CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 694

Identification in Search Models with Social Information

Niccolò Lomys and Emanuele Tarantino

November 2023

University of Naples Federico II | **University of Salerno** | **Bocconi University, Milan**

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 694

Identification in Search Models with Social Information

Niccolò Lomys* **and Emanuele Tarantino†**

Abstract
We theoretically study the problem of a researcher seeking to identify and estimate the search cost distribution when a share of agents in the population observes some peers' choices. To begin with, we show that social information changes agents' optimal search and, as a result, the distributions of observable outcomes identifying the search model. Consequently, neglecting social information leads to non-identification of the search cost distribution. Whether, as a result, search frictions are under or overestimated depends on the dataset's content. Next, we present empirical strategies that restore identification and correct estimation. First, we show how to recover robust bounds on the search cost distribution by imposing only minimal assumptions on agents' social information. Second, we explore how leveraging additional data or stronger assumptions can help obtain more informative estimates.

JEL Classification: C1; C5; C8; D1; D6; D8.

Keywords: Search & Learning; Social Information; Identification; Networks; Robustness; Partial Identification.

Acknowledgments: We thank various seminars, workshops, and conference audiences, and, in particular, Michele Bisceglia, Federico Carlini, Daniel Ershov, Matteo Escudé, Satoshi Fukuda, Daniel Garcia, Cristina Gualdani, Ole Jann, Philipp Kircher, Lorenzo Magnolfi, José Luis Moraga-Gonzàlez, Philipp Peitler, Jacopo Perego, and Yutec Sun for their helpful comments. We thank Xingwei Gan for his excellent research assistance. This work was done in part while Niccolò Lomys was visiting the Simons Institute for the Theory of Computing at UC Berkeley (program on Data-Driven Decision Processes). Niccolò Lomys acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreements N° 714147 and N° 714693).

* CSEF and Università degli Studi di Napoli Federico II. Email: niccolomys@gmail.com.

† Luiss University, EIEF, and CEPR. Email: etarantino@luiss.it

1 Introduction

Agents seldom search in isolation: social information—i.e., the choices and experiences of other agents—is readily available via direct observation, communication, and social networks. An extensive theoretical and empirical literature on social learning (see, e.g., Mobius and Rosenblat, 2014; Golub and Sadler, 2016; Bikhchandani, Hirshleifer, Tamuz, and Welch, 2022, for comprehensive reviews) documents how heavily agents rely on the information of others in shaping their beliefs and behavior. For example, Bailey, Johnston, Kuchler, Stroebel, and Wong (2022) show that a new phone purchase by a friend on Facebook significantly affects an individual's demand for a phone of the same brand. Cai, Chen, and Fang (2009) find that restaurant patrons are more likely to order the goods presented to them as the most popular. Similarly, for the movie market, Moretti (2011) finds that an unexpected increase in box office revenues during the first week has a persistent and significant effect on future attendance. More broadly, these findings are consistent with survey evidence documenting that referrals on social media influence purchase behavior (Forbes, 2012, 2022).

Despite the wealth of evidence highlighting how strongly social information influences individual behavior, virtually all empirical search models assume agents search in isolation, ignoring the information in their peers' decisions (see, e.g., Honka, Hortaçsu, and Wildenbeest, 2019; Ursu, Seiler, and Honka, 2023, for recent accounts of the empirical search literature). However, if social information changes the distributions that generate observable individual outcomes identifying the search model, neglecting its presence may result in the non-identification of search cost distributions and, hence, their inconsistent estimation.

Understanding whether, and if so, under which conditions, neglecting social information leads to the incorrect quantification of search frictions is a central question for policy analysis; so is exploring the scope of potential remedies that help fix such misguided conclusions. The reason is that search costs are a primary determinant of agents' choice, pricing behavior, and market outcomes. Hence, their correct quantification is a fundamental input for reliable empirical investigations, such as computing price elasticities, assessing market competitiveness, and performing counterfactual analyses in regulated markets.

In this paper, we theoretically investigate these questions by considering a researcher who has access to standard datasets on observable individual outcomes related to the solution of a search problem, e.g., searching for a product or information about it. A share of individuals in the population that generates the dataset have social information—i.e., observe the choices of some peers—before engaging in their search. Our goal is to determine the assumptions the researcher can or should make about the presence of social information to identify and estimate (features of) the search cost distribution. First, we show that neglecting social information, even in its simplest form, leads to non-identification and inconsistent estimation of the search cost distribution. Next, we present empirical strategies that allow the researcher to restore identification and obtain consistent estimates.

We illustrate the logic for the non-identification and biased estimation results with two datasets commonly available to empirical researchers.

Example 1: Data on Choice. Agents search online retailers for the best price to purchase a specific mobile phone. The researcher observes the price at which each agent buys the phone at various retailers. Some agents pay a low price, whereas others pay a high price for the same phone. Before searching, some agents observe the retailer where some peers have purchased the phone.

If the researcher assumes agents search in isolation, she infers low search costs for all agents paying a low price. However, some agents observe the retailer where some of their peers have purchased. Since the peers have possibly searched other retailers, they infer that that particular retailer is likely offering a good bargain. Therefore, they buy at a low price because they exploit social information and not because they have a low search cost. Formally, the probability that agents in the population buy at a low price is greater with social information than without. As a result, by neglecting social information, the search cost distribution is not identified, and the researcher underestimates search costs.

Example 2: Data on the Number of Searches. Agents search for restaurants online. The researcher observes the number of searches each agent conducts before choosing a restaurant. Some agents search multiple restaurants, whereas others search only once. Before searching, some agents observe online referrals of a specific restaurant by some peers.

If the researcher assumes agents search in isolation, she infers high search costs for all agents searching only once. However, some agents observe their peers' referrals. Since the peers have possibly searched for multiple restaurants before posting on social media, they infer that that specific restaurant is likely offering high-quality meals. Therefore, they search only once because they exploit their social information and not because they have a high search cost. Formally, the probability that agents in the population search only once is greater with social information than without. As a result, by neglecting social information, the search cost distribution is not identified, and the researcher overestimates search costs.

The two examples show that, under the standard assumption that all agents are isolated, failure to account for social information leads to non-identification of the search cost distribution. Whether, as a result, search costs are under or overestimated depends on the dataset's content (data on choice as opposed to data on the number of searches), thus highlighting the non-trivial role that social information can play in empirical search models. However, we show that the researcher can restore identification and consistent estimation by taking two alternative approaches. First, the researcher can recover robust bounds on the search cost distribution by imposing minimal assumptions on agents' amount and type of social information. Second, the researcher can recover more informative estimates with access to additional data or by leveraging stronger assumptions on the environment.

