuristic, whereas they have the opposite effect for random search. Note that as the cost of search decreases (stopping threshold increases) the Gini coefficients of the two processes gradually converge to those of the perfect knowledge markets, which are $G = \{1, 0.98, 0.96, 0.93, 0.9, 0.85, 0.8, 0.72, 0.62, 0.46, 0.18\}$ for diversity of preferences $d = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, respectively.

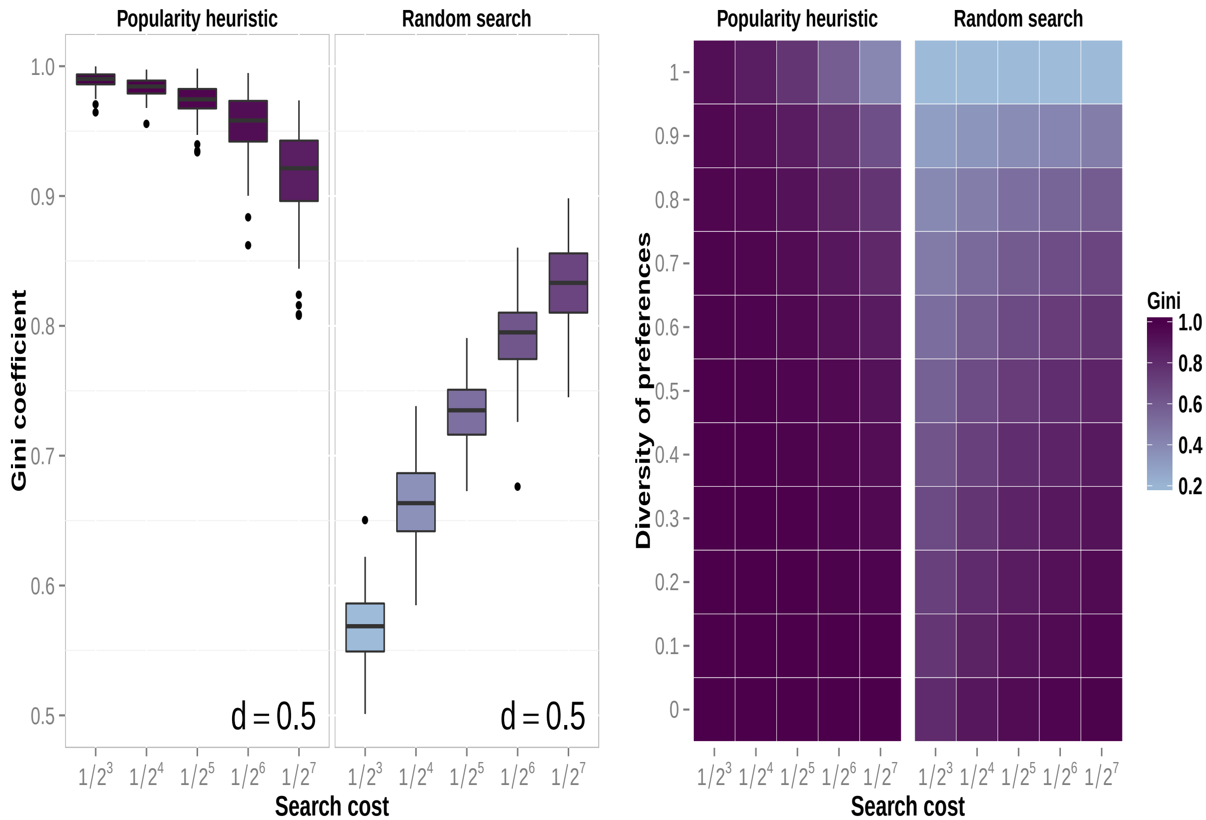

Figure 2: The average Gini coefficient in a market with $d = 0.5$. In the popularity heuristic market a higher search cost (lower stopping threshold) leads to more inequality in the market. In contrast, in the random search market it implies a more egalitarian distribution. As the cost of search drops the Gini coefficient of the two processes converges to that of a perfect knowledge market. Right: The average Gini coefficient for all 55 popularity heuristic and random search markets.

3.2 Outcome unpredictability and luck

How accurately could we predict the popularity (or market shares) of the alternatives if we were able to rerun the entire market with the same initial conditions? To answer this question we generated 10 different distributions of objective utilities, which we refer to as worlds, and simulated each of these worlds 10 times, every time with new agents. In this way we balanced the variability that could be caused by the environment — consider a randomly generated world with an extraordinarily good alternative — ,with the need for additional replications of the same environment. Following Salganik et al. (2006) we defined the unpredictability related to a specific alternative as $v_i = \sum_{j=1}^{W} \sum_{k=j+1}^{W} \mid m_{ij} - m_{ik} \mid / \binom{W}{2}$, where m_{ij} is the market share of an alternative i, in world j and W is the number of worlds. We then defined the overall unpredictability in an economy as $V = \sum_{i=1}^{n} v_i/2$. Following this formula the results have an intuitive interpretation that holds for any number of alternatives. Unpredictability 1 corresponds to the case where the entire market share goes to different alternatives in every single replication (world) of the simulation. Unpredictability

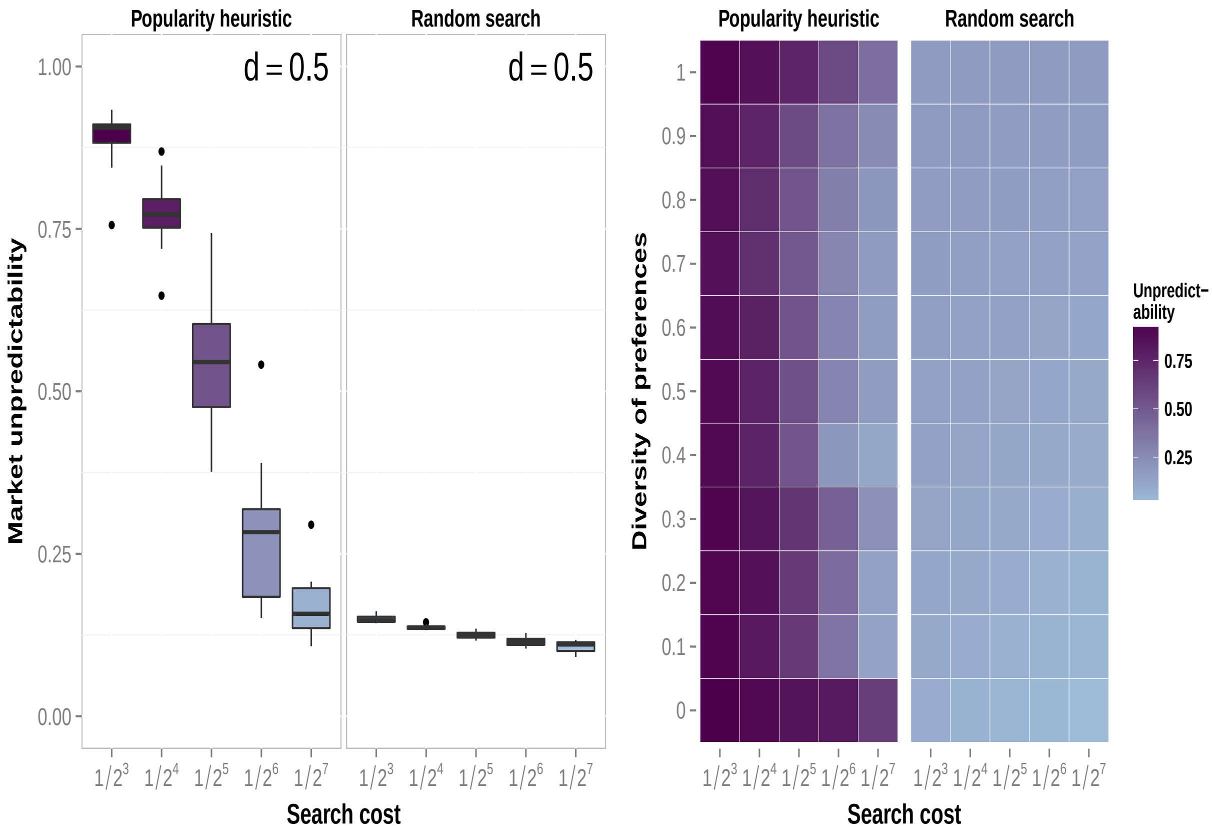

Figure 3: Left: The unpredictability in the market as a function of the cost of search when d = 0.5. For both the popularity heuristic and random search markets, the unpredictability in the market decreases as a function of cost of search, converging to 0, which corresponds to the perfect knowledge market. The convergence is much steeper in the popularity heuristic market, where for high cost of search the market is very unpredictable. The variability of the unpredictability coefficient is large in the popularity heuristic market, whereas it is negligible in the random search market. Right: The unpredictability coefficient for all 55 conditions in the popularity heuristic and random search markets.

0 corresponds to the case where the market share is distributed identically in all the replications (worlds) of the simulation.[4]

As depicted in Figure 3, both search processes converge toward unpredictability 0 as the cost of search decreases (threshold increases), which corresponds to a perfect knowledge market. The popularity heuristic markets are very unpredictable for high costs but become fairly predictable for low costs. In contrast, the unpredictability reduces at a slow pace in the random search markets. Similar results are obtained for all values of d. In the popularity heuristic markets, unpredictability as a function of diversity of preferences is bimodal — the lowest levels of unpredictability are

[4]Salganik et al. (2006) and Salganik and Watts (2008, 2009) defined the overall unpredictability as $V = \sum_{i=1}^{n} v_i/S$ where S is the number of alternatives in the market (in their case, songs). Borghesi and Bouchaud (2007) and Krumme et al. (2012), who have developed models that can capture the dynamics of the Music Lab experiment also used this formula in their studies. However, according to this formula the maximum degree of unpredictability depends on the number of alternatives available in the market leading to results that can't be compared between studies with a different number of alternatives.

obtained for intermediate levels of d, while it increases as d goes to 0 and 1. Reasons for higher unpredictability in extremes differ. As d goes to 0, agent preferences become correlated and first movers are dictators of the market, but what agent is the first mover is random from world to world. On the other hand, as d goes to 1, agent preferences differ a lot and each world is driven by idiosyncratic consumer preferences, which are world specific. In contrast, for intermediate levels of uncertainty agents search a bit longer and for varying lengths down the same *search path*. This leads to the accumulation of more valuable social information. Different agents often choose the same alternatives with a high objective utility component. Note that while both environments are uncertain, random search markets are predictable in the long run, whereas the popularity heuristic markets are not. The popularity heuristic markets are nonergodic, which means that small events have irreversible consequences in the course of history, whereas the random search markets are ergodic random processes that converge to approximately the same results for large samples (for more details on non-ergodicity see Arthur, 1989).