We formalize these insights with a stylized version of Weitzman (1979)'s sequential search model. Each agent in the population must choose between two alternatives whose utilities, high or low for simplicity, are i.i.d. draws. Before searching, the agent knows the utility distribution but not the realized utilities. Searching for an alternative reveals its utility to the agent. After the first free search, the agent decides whether to search for the second alternative at a cost drawn from some distribution. In this baseline setting, all agents act in isolation, in line with the standard approach of the empirical search literature.

To model social information, we develop two polar generalizations of the baseline search model. First, we consider an environment—the *simple social information* setting—in which, before searching, each agent in the population observes with some probability the choice of one of her peers, but neither the peer's search behavior nor her search cost. The peer is isolated and faces the same utility realizations as the agent. Thus, the agent draws some imperfect inferences about realized utilities from the choice of her peer. We rely on this parsimonious setting to transparently illustrate how social information changes the researcher's problem of identifying and estimating search cost distributions. Second, we consider an environment—the *general social information* setting—in which the researcher imposes only minimal assumptions on the structure of social information. We rely on this setting to show that the researcher can recover robust bounds on search cost distributions while remaining agnostic on the amount and type of social information agents have access to.

To begin with, we show that the optimal search of an agent with simple social information differs from that of an isolated agent in two ways. First, whereas the latter is indifferent about which alternative to search first, the former is not. In the eyes of an agent with simple social information, the utility distribution of the alternative chosen by her peer first-order stochastically dominates that of the other because the peer already searches for both with positive probability. Thus, an agent with simple social information searches for the alternative chosen by her peer first. Hence, she samples a high-utility alternative at the first search with the same probability an isolated agent chooses such an alternative after stopping her search. Second, the expected gain from the second search for an agent with social information is lower than that of an isolated agent. The reason is that the second search is valuable only if the peer searches only once. Thus, social information reduces the incentive to search.

These two differences imply that agents with simple social information choose a high-utility alternative and search only once with greater probability than isolated agents. Hence, social information alters the distributions of the observable outcomes that identify the search model. Our first main result follows: ignoring social information, even in its simplest form, leads to non-identification and biased estimation of search cost distributions.

We illustrate these results within the simple social information setting. Although admittedly simplistic, this setting allows us to transparently identify the driving forces behind the insight—i.e., that agents with social information are more likely to choose a high-utility alternative and conduct fewer searches than isolated agents. These forces, however, would emerge in any bandit model in which agents trade off the exploitation of social information with independent exploration (namely, search) for the best alternative. Therefore, although deriving the same results in other or more general specifications of the search model may become algebraically more complicated (hence, less transparent), and the precise quantitative effect may differ from one specification to another, the conceptual insights would remain unchanged. We illustrate these considerations by discussing extensions of our analysis to various salient generalizations of the setting with simple social information.

In the second part of the paper, we present empirical strategies that help restore identification and correct estimation of the search cost distribution. These remedies form our second main set of results. We distinguish between two approaches.

Under the first approach, which relies on the general social information setting, we show how the researcher can construct robust and informative bounds on the search cost distribution when the amount and type of agents' social information remain unobserved. This approach is motivated by the observation that real social networks, communication channels, and informational externalities among agents are complex. As a result, it may be hard to make reliable assumptions about the specific structure of agents' social information.

To construct robust bounds, the researcher must consider the implications of two opposite assumptions on the observable outcomes that identify the search model: each agent acts in isolation; social information is so abundant that it suffices to each agent to choose the alternative with the highest realized utility at the first search. No matter the exact amount of agents' social information or where it comes from, it will always be between these extremes. Hence, the probability that agents choose an alternative with a given utility level is always bounded (from above and below) by those implied by these two opposite assumptions on the data-generating process, and so is the probability that agents conduct a single search.

This robust partial-identification approach is valuable for three reasons. First, it allows the researcher to remain agnostic about many features of the environment about which it may be hard to formulate credible assumptions. In particular, the researcher need not specify how many peers agents interact with and whether social information comes from observational or communication learning, advertising, ratings, or reviews, among other assumptions. Second, the method's empirical implementation is intuitive and straightforward. To construct bounds on the search cost distribution under general social information, the researcher only needs to estimate the search model twice, once with data on the number of searches and once with data on choice, under the standard assumption in the empirical search literature that all agents act in isolation. The resulting estimates correspond to the minimal and the maximal values that the features of the search cost distribution that are of interest can take. Third, besides providing a robustness benchmark, how far apart the bounds derived with this approach are is also a measure of how misguided conclusions can be when neglecting social information.

In the second approach, we outline three empirical strategies that allow the researcher to obtain more informative estimates of the search cost distribution. These strategies, however, require the researcher to have access to additional data or be willing to impose stronger assumptions on the empirical environment.

First, the researcher can point-identify the search cost distribution if she has access to agent-level data distinguishing isolated agents from agents with simple social information. Such data, however, are hardly available. Second, the researcher can partially identify the search cost distribution by estimating offline the share of agents with simple social information. For instance, the researcher can acquire detailed network data on the agents with access to social media or survey evidence on agents' reliance on the choices of their peers for purchase decisions. Third, the researcher can partially identify the search cost distribution with no information on the share of agents with simple social information or jointly identify both of such primitives.

We explore how the researcher can leverage additional data or stronger assumptions to obtain more informative estimates of the search cost distribution in the simple social information setting. However, our analysis provides general insights into the data requirements or the assumptions that help identify search cost distributions in other specifications of the search model one may develop to tailor specific empirical applications.

Related Literature. Recovering economic primitives from observable outcomes has a longstanding tradition in Economics (see, e.g., the revealed preference literature, Chambers and Echenique, 2016). Our approach draws inspiration from recent theoretical work on identification, such as: Heidhues and Strack (2021) for the identification of present bias from the timing of choices; Bergemann, Brooks, and Morris (2022) for counterfactual predictions with latent information; Liu and Netzer (2023) for the identification of happiness measures from ordered response data; Libgober (2023) for the identification of information structures from posterior beliefs; Kang and Vasserman (2022) for the construction of bounds on welfare estimates that are robust to functional form assumptions on consumer demand; Shmaya and Yariv (2016); Deb and Renou (2021); De Oliveira and Lamba (2023) for the testable implications of learning on observed choices; Heumann (2019) for informationally robust comparative statics. None of these papers, however, considers the identification of search models nor studies the role of social information.