3.3 Welfare analysis

In Figure 4, we present the average utility obtained in 55 different popularity heuristic conditions. First, note that in random search the average utility in the market does not depend on the diversity of preferences in the population but merely on the search cost variable. As a result, there are only five relevant conditions that are perfectly equivalent, in terms of average utility and cost of search, with the popularity heuristic conditions with $d = 1$. When $d = 1$ the environment is recreated randomly for each agent. Thus, even though the agents search according to the popularity order, they essentially behave the same as agents that search randomly. In these conditions the agents incur the highest average total cost of search and earn the lowest average net utility from the market. To find out the extent to which the popularity heuristic outperforms random search in various diversity of preferences conditions, the reader can simply compare them with the $d = 1$ condition. Also note that in a perfect knowledge market ($c = 0$), the agents always consume the alternative with the highest utility in the market, which on average has utility $u = 2.51$.

We find that as d decreases, the cost of search steadily declines, as well. When all the agents have exactly the same preferences, the first agent pays total search cost $k * c$ (where k stands for the number of alternatives sampled until the agent encounters an alternative with utility higher than T). All the remaining agents simply imitate the decision made by the first agent and pay c only once. Following the terminology introduced by Page (2006), the process is initial-outcome dependent. All the agents deterministically follow the choice of the first individual. Surprisingly, although the average agent in markets with $d > 0$ incurs a higher average cost of search than the average agent in a market with $d = 0$, this loss is outweighed by the benefits of choosing alternatives with higher gross utility. For high search costs, for example, intermediate levels of d lead to the highest average net utility. As the cost of search declines, however, the amount of beneficial diversity of preferences decreases. Consequently, the differences in the average utility among the conditions decline, as can be seen in the rightmost columns of the net utility heat map