Our approach to general social information is close in spirit to the theoretical literature on robust predictions in incomplete information or extensive form games (Bergemann and Morris, 2016; Doval and Ely, 2020) and the econometric literature on partial identification (see, e.g., Manski, 2003; Molinari, 2020; Kline and Tamer, 2023, for comprehensive surveys). Like other papers in this literature, our analysis of robust bounds on the search cost distribution under relatively weak restrictions aims to provide transparency for the mapping between modeling assumptions and subsequent econometric conclusions. For instance, recent work uses the Bayes correlated equilibrium notion of Bergemann and Morris (2016) to develop informationally robust identification and estimation strategies. Examples are: Magnolfi and Roncoroni (2022) for entry games; Syrgkanis, Tamer, and Ziani (2021) for auctions; Gualdani and Sinha (2023) for single-agent models of voting; Canen and Song (2023) for counterfactual analyses.

Recent work analyzes how social learning affects individual search behavior (see, e.g., Kircher and Postlewaite, 2008; Galeotti, 2010; Hendricks, Sorensen, and Wiseman, 2012; Mueller-Frank and Pai, 2016; Garcia and Shelegia, 2018; Lomys, 2023). We use these results to motivate the importance of social information in shaping the search process. However, our goals are distinct, as none of these papers studies econometric identification problems. More broadly, by making a first step into understanding how social information can change the identification of search models, our theoretical analysis can provide the well-established empirical search literature (see, e.g., Honka et al., 2019; Ursu et al., 2023, for comprehensive reviews) with insights into how to account for social information in specific applications.

2 Setup

In this section, we present the setup. First, we describe hypothetical datasets on observable individual outcomes related to the solution of search problems, e.g., searching for a product or information about it. Such datasets are routine in the empirical search literature (see, e.g., Honka et al., 2019). Next, we consider alternative assumptions about the data-generating process. Most of these assumptions are standard in the empirical search literature. Our focus, instead, is on the assumptions about the presence of social information. The goal is to characterize the conditions that allow a researcher to identify and estimate the search cost distribution as a function of such assumptions.

2.1 Data

Consider the following canonical search environment. In each of countably many search problems, indexed by $n \in \mathbb{N}$, agent n must select an alternative from a finite set X. Let u_n^x denote agent n's (indirect) utility from an alternative $x \in X$. Utilities are drawn from some distribution with finite support $U \subseteq \mathbb{R}_+$. Agent n knows the utility distribution but not the realized utilities, about which she collects information via a costly search.

Examples. To fix ideas, consider the following examples.

- Suppose $u_n^x := \hat{u} - p_n^x$, where $\hat{u} > 0$ is a constant valuation for all alternatives common to all agents, and p_n^x is the price paid by agent n for alternative x. This specification corresponds to a price search model for homogeneous goods with identical agents and ex-ante identical firms.

- Suppose $u_n^x := \varepsilon_n^x$, where ε_n^x captures the idiosyncratic valuation of agent n for alternative x, i.e., how well alternative x fits, or matches, agent n's needs. This specification corresponds to a match-value search model with ex-ante identical agents and firms.

By suitably specifying the general model to capture the features of interest, one can similarly accommodate other classes of search models. ●

A researcher observes either or both of the following standard datasets.

- *Data on Choice.* For each possible utility $u \in U$, the researcher observes the empirical distribution of agents who choose an alternative with utility u. Data on choice are readily available for price search models: the researcher needs to observe the shares of transactions occurring at different prices. Data on choice, instead, may not always be available for match-value search models, as match values may not be observable.

- *Data on the Number of Searches.* Let $K := |X|$. The researcher observes the empirical distribution of agents who conduct k searches for all $k \in \{1, \dots, K\}$. Data on the number of searches are readily available for price and match-value search models. Information about prices or match values is not necessary.

Given the available dataset, the researcher wants to identify and estimate the search cost distribution. To do so, the researcher makes assumptions about the search environment and agents' behavior to solve their search problem. These assumptions form the data-generating process. Based on these assumptions and the observable model implications that result, the researcher determines the conditions that identify the search cost distribution.

2.2 Data-Generating Process

We first lay down some basic assumptions in a canonical empirical search model. For simplicity, suppose $X := \{0, 1\}$ and $U := \{\underline{u}, \overline{u}\}$, where $\underline{u} < \overline{u}$. Suppose utilities are i.i.d. across alternatives within search problem n and across search problems. Let $\alpha := \mathbb{P}(u_n^x = \overline{u}) \in (0, 1)$, $\Delta u := \overline{u} - \underline{u}$, and $\neg x$ denote the alternative in X other than x. Suppose the researcher knows or can consistently estimate the utility distribution.

Each agent n is Bayesian and collects information via costly sequential search with recall à la Weitzman (1979). First, agent n decides which alternative to search first, $s_n^1 \in X$. By searching alternative s_n^1, agent n perfectly learns its realized utility $u_n^{s_n^1}$. Next, agent n decides whether to search for the remaining alternative, $s_n^2 = \neg s_n^1$, and perfectly learn its realized utility $u_n^{\neg s_n^1}$, or to discontinue the search, $s_n^2 = d$. Finally, agent n chooses an alternative a_n from the set S_n of alternatives she has searched.

The first search is free. The second search costs c_n, known to agent n.[1] Search costs—i.i.d. across agents—are drawn from a probability distribution F with full support

[1] All results remain unchanged if all searches are costly, and each agent must take an alternative so that she must search at least once.

on $[0, \bar{c}]$. We assume $\alpha \Delta u < \bar{c}$. Absent this assumption, the search problem is trivial: an agent would always search for both alternatives, irrespective of her search cost.

Agent n maximizes the difference between the utility of the chosen alternative and the incurred search cost: $u_n^{a_n} - c_n(|S_n| - 1)$.

Isolated Agents and Social Information. The empirical search literature typically assumes that all agents act in isolation, i.e., they solve the search problem as described above. We contrast identification and estimation under such an assumption to those when agents have social information, i.e., they observe the choices of some other agents before engaging in their search. In particular, we consider two polar cases for social information.

Simple Social Information. To model the simplest form of social information, we consider an agent that observes the choice of another agent in her social network with the same preferences (or utility draw). Formally, let $\theta_n \in \{\mathrm{I}, \mathrm{S}\}$ be the type of search problem n. Types θ_n are i.i.d. across search problems. Let $\gamma := \mathbb{P}(\theta_n = \mathrm{I}) \in (0, 1)$.

- If $\theta_n = \mathrm{I}$, agent n is isolated. Her search problem is as described above.
- If $\theta_n = \mathrm{S}$, agent n has simple social information. Before searching sequentially as described above, agent n observes the alternative a_{n_0} chosen by a fictitious Bayesian agent n_0—a different n_0 for each n with $\theta_n = \mathrm{S}$—who: (i) is isolated, $\mathbb{P}(\theta_{n_0} = \mathrm{I}) = 1$; (ii) has the same realized utilities as agent n, $(u_{n_0}^0, u_{n_0}^1) = (u_n^0, u_n^1)$; (iii) has idiosyncratic search cost c_{n_0} drawn independently of c_n from the same distribution. Agent n, however, observes neither agent n_0's search cost nor n_0's search decisions.[2]

General Social Information. To formalize the weakest assumptions on social information, we bound its amounts at the agent level between two extremes. Formally, let θ_n denote the "amount" of agent n's social information. Amounts θ_n are i.i.d. across search problems and bounded by the minimal and the maximal social information. Agent n's social information is minimal, denoted by $\theta_n = \underline{\theta}$, if agent n is isolated. Agent n's social information is maximal, denoted by $\theta_n = \overline{\theta}$, if agent n chooses the alternative with the highest realized utility at the first search just by exploiting her social information.

2.3 Identification and Estimation under Social Information

In what follows, we analyze how the identification and estimation problem of the researcher changes when agents have social information compared to when all agents are isolated.

In Section 3, we assume the Simple Social Information setting is the data-generating process. Two insights emerge from the analysis. First, under the standard assumption that all agents are isolated, failure to account for social information leads to non-identification

[2] All our insights remain qualitatively the same if agent n_0's realized utilities correlate with agent n's realized utilities but are not necessarily identical. This alternative assumption, however, would make the analysis algebraically more complicated and, hence, less transparent. See the discussion in Section 3.4.

of the search cost distribution. Second, whether, as a result, search costs are under or over-estimated depends on the dataset's content (choice as opposed to number of searches), thus highlighting the non-trivial role that social information can play in empirical search models.

In Section 4, we present approaches to restore identification and consistent estimation. First, under the robust General Social Information approach, the researcher can recover informative bounds on the search cost distribution while imposing minimal assumptions on the environment. Second, the researcher can exploit additional information or further structure on the environment to obtain more informative estimates.

3 Non-Identification with Simple Social Information

In this section, under the Simple Social Information data-generating process, we first characterize how optimal decisions differ between search problems of type S and type I. Second, we clarify how these differences modify the probability distributions generating observable individual outcomes related to the solution of agents' search problems. Third, we show how these changes result in the lack of identification and biased estimation of the search cost distribution if the researcher neglects social information. Finally, we discuss the robustness of our results to natural generalizations of the baseline model.

3.1 Optimal Decisions

3.1.1 Search Problem of Type I

Suppose $\theta_n = \mathrm{I}$, that is, agent n is isolated.

First Search Stage. Since the utilities of the two alternatives are i.i.d., agent n decides uniformly at random which alternative to search first: $s_n^1 = \frac{1}{2} \circ 0 + \frac{1}{2} \circ 1$, where $\sum_x \xi(x) \circ x$ denotes the mixture assigning probability $\xi(x)$ to alternative x. Breaking indifferences uniformly at random captures that labels do not convey information about alternatives' utilities or agents' behavior.[3]

Since utilities are i.i.d.,

$$u_n^{s_n^1} = \begin{cases} \overline{u} & \text{with probability } \alpha \\ \underline{u} & \text{with probability } 1-\alpha \end{cases} . \tag{1}$$

Second Search Stage. Agent n searches for the second alternative if and only if the

[3] All our insights remain qualitatively the same with different tie-breaking rules, but their quantification might change. Removing tie-breaking or equilibrium selection assumptions in structural econometric models is an interesting question and typically requires partial identification approaches. We note here that our analysis under General Social information in Section 4.1 does not require formulating any assumption on the tie-breaking criterion adopted by the agents in the population.

expected gain from doing so is greater than her search cost.[4] Given the utility of the first searched alternative, $u_n^{s_n^1}$, such a gain is

$$V_{\mathrm{I}}\left(u_n^{s_n^1}\right) := \mathbb{E}\left[\max\left\{u - u_n^{s_n^1}, 0\right\}\right] = \begin{cases} 0 & \text{if } u_n^{s_n^1} = \overline{u} \\ \alpha \Delta u & \text{if } u_n^{s_n^1} = \underline{u} \end{cases} . \tag{2}$$

Thus,

$$s_n^2 = \begin{cases} d & \text{if } u_n^{s_n^1} = \overline{u} \\ d & \text{if } u_n^{s_n^1} = \underline{u} \text{ and } c_n \geq V_{\mathrm{I}}(\underline{u}) \\ \neg s_n^1 & \text{if } u_n^{s_n^1} = \underline{u} \text{ and } c_n < V_{\mathrm{I}}(\underline{u}) \end{cases} .$$

In words, agent n discontinues the search if (i) the utility of the first searched alternative is high, or (ii) the utility of the first searched alternative is low, and her search cost is not smaller than the expected gain from the second search. Agent n searches for the remaining alternative otherwise.

Choice Stage. Agent n chooses the best alternative among those she searched, randomizing uniformly if indifferent:

$$a_n = \begin{cases} s_n^1 & \text{if } s_n^2 = d \\ \neg s_n^1 & \text{if } s_n^2 = \neg s_n^1 \text{ and } u_n^{\neg s_n^1} = \overline{u} \\ \frac{1}{2} \circ 0 + \frac{1}{2} \circ 1 & \text{if } s_n^2 = \neg s_n^1 \text{ and } u_n^{\neg s_n^1} = \underline{u} \end{cases} . \tag{3}$$

Decision Tree. Figure 1 summarizes agent n's decisions in a search problem of type I. Agent n decides uniformly at random which alternative to search first and observes the utility associated with such an alternative. Next, if her search cost is smaller than the expected gain from an additional search, agent n searches for the second alternative and observes the utility associated with such an alternative. Agent n discontinues the search otherwise. Finally, agent n chooses the best alternative among those she searched, randomizing uniformly if indifferent. At the end of each terminal node of the decision tree, we report the utility of the chosen alternative.