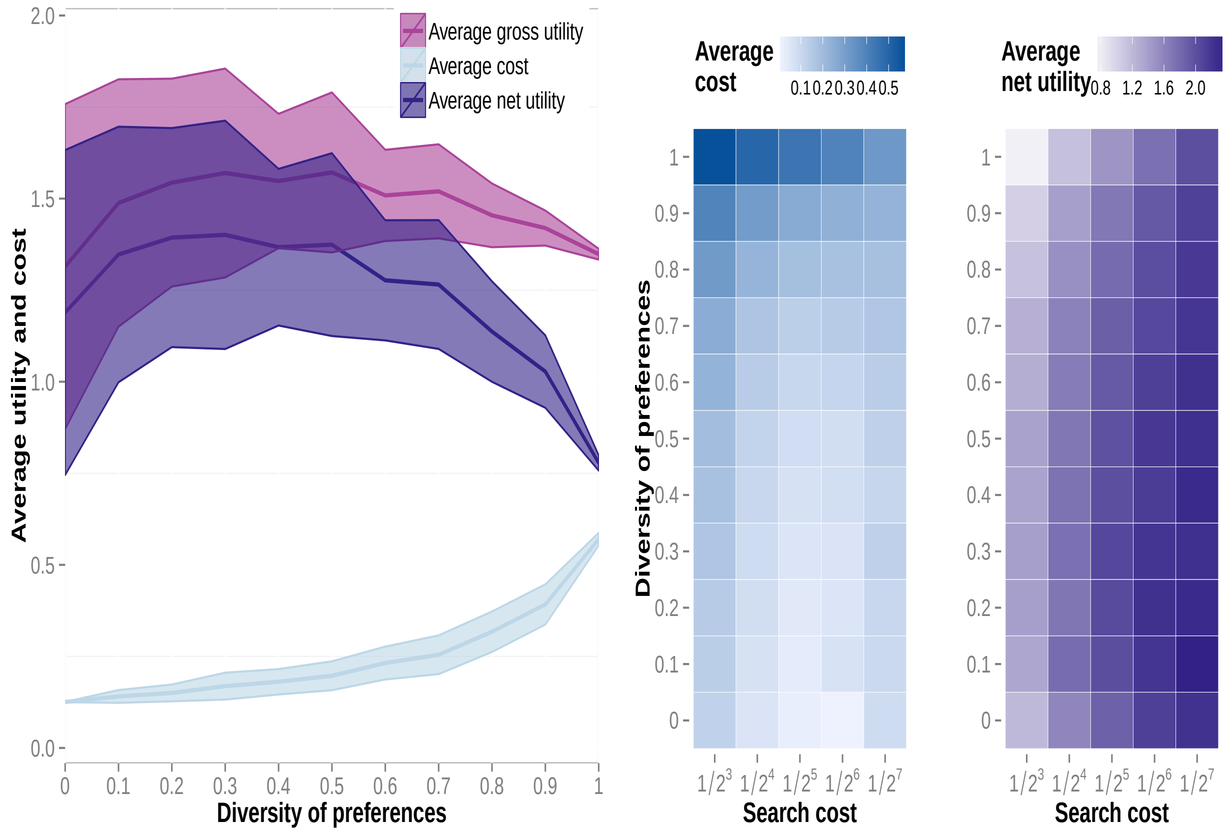

Figure 4: Left: The average gross and net utility of the selected alternative and the average total cost of search for search cost $1/2^3$. The ribbons represent the variability in 100 repetitions of the simulation. The average net utility in the market is the highest for intermediate levels of diversity of preferences. Right: The average net utility and the average total cost of search in the 55 popularity heuristic conditions. The graph on the left side corresponds to the leftmost columns of the heat maps on the right side.

(Figure 4). Finally, because as d decreases agents tend to settle more often on the choices made by the very first agent(s), and there is luck in who moves first, a lower d implies more variability in the average utility in the market. The first agent(s) may settle on a very good or a simply passable alternative. This is illustrated by the ribbons in the left panel of Figure 4.

This analysis revealed an unexpected result. Although the costs of search are lowest in markets with no diversity of preferences, the average net utility is higher in economies with moderate diversity. What is the process that leads to an increase in the average obtained utility in the market in such conditions? To gain insight into the process, we examined the evolution of net utility in the market. We plotted the average net utility enjoyed by the agents $\sum_{i=1}^{n} u_i/b$ in sequential blocks of 25 agents for search cost $1/2^3$ (left panel in Figure 5). This analysis reveals that in contrast to a market without diversity where the average utility remains constant, in markets with some diversity the average utility continues to increase in the first agent blocks. As discussed above, in a market with no diversity whatsoever, everybody is satisfied with a decision made by the first individual. Occasionally this turns out to be an excellent alternative, but more often it is located just off the satisficing threshold. In contrast, in markets with some diversity of preferences, alternatives with a

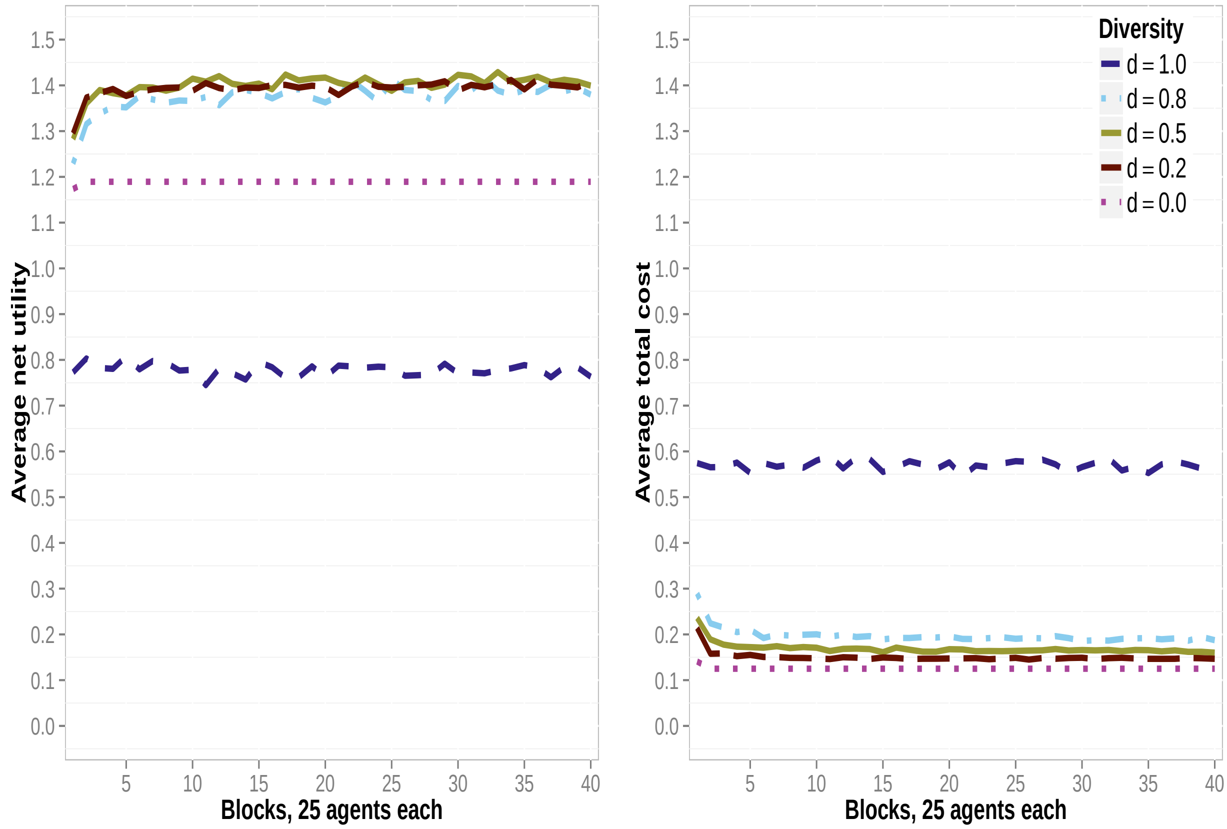

Figure 5: The average net utility in the market $\sum_{i=1}^{l} u_i/25$ (left panel) and the average total cost of search $\sum_{i=1}^{l} k_i * c/25$ (right panel), where l was set to blocks of 25 agents, for search cost $1/2^3$. When $0 < d < 1$, the average net utility in the market tends to increase as more agents make their choices, but with diminishing returns with each additional agent. The average total cost in the market has the opposite pattern: it decreases with diminishing returns with each additional agent. The largest increase (decline, respectively) occurs in the very first agent blocks.