3.1.2 Search Problem of Type S

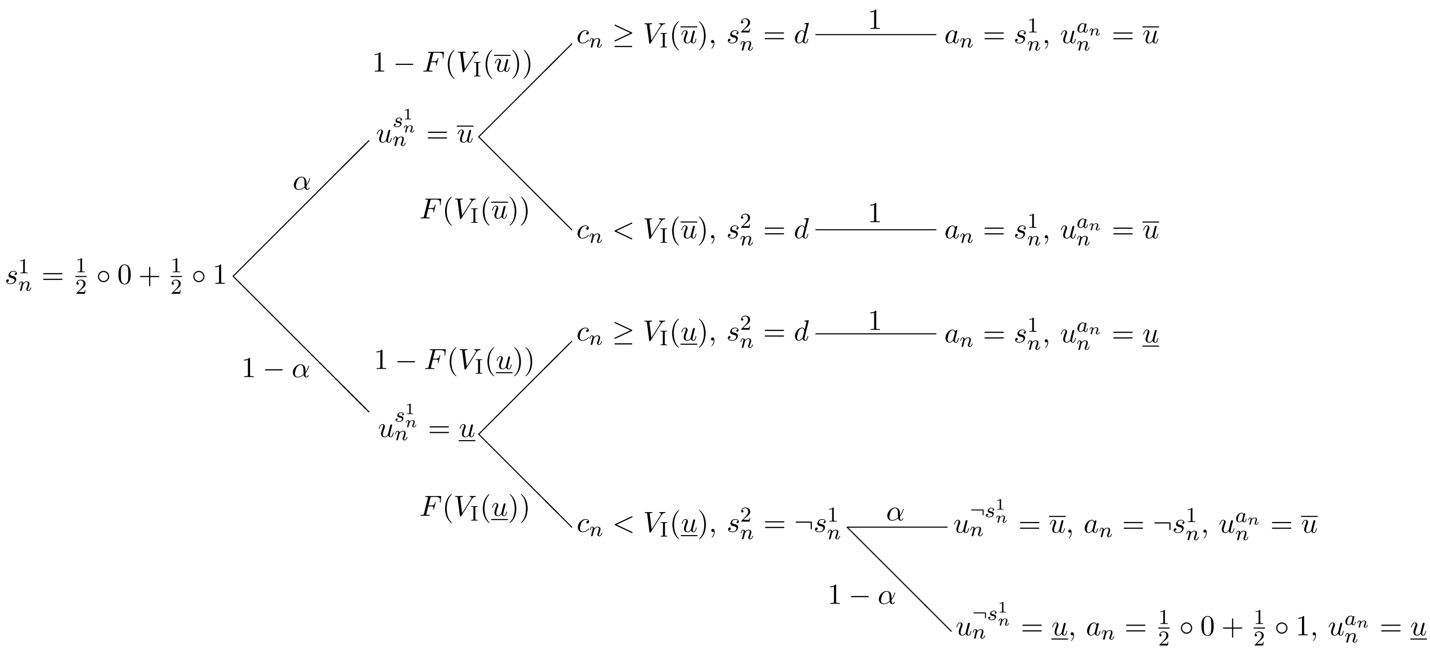
Suppose $\theta_n = \mathrm{S}$, that is, agent n has simple social information.

First Search Stage. Agent n's belief about the alternatives' utilities depends on agent n_0's choice in a search problem of type I. Two cases each occur with positive probability:

- Agent n_0 did not search for alternative $\neg a_{n_0}$. If so, agent n_0's choice is uninformative about the utility of alternative $\neg a_{n_0}$.

[4] Assuming that agents do not search for the second alternative in case of indifference is without loss of generality since this case is non-generic in the parameter space.

Figure 1: Decision Tree for a Search Problem of Type I.



- Agent n_0 searched for alternative $\neg a_{n_0}$. If so, it must be that $u_n^{a_{n_0}} \geq u_n^{\neg a_{n_0}}$ and, with positive probability, $u_n^{a_{n_0}} > u_n^{\neg a_{n_0}}$. That is, agent n_0's choice reveals alternative a_{n_0} to be superior to alternative $\neg a_{n_0}$, and strictly so with positive probability.

As a result, agent n's belief about the utility of alternative a_{n_0} strictly first-order stochastically dominates her belief about the utility of alternative $\neg a_{n_0}$. Hence, by Weitzman (1979)'s optimal search rule, as extended by Gergatsouli and Tzamos (2023) to correlated alternatives' utilities, agent n searches for alternative a_{n_0} first: $s_n^1 = a_{n_0}$.[5]

Therefore,

$$u_n^{s_n^1} = u_{n_0}^{a_{n_0}} = \begin{cases} \overline{u} & \text{with probability } \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})) \\ \underline{u} & \text{with probability } 1 - \alpha - \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})) \end{cases}, \tag{4}$$

where the probabilities are calculated from Figure 1. Specifically, simply by exploiting her social information—i.e., the observation of agent n_0's choice—an agent n with simple social information samples a high-utility alternative at the first search with greater probability than an isolated agent, such as agent n_0 (compare equations (1) and (4)). In particular, an agent with simple social information samples a high-utility alternative at the first search with the same probability that an isolated agent chooses a high-utility alternative at the end of her search. The latter probability is the sum of two probabilities: (i) α is the probability that agent n_0 samples a high-utility alternative at the first search; (ii) $\alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u}))$ is the probability that agent n_0 samples a low-utility alternative at the first search (i.e., $1-\alpha$), searches for the second alternative (i.e., $F(V_{\mathrm{I}}(\underline{u}))$), and such alternative has high

[5]The alternatives' utilities need no longer be independent in the eyes of an agent with social information.

utility (i.e., α). This is the *first difference* between search problems of types I and S.

Second Search Stage. Agent n searches for the second alternative if and only if the expected gain from doing so is greater than her search cost. Such a gain depends on the probability that agent n_0 did not search for alternative $\neg s_n^1$ given that an alternative with utility $u_n^{s_n^1}$ was chosen, denoted by $P\Big(u_n^{s_n^1}\Big)$. With remaining probability, agent n_0 searched for alternative $\neg s_n^1$ but chose alternative s_n^1, in which case alternative s_n^1 is non-inferior by revealed preference. Thus, agent n's expected gain from the second search is

$$V_{\mathrm{S}}\Big(u_n^{s_n^1}\Big) := P\Big(u_n^{s_n^1}\Big)\mathbb{E}\Big[\max\{u - u_n^{s_n^1}, 0\}\Big] = P\Big(u_n^{s_n^1}\Big)V_{\mathrm{I}}\Big(u_n^{s_n^1}\Big).$$

From Bayes rule and Figure 1,

$$P(\underline{u}) := \mathbb{P}\Big(s_{n_0}^2 = d \mid u_{n_0}^{a_{n_0}} = \underline{u}\Big) = \frac{1 - F(V_{\mathrm{I}}(\underline{u}))}{1 - F(V_{\mathrm{I}}(\underline{u})) + (1-\alpha)F(V_{\mathrm{I}}(\underline{u}))},$$

and so

$$V_{\mathrm{S}}\Big(u_n^{s_n^1}\Big) = \begin{cases} 0 & \text{if } u_n^{s_n^1} = \overline{u} \\ \frac{(1-F(V_{\mathrm{I}}(\underline{u})))\alpha\Delta u}{1-F(V_{\mathrm{I}}(\underline{u}))+(1-\alpha)F(V_{\mathrm{I}}(\underline{u}))} & \text{if } u_n^{s_n^1} = \underline{u} \end{cases}. \tag{5}$$

Thus,

$$s_n^2 = \begin{cases} d & \text{if } u_n^{s_n^1} = \overline{u} \\ d & \text{if } u_n^{s_n^1} = \underline{u} \text{ and } c_n \geq V_{\mathrm{S}}(\underline{u}) \\ \neg s_n^1 & \text{if } u_n^{s_n^1} = \underline{u} \text{ and } c_n < V_{\mathrm{S}}(\underline{u}) \end{cases}.$$

Again, agent n discontinues the search if (i) the utility of the first searched alternative is high, or (ii) the utility of the first searched alternative is low, and her search cost is not smaller than the expected gain from the second search. Agent n searches for the remaining alternative otherwise.