higher objective utility component tend to be gradually pulled toward the beginning of the search order. Because preferences are not perfectly correlated, agents will search a bit further until they encounter a satisficing alternative. The *search path* followed by the agents gradually improves in terms of the encountered objective utilities after the decisions of several first agents. Consequently, alternatives with higher objective utility components, which in some cases are even higher than the threshold, tend to conquer the first spots in the *search path*. Thus, the agents deciding later in time have a more useful popularity signal, require fewer search steps to find an alternative that satisfies Condition (Equation) 4, and tend to settle on much better alternatives. In essence, some diversity of preferences allows the agents to benefit from the informational externalities that are inherent in the sequential social interaction. By searching and deciding, the agents unintentionally provide useful information to the individuals deciding after them. The increase of the average net utility in the economy is observed for all diversity of preferences conditions, but it is strongest when the cost of search is high (or the satisficing threshold is low). The reason behind this is clear: With high search cost, the agents search relatively shallowly, leaving a greater potential for further improvement by reordering the alternatives in the *search path*.

4 Discussion

Sixto Diaz Rodriguez, a talented folk/rock musician from Detroit who many thought destined for great success, released two promising albums at the beginning of the 1970s. To the disappointment of his producer, his albums sold only a few copies in the United States and most other countries in the world. However, Rodriguez's music was an absolute "hit" in South Africa, Botswana, Australia, and New Zealand. Rodriguez himself was completely unaware of his superstar status in these countries until he was invited to tour in Australia in the late 1970s and in South Africa only in the 1990s.

In the absence of counterfactual evidence, like the different regional markets in the story of Rodriguez, we could quickly conclude that differences in quality are enough to explain the success or failure in a market. The Music Lab experiments and recent models of social influence have demonstrated that we have to take into account the social influence and sampling processes that lead to rich-get-richer dynamics in a market. With these effects, exactly the same initial conditions can lead to different end-market outcomes where only few high-quality alternatives will turn into hits. Yet, several questions remain unresolved in regard to the mechanisms that generate the rich-get-richer dynamics. Without a process model of the cognitive and social influence processes, modeling results cannot generalize to new conditions of interest.

In this paper, we modeled the steps in the decision-making process followed by consumers when deciding which alternative to choose. Consumers followed a sequential search process (Simon, 1955a; DeGroot, 1970), a model with extensive applications in economics and marketing (Moorthy, Ratchford, & Talukdar, 1997). We introduced social information into the model through a popularity heuristic, whereby individuals examined alternatives according to aggregate popularity information, a signal used widely in real-life cultural markets — number of citations and book bestseller lists being prime examples. We then used agent-based simulations to scale up from individual agents to the market outcomes. This method has the advantage that counterfactual worlds can be created very cheaply. This way, our simulations can generate predictions for environments with conditions different from the data at hand. We simulated the collective dynamics in the market as a function of two crucial variables — the diversity of taste in the environment and the satisficing threshold employed by the agents. By comparisons to a random search market and a perfect information market our model enables us to measure the exact impact of social influence on the patterns of collective behavior.

4.1 Luck and dynamics of success in cultural markets

Stochastic process models suggest that the success of an alternative in the market is a matter of pure luck or merely early arrival in the market (Price, 1976). Differences in quality are completely inconsequential with respect to success. In contrast, in models in which agents have full knowledge about utilities, the eventual market shares of the alternatives are completely determined by utilities (Rosen, 1981). Salganik et al.'s (2006) experiments indicated that both quality and luck play a role.

Our model pins down an exact cumulative advantage mechanism by which quality (objective utility) and luck play out. Success in the market depends on both the objective utility of an alternative and the decisions of a few influential decision makers in the market. In our model, conquering one of the first positions in the search order confers a significant advantage. Alternatives with a high objective utility component stand a better chance of reaching these positions, which through social interactions may attract many more consumers over time. This setting led us to some novel insights on the inner workings of the market. For example, in our model, the decision makers unavoidably remain unaware of the existence of many high-quality alternatives the alternatives that will be pulled to the first position of the popularity list are only a subset of all the high-quality alternatives and it is probable that the alternative with the highest objective utility is not among them. This provides a rationale for firms competing to place their product along the consumers *search paths* (e.g., Bagwell & Ramey, 1994; Armstrong et al., 2009). A more important benefit of using this setting was that it allowed us to analyze welfare outcomes for the market, which we discuss in details in the following section.

4.2 Exploration, imitation and diversity

Search models exemplify the trade-off between exploiting the best alternative found thus far and further exploring with the hope of identifying better ones (March, 1991). Exploration is based on trial and error and entails a high cost for individual decision makers. Thus, in sequential decision-making environments self-interested decision makers are often better off imitating the choices of individuals who have already incurred the costs of exploration. As an example, consider our model in the conditions with no diversity of preference. Although imitation makes sense at the individual level, it deprives the group of additional information that could have been collected by individual explorers (Rogers, 1988). This is why in eusocial species such as bees and ants, in which individuals have evolved to maximize the fitness of the entire group, the members of the group occasionally explore new alternatives for hive locations or better paths to reach food sources. Information cascade and herding models developed by economists portray the shortcomings of imitation and social influence and stress the potential losses of collective welfare (e.g., Bikhchandani et al., 1992). In contrast, social learning studies tend to focus on the effectiveness of imitation strategies at the individual level (e.g., Laland, 2004; Richerson & Boyd, 2008; Rendell et al., 2010). Depending on the quality of feedback, and the pay-off structure in the environment, imitation can lead to either passable or very good results at the aggregate level.

Even in environments where individuals are self-interested there are mechanisms that can keep imitation forces in check and reap some of the benefits of the independent collection of information. For example, under different sets of conditions a natural equilibrium evolves between explorers (or producers) and imitators (or scroungers Conlisk, 1980; Kameda & Nakanishi, 2002; Rogers, 1988). Further, in some search environments, barriers to communication in the form of a sparser communication network among agents turn out to be beneficial at the collective level. They encourage individuals to explore more, in this way supplying useful information to the group (Fang, Lee, &

Schilling, 2010; Lazer & Friedman, 2007; Mason, Jones, & Goldstone, 2008). We believe that the interplay between diversity and costly sequential search provides another such example. Because preferences differ and agents can directly experience the utility of alternatives for them, they sample alternatives until they encounter a satisficing one, rather than unconditionally imitating the choices of individuals deciding before them. The popularity information influences the order of search, but crucially agents still sample some alternatives, which ensures that some new information is stored in their choices. Over time this behavior improves the *search path* and increases the welfare of the collective (see Figure 5). Note, however, that too much diversity makes the signals less informative since preferences become completely uncorrelated, which then leads to a decrease in the aggregate welfare.

Similar counter-intuitive beneficial effects of diverse preferences have been shown in an earlier sequential model of social learning (Goeree, Palfrey, & Rogers, 2006) and a recent publication on the dynamics of the academic reviewing process (Park, Peacey, & Munafò, 2014). In line with our account, in these models diversity incentivizes agents to use individual information, in this way providing useful information to the group. In all three examples diversity can be seen as a beneficial hurdle to communication that helps the agents internalize the strong informational externalities. This mechanism is distinct from alternative accounts of the beneficial role of diversity in groups that are based on complementarities in the problem-solving capacities of the agents in the group (Clearwater, Huberman, & Hogg, 1991; Hong & Page, 2004), or complementarities in the information possessed by diverse members of a group (Conradt, List, & Roper, 2013; Davis-Stober, Budescu, Dana, & Broomell, 2014).

Overall, in our study socially informed search led to a drastic improvement in the overall welfare as compared to random search when the agent preferences were correlated. As illustrated in Figure 1, random search generates much more useful information, but it does not allow an individual agent to save any costs. Thus, an agent would prefer to be in a market of random searchers but would choose using the popularity heuristic. In our model lower costs of search drastically changed the performance of the popularity heuristic, led to much better aggregate welfare outcomes, and lessened the extent to which additional diversity can be beneficial for the collective.

4.3 Revisiting the Music lab experiment

The Music Lab illustrated how history can be rerun in the lab (Salganik et al., 2006). The authors conducted the experiment with clearly stated hypotheses, but without specifying an individual decision-making model beforehand. They suggested that agent-based models could complement their research and generate predictions for other possible conditions of the experiment— a different number of alternatives, or participants or an evolving market (Salganik & Watts, 2009). As already discussed in the Introduction, two computational models have attempted to capture the decision-making processes of the participants in the experiment (Borghesi & Bouchaud, 2007; Krumme et al., 2012). Fitting the parameters of the models to the data, they have reproduced the findings of the experiment.[5] Yet it is questionable to what extent these models generalize to new environments. Our model generates clear-cut hypotheses about the direction of change of all the main measures of the Music Lab experiments. Thus, it could serve as a guide for future experimentation or the collection of field data. For example, our model would predict that when the number of alternatives increases, the inequality and unpredictability in the market should also increase. Similarly, it would predict more inequality and unpredictability in markets in which consumers have more homogeneous preferences. We believe that large-scale experiments in which participants have limited monetary or time budgets would further help test our theory and uncover the individual decision-making processes.