By comparing equations (2) and (5) for $u_n^{s_n^1} = \underline{u}$, we observe that $V_{\mathrm{S}}(\underline{u}) < V_{\mathrm{I}}(\underline{u})$: if the first searched alternative has low utility, the expected gain from the second search for an agent with simple social information is smaller than that for an isolated agent. That is, even the simplest form of social information decreases an agent's incentives to engage in independent exploration, i.e., to search for the second alternative. This is so because the expected gain from the second search for an agent n with social information is a "discounted" version of that of an isolated agent. The discounting term $P(\underline{u})$ is the probability that agent n_0 did not search for alternative $\neg s_n^1$ given that an alternative with utility $\underline{u}$ was chosen, as only in this case the second search is valuable.

Since F has full support, $V_{\mathrm{S}}(\underline{u}) < V_{\mathrm{I}}(\underline{u})$ implies that $F(V_{\mathrm{S}}(\underline{u})) < F(V_{\mathrm{I}}(\underline{u}))$: an agent with simple social information is more likely to discontinue the search after sampling a low-utility alternative at the first search than an isolated. This is the *second difference* between search problems of types I and S.

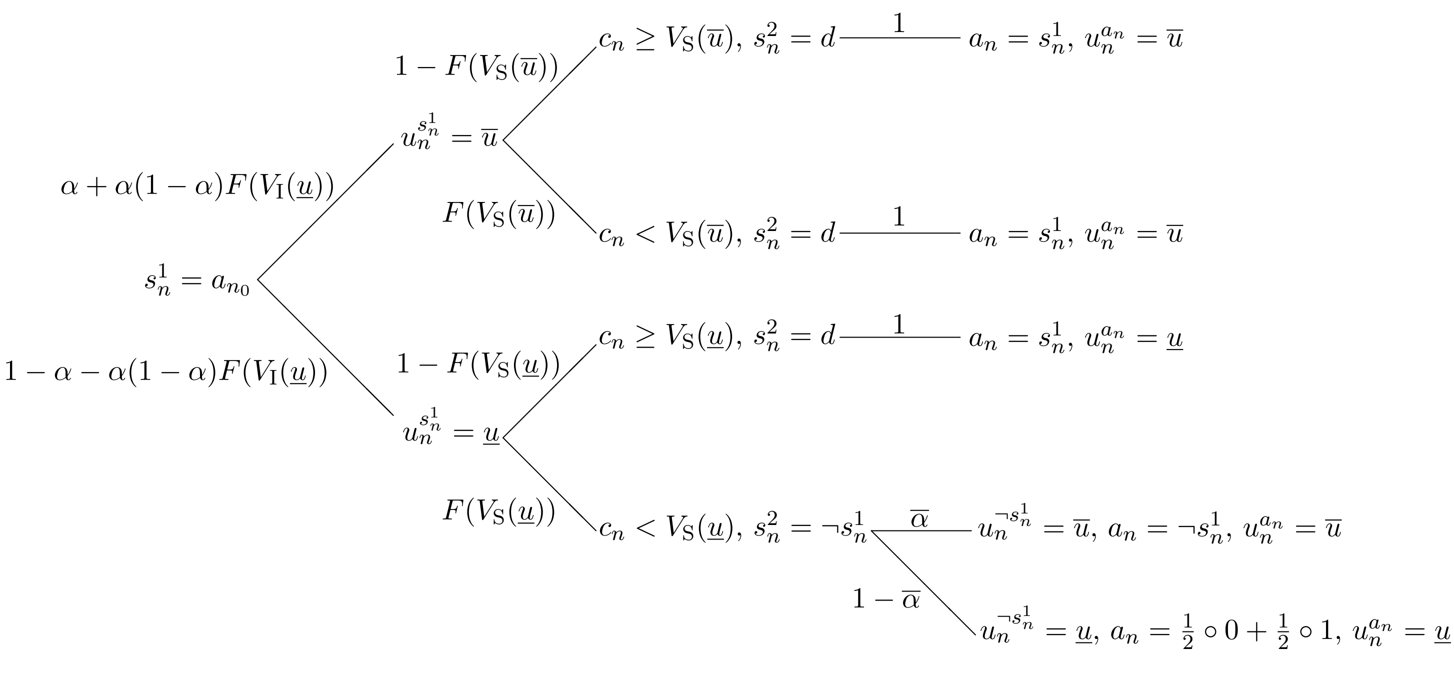
Choice Stage. Optimal choice is as in a search problem of type I (see equation (3)).

Decision Tree. Figure 2 summarizes agent n's decisions in a search problem of type S. In the game tree, we use the following notation

$$\overline{\alpha} := \mathbb{P}\Big(u_{n_0}^{\neg a_{n_0}} = \overline{u} \mid u_{n_0}^{a_{n_0}} = \underline{u}\Big) = \frac{1-\alpha}{1-\alpha F(V_{\mathrm{I}}(\underline{u}))},$$

where the equality holds by Bayes rule and Figure 1. Agent n begins the search from the alternative chosen by agent n_0 and observes the utility associated with such an alternative. By exploiting her social information, agent n is more likely to search first for a high-utility alternative than an isolated agent. Next, if her search cost is smaller than the expected gain from an additional search, agent n searches for the second alternative and observes the utility associated with such an alternative. Agent n discontinues the search otherwise. The expected gain from the second search is discounted by the probability that agent n_0 did not search the other alternative. As a result, agent n is more likely to discontinue her search conditional on the first searched alternative having utility $\underline{u}$ than an isolated agent. Finally, agent n chooses the best alternative among those she searched, randomizing uniformly if indifferent. At the end of each terminal node of the decision tree, we report the utility of the chosen alternative.

Figure 2: Decision Tree for a Search Problem of Type S.



3.2 Comparison of Optimal Decisions

The two differences between types of search problems have the following implications on the observables, from which the results about the lack of identification and biased estimation we present in the next subsection follow.