4.4 Managing social influence

Creators, producers, and other individual actors in the markets for cultural products are primarily interested in maximizing their own benefits from the market. In contrast, the designers of online cultural marketplaces (such as Amazon) or public organizations might be interested in maximizing the welfare of the entire consumer population. To this end, the designers may attempt to manage social influence by channeling information selectively to the agents in the decision-making sequence. An early example of large-scale social influence management is provided by the work of Wu and Huberman (2008), who examined the social network Twitter and studied the impact of different post-ordering algorithms on the total number of clicks. How should the social influence be structured if the goal is to maximize the sum of the utilities of the participants in the market? Our model reveals that consumers deciding early on in the sequence have a large impact on the average utility in the economy. In addition, consumers who act independently of each other supply more information to the consumers deciding after them. Hence, a possible approach would be to give to n agents in the beginning of the sequence the task of randomly exploring the alternatives. Then, the remaining agents would follow the *search path* shaped by those first agents. Such a tentative analysis could be easily conducted within the current framework or its extensions in the future.

4.5 Extensions and limitations

Hitherto, we have opted for a minimal and psychologically plausible process model that is still able to capture the stylized facts of the markets for cultural products. Clearly, in many actual markets the consumer decision-making processes are psychologically richer than in our account. For example, how would social influence play out if consumers had access to further information such as the average rating from other users in addition to the popularity information?[6] Similarly, how would awareness or recognition of some alternatives change an individuals search and choice strategy and as a result the collective behavior? Or, how would the collective patterns change

[5] Note that we could fix several parameters of the model, such as the number of alternatives selected by each agent, the diversity and the satisficing threshold in order to reproduce as closely as possible the aggregate dynamics.

[6] Characteristically, Netflix implements a ranking algorithm that combines popularity with other non-social information to order the alternatives for its users.

if consumers accessed information locally from their group of friends and acquaintances (Pachur, Rieskamp, & Hertwig, 2005) rather than globally from a general bestseller list. Our model provides a framework within which such psychological insights and their implications for aggregate welfare could be studied in the future. Extension of the model could take into account other sources of information, friendship networks or product awareness. Such extensions would also be crucial for further understanding and then managing social influence in real-world marketplaces.

4.6 Concluding remarks

To understand most real-world phenomena, like the initial failure and the eventual triumph of Sixto Diaz Rodriguez, it does not suffice to put forward as-if models that disregard the cognitive and social influence processes. We also need to understand the processes that determine individual behavior and lead to collective macroscopic patterns. The study of collective behavior has been gaining momentum over the last few decades. New methodological possibilities such as the accessibility of big data (e.g., Wu & Huberman, 2007; Bentley, O'Brien, & Brock, 2014), the possibility to conduct large-scale experiments over the Internet (e.g., Salganik et al., 2006), the rise of agent-based simulations (e.g., Schelling, 1978), or the coupling of experimentation with simulation methods (e.g., Moussaïd, Kämmer, Analytis, & Neth, 2013) have contributed to its development (also see Goldstone & Janssen, 2005; Goldstone & Gureckis, 2009).

In this paper we reported on our development of a parsimonious process model of collective behavior based on a well-established model of individual decision making that can capture the empirical observations from the markets for cultural products better than in previous accounts. We studied the rich-get-richer dynamics and the average welfare as a function of the diversity of taste in the consumer population and the satisficing threshold employed by the decision makers. Finally we quantified the impact of social influence by comparing this environment to environments without social interaction. Our model can inform the decisions of a host of stakeholders participating in the markets: producers, for instance, betting on which music albums will succeed and figuring out how to distribute the risk among them, or marketers planning to launch a campaign for a product that has not been selling well. It could be a tool in the hands of policy makers who need to design and justify scholarship schemes to support promising yet unknown artists. Moreover, it provides insight into how the markets could be designed to improve the average welfare of the consumers acting in them. Last but not least, the model can be used to generate new hypotheses for future large-scale studies such as the Music Lab experiment, and it provides a framework within which additional psychological processes that determine the market dynamics might be investigated in the future.

References

Adler, M. (1985). Stardom and talent. *The American Economic Review*, *75*(1), 208–212.

Analytis, P. P., Barkoczi, D., & Herzog, S. M. (2015). You' re special, but it doesn't matter if you're a greenhorn: Social recommender strategies for mere mortals. In *Proceedings of the 37th annual conference of the cognitive science society.*

Analytis, P. P., Kothiyal, A., & Katsikopoulos, K. (2014). Multi-attribute utility models as cognitive search engines. *Judgment and Decision Making*, *9*(5), 403–419.

Armstrong, M., Vickers, J., & Zhou, J. (2009). Prominence and consumer search. *The RAND Journal of Economics*, *40*(2), 209–233.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, *99*(394), 116–131.

Bagwell, K., & Ramey, G. (1994). Coordination economies, advertising, and search behavior in retail markets. *The American Economic Review*, *84*(3), 498–517.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, *107*(3), 797–817.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bentley, R. A., O'Brien, M. J., & Brock, W. A. (2014). Mapping collective behavior in the big-data era. *Behavioral and Brain Sciences*, *37*(01), 63–76.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Borghesi, C., & Bouchaud, J.-P. (2007). Of songs and men: a model for multiple choice with herding. *Quality & Quantity*, *41*(4), 557–568.

Brock, W. A., & Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, *68*(2), 235–260.

Chow, Y. S., Robbins, H., & Siegmund, D. (1971). *Great expectations: The theory of optimal stopping.* Boston, MA: Houghton Mifflin Boston.

Chung, K. H., & Cox, R. A. (1994). A stochastic model of superstardom: An application of the Yule distribution. *The Review of Economics and Statistics*, *76*(4), 771–775.

Clearwater, S. H., Huberman, B. A., & Hogg, T. (1991). Cooperative solution of constraint satisfaction problems. *Science*, *254*, 1181–1183.

Conlisk, J. (1980). Costly optimizers versus cheap imitators. *Journal of Economic Behavior & Organization*, *1*(3), 275–293.

Conlisk, J. (2003). A note on convergence of adaptive satisficing to optimal stopping. *Journal of Political Economy*, *111*(6), 1353–1360.

Conradt, L., List, C., & Roper, T. J. (2013). Swarm intelligence: When uncertainty meets conflict. *The American Naturalist*, *182*(5), 592–610.

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, *1*(2), 97-101.

DeGroot, M. (1970). *Optimal statistical decisions.* New York, NU: McGraw-Hill.

De Vany, A. (2004). *Hollywood economics: How extreme uncertainty shapes the film industry.* London, England:Routledge.

Elster, J. (1989). *Nuts and bolts for the social sciences.* Cambridge England: Cambridge Univ Press.

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organization Science*, *21*(3), 625–642.

Goel, S., Reeves, D. M., Watts, D. J., & Pennock, D. M. (2010). Prediction without markets. In *Proceedings of the 11th acm conference on electronic commerce* (pp. 357–366).

Goeree, J. K., Palfrey, T. R., & Rogers, B. W. (2006). Social learning with private and common values. *Economic Theory*, *28*(2), 245–264.

Goldstone, R. L., & Gureckis, T. M. (2009). Collective behavior. *Topics in Cognitive Science*, *1*(3), 412–438.

Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, *9*(9), 424–430.

Granovetter, M., & Soong, R. (1986). Threshold models of interpersonal effects in consumer demand. *Journal of Economic Behavior & Organization*, *7*(1), 83–99.

Hendricks, K., & Sorensen, A. (2009). Information and the skewness of music sales. *Journal of Political Economy*, *117*(2), 324–369.

Hendricks, K., Sorensen, A., & Wiseman, T. (2012). Observational learning and demand for search goods. *American Economic Journal: Microeconomics*, *4*(1), 1–31.

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1191-1206.

Hey, J. D. (1987). Still searching. *Journal of Economic Behavior & Organization*, *8*(1), 137–144.

Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less more? *Journal of Mathematical Psychology*, *49*(2), 115–124.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(46), 16385–16389.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Kameda, T., & Nakanishi, D. (2002). Cost–benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evolution and Human Behavior*, *23*(5), 373–393.

Krumme, C., Cebrian, M., Pickard, G., & Pentland, S. (2012). Quantifying social influence in an online cultural market. *PLOS ONE*, *7*(5), e33785.

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667–694.

Leibenstein, H. (1950). Bandwagon, snob, and Veblen effects in the theory of consumers' demand. *The Quarterly Journal of Economics*, 183–207.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

Mason, W. A., Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, *137*(3), 422-433.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, *84*(1), 113–126.

Merton, R. K. (1968). The matthew effect in science. *Science*, *159*(3810), 56–63.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, *1*(2), 226–251.

Moorthy, S., Ratchford, B. T., & Talukdar, D. (1997). Consumer information search revisited: Theory and empirical analysis. *Journal of Consumer Research*, *23*(4), 263–277.

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLOS ONE*, *8*(11), e78433.

Newman, M. E. (2005). Power laws, pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.

Pachur, T., Rieskamp, J., & Hertwig, R. (2005). The social circle heuristic: Fast and frugal decisions based on small samples. In *Proceedings of the 26th annual conference of the cognitive science society. mahwah, nj: Erlbaum* (pp. 1077–1082).

Park, I.-U., Peacey, M. W., & Munafò, M. R. (2014). Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*, *506*(7486), 93–96.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306.

Rapoport, A., & Tversky, A. (1970). Choice behavior in an optional stopping task. *Organizational Behavior and Human Performance*, *5*(2), 105–120.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., ... Laland, K. N. (2010). Why copy others? insights from the social learning strategies tournament. *Science*, *328*(5975), 208–213.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, 900–926.

Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution.* Chicago, IL: University of Chicago Press.

Rogers, A. R. (1988). Does biology constrain culture? *American Anthropologist*, *90*(4), 819–831.

Rosen, S. (1981). The economics of superstars. *The American economic review*, *71*(5), 845–858.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*(5762), 854–856.

Salganik, M. J., & Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*, *71*(4), 338–355.

Salganik, M. J., & Watts, D. J. (2009). Web-based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science*, *1*(3), 439–468.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. WW Norton.

Simon, H. A. (1955a). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Simon, H. A. (1955b). On a class of skew distribution functions. *Biometrika*, 425–440.

Telser, L. G. (1973). Searching for the lowest price. *American Economic Review*, *63*(2), 40–49.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, *393*(6684), 440–442.

Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences*, *104*(45), 17599–17601.

Wu, F., & Huberman, B. A. (2008). Popularity, novelty and attention. In *Proceedings of the 9th acm conference on electronic commerce* (pp. 240–245).

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, F. R. S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *213*(402-410), 21–87.