Observable Outcomes on Choice. An agent with simple social information is more likely to choose a high-utility alternative than an isolated agent:

$$\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{S}) > \mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{I}). \tag{6}$$

Inequality (6) follows from two observations. First, recall that an agent with simple social information samples a high-utility alternative at the first search with the same probability that an isolated agent chooses a high-utility alternative at the end of her search (recall the first difference between search problems of types I and S). Thus, the probability that an agent with simple social information chooses a high-utility alternative can be no smaller than that of an isolated agent. Second, if the utility of the first searched alternative is low, which occurs with positive probability, an agent with simple social information chooses a high-utility alternative if her search cost is sufficiently low (i.e., smaller than $V_{\mathrm{S}}(\underline{u})$) and the utility of the second searched alternative is high, which again occurs with positive probability.

Observable Outcomes on the Number of Searches. An agent with simple social information is more likely to discontinue her search than an isolated agent:

$$\mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{S}\Big) > \mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{I}\Big). \tag{7}$$

Agent n discontinues the search if either the utility of the first searched alternative is high or the utility of the first searched alternative is low, and her search cost is sufficiently high. Thus, inequality (7) follows from two observations. First, an agent with simple social information samples a high-utility alternative at the first search with a greater probability than an isolated agent—recall the first difference between search problems of types I and S. Second, an agent with simple social information discontinues the search after sampling a low-utility alternative at the first search with greater probability than an isolated agent—recall the second difference between search problems of types I and S.

3.3 Lack of Identification and Biased Estimation

With a binary utility distribution, if all agents are isolated, the researcher can identify and estimate the share $F(V_{\mathrm{I}}(\underline{u}))$ of agents with search costs below the threshold $V_{\mathrm{I}}(\underline{u})$. In our setting, this is the economically relevant primitive to measure search frictions because $V_{\mathrm{I}}(\underline{u})$ separates searchers—agents n with low search costs, $c_n \leq V_{\mathrm{I}}(\underline{u})$, that always search for the second alternative whenever the utility of the first search alternative is low—from non-searchers—agents n with high search costs, $c_n > V_{\mathrm{I}}(\underline{u})$, that never do so.

For each dataset, we show how to identify and estimate $F(V_{\mathrm{I}}(\underline{u}))$ when all agents are isolated. Next, we explain why identification fails if a positive share of search problems in the data-generating process is of type S and the researcher neglects social information. Depending on the dataset, neglecting social information may lead to under- or overestimation of the share $F(V_{\mathrm{I}}(\underline{u}))$ of agents with low search costs. Under- or overestimation of $F(V_{\mathrm{I}}(\underline{u}))$ corresponds to, respectively, over- or underestimation of the level of search costs in the population or, equivalently, the search frictions in the environment.

3.3.1 Data on Choice

For some sample size N, the researcher observes $\overline{u}_N$, the empirical distribution of agents who choose an alternative with utility $\overline{u}$:

$$\overline{u}_N := \frac{\sum_{n=1}^{N} \mathbb{1}_{\{u_n^{a_n}=\overline{u}\}}}{N}.$$

Preliminary Observations. We first characterize the probability that an agent chooses an alternative with high utility in the data-generating process. Since a share γ of agents in the population has simple social information, the law of total probability implies that

$$\mathbb{P}(u_n^{a_n}=\overline{u}) = \mathbb{P}(u_n^{a_n}=\overline{u} \mid \theta_n=\mathrm{I})\gamma + \mathbb{P}(u_n^{a_n}=\overline{u} \mid \theta_n=\mathrm{S})(1-\gamma). \tag{8}$$

By Figures 1 and 2,

$$\mathbb{P}(u_n^{a_n}=\overline{u} \mid \theta_n=\mathrm{I}) = \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})), \tag{9}$$

and

$$\mathbb{P}(u_n^{a_n}=\overline{u} \mid \theta_n=\mathrm{S}) = \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})) + (1-\alpha)^2 F(V_{\mathrm{S}}(\underline{u})). \tag{10}$$

Hence, by equations (8)–(10),

$$\mathbb{P}(u_n^{a_n}=\overline{u}) = \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})) + (1-\gamma)(1-\alpha)^2 F(V_{\mathrm{S}}(\underline{u})). \tag{11}$$

Moreover, by the strong law of large numbers,

$$\overline{u}_N \xrightarrow{a.s.} \mathbb{E}[\overline{u}_N] = \mathbb{P}(u_n^{a_n}=\overline{u}). \tag{12}$$

That is, $\overline{u}_N$ is an unbiased and strongly consistent estimator of $\mathbb{P}(u_n^{a_n}=\overline{u})$, the probability that an agent in the population chooses an alternative with high utility.

Identification and Estimation when All Agents Are Isolated. Suppose all agents are isolated, i.e., $\gamma=1$. By equations (8) and (9), $F(V_{\mathrm{I}}(\underline{u}))$ is identified by

$$\mathbb{P}(u_n^{a_n}=\overline{u}) = \mathbb{P}(u_n^{a_n}=\overline{u} \mid \theta_n=\mathrm{I}) = \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})), \tag{13}$$

or, equivalently,

$$F(V_{\mathrm{I}}(\underline{u})) = \frac{\mathbb{P}(u_n^{a_n}=\overline{u} \mid \theta_n=\mathrm{I}) - \alpha}{\alpha(1-\alpha)}. \tag{14}$$

Replacing $\mathbb{P}(u_n^{a_n}=\overline{u})$ with its sample analog $\overline{u}_N$ in equation (14), we have

$$\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N := \frac{\overline{u}_N - \alpha}{\alpha(1-\alpha)}, \tag{15}$$

which, by the convergence and the equality in (12), is an unbiased and strongly consistent estimator of $F(V_{\mathrm{I}}(\underline{u}))$.

Lack of Identification and Biased Estimation with Simple Social Information. The next proposition summarizes the identification and estimation of $F(V_{\mathrm{I}}(\underline{u}))$ with data on choice when a positive share of search problems is of type S, i.e., $\gamma < 1$, but the researcher assumes that all agents are isolated, i.e., $\gamma = 1$.

Proposition 1. *Let $\gamma < 1$. Suppose the researcher observes data on choice and assumes $\gamma = 1$. Then:*

1. *$F(V_{\mathrm{I}}(\underline{u}))$ is not identified by equation (13) or, equivalently, (14).*
2. *The estimator $\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N$ in equation (15) is biased and inconsistent, and search costs are underestimated.*

Proof. [*Part 1.*] By equations (6) and (8), $\mathbb{P}(u_n^{a_n} = \overline{u}) > \mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{I})$. That $F(V_{\mathrm{I}}(\underline{u}))$ is not identified by equation (13) or, equivalently, (14) follows.

[*Part 2.*] To begin, note that

$$\begin{aligned}\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N &\xrightarrow{a.s.} \mathbb{E}\Big[\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N\Big] \\ &= F(V_{\mathrm{I}}(\underline{u})) + \frac{(1-\gamma)(1-\alpha)F(V_{\mathrm{S}}(\underline{u}))}{\alpha} \\ &> F(V_{\mathrm{I}}(\underline{u})),\end{aligned} \tag{16}$$

where: the equality holds by equation (11); The inequality holds because $F(V_{\mathrm{S}}(\underline{u})) > 0$. That $\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N$ is biased and inconsistent, and search costs are underestimated, follows.■

Social information increases the probability that agents in the population choose a high-utility alternative, and such a probability is used to identify $F(V_{\mathrm{I}}(\underline{u}))$ with data on choice. Hence, if the researcher neglects social information, $F(V_{\mathrm{I}}(\underline{u}))$ is not identified, and search costs are underestimated.

By assuming all agents are isolated, the researcher infers low search costs for all agents who choose a high-utility alternative. Some of these agents, however, secure a high utility because they exploit their social information and not because of low search costs. Thus, the researcher overestimates the share $F(V_{\mathrm{I}}(\underline{u}))$ of agents with low search costs or, equivalently, underestimates the search frictions in the environment.

3.3.2 Data on the Number of Searches

For some sample size N, the researcher observes d_N, the empirical distribution of agents who conduct only one search:

$$d_N := \frac{\sum_{n=1}^{N} \mathbb{1}_{\{s_n^2 = d\}}}{N}.$$

Preliminary Observations. We first characterize the probability that an agent discontinues the search in the data-generating process. Since a share γ of agents in the population has simple social information, the law of total probability implies that

$$\mathbb{P}\Big(s_n^2 = d\Big) = \mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{I}\Big)\gamma + \mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{S}\Big)(1-\gamma). \tag{17}$$

By Figures 1 and 2,

$$\mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{I}\Big) = \alpha + (1-\alpha)[1 - F(V_\mathrm{I}(\underline{u}))], \tag{18}$$

and

$$\mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{S}\Big) = \alpha + (1-\alpha)[1 - F(V_\mathrm{S}(\underline{u}))] + \alpha(1-\alpha)F(V_\mathrm{I}(\underline{u}))F(V_\mathrm{S}(\underline{u})). \tag{19}$$

Hence, by equations (17)–(19),

$$\begin{aligned}\mathbb{P}\Big(s_n^2 = d\Big) = \alpha &+ (1-\alpha)[1 - \gamma F(V_\mathrm{I}(\underline{u})) - (1-\gamma)F(V_\mathrm{S}(\underline{u}))] \\ &+ (1-\gamma)\alpha(1-\alpha)F(V_\mathrm{I}(\underline{u}))F(V_\mathrm{S}(\underline{u})).\end{aligned} \tag{20}$$

Moreover, by the strong law of large numbers,

$$d_N \xrightarrow{a.s.} \mathbb{E}[d_N] = \mathbb{P}\Big(s_n^2 = d\Big). \tag{21}$$

That is, d_N is an unbiased and strongly consistent estimator of $\mathbb{P}(s_n^2 = d)$, the probability that an agent in the population discontinues the search.

Identification and Estimation when All Agents Are Isolated. Suppose all agents are isolated, i.e., $\gamma = 1$. By equations (17) and (18), $F(V_\mathrm{I}(\underline{u}))$ is identified by

$$\mathbb{P}\Big(s_n^2 = d\Big) = \mathbb{P}\Big(s_n^2 = d \mid \theta_n = \mathrm{I}\Big) = \alpha + (1-\alpha)[1 - F(V_\mathrm{I}(\underline{u}))], \tag{22}$$

or, equivalently,

$$F(V_\mathrm{I}(\underline{u})) = \frac{1 - \mathbb{P}(s_n^2 = d)}{1-\alpha}. \tag{23}$$

Replacing $\mathbb{P}\Big(s_n^2 = d\Big)$ with its sample analog d_N in equation (23), we have

$$\widehat{F(V_\mathrm{I}(\underline{u}))}_N := \frac{1 - d_N}{1-\alpha}, \tag{24}$$

which, by the convergence and the equality in (21), is an unbiased and strongly consistent estimator of $F(V_\mathrm{I}(\underline{u}))$.

Lack of Identification and Biased Estimation with Simple Social Information. The next proposition summarizes the identification and estimation of $F(V_\mathrm{I}(\underline{u}))$ with data on the number of searches when a positive share of search problems is of type S, i.e., $\gamma < 1$, but the researcher assumes that all agents are isolated, i.e., $\gamma = 1$.

Proposition 2. *Let $\gamma < 1$. Suppose the researcher observes data on the number of searches and assumes $\gamma = 1$. Then:*

1. $F(V_{\mathrm{I}}(\underline{u}))$ *is not identified by equation* (22) *or, equivalently,* (23).
2. *The estimator* $\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N$ *in equation* (24) *is biased and inconsistent, and search costs are overestimated.*

Proof. [*Part 1.*] By equations (7) and (17), $\mathbb{P}\left(s_n^2 = d\right) > \mathbb{P}\left(s_n^2 = d \mid \theta_n = \mathrm{I}\right)$. That $F(V_{\mathrm{I}}(\underline{u}))$ is not identified by equation (22) or, equivalently, (23) follows.

[*Part 2.*] To begin, note that

$$\begin{aligned}\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N &\xrightarrow{a.s.} \mathbb{E}\left[\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N\right] \\ &= F(V_{\mathrm{I}}(\underline{u})) + (1-\gamma)\Big\{F(V_{\mathrm{S}}(\underline{u})) - [1 + \alpha F(V_{\mathrm{S}}(\underline{u}))]F(V_{\mathrm{I}}(\underline{u}))\Big\} \\ &< F(V_{\mathrm{I}}(\underline{u})),\end{aligned} \tag{25}$$

where: the equality holds by equation (20); the inequality holds because $F(V_{\mathrm{S}}(\underline{u})) < F(V_{\mathrm{I}}(\underline{u}))$. That $\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N$ is biased and inconsistent, and search costs are overestimated, follows. ■

Social information increases the probability that agents in the population search only once, and such a probability is used to identify $F(V_{\mathrm{I}}(\underline{u}))$ with data on the number of searches. Hence, if the researcher neglects the presence of social information, $F(V_{\mathrm{I}}(\underline{u}))$ is not identified, and search costs are overestimated.

By assuming all agents are isolated, the researcher infers high search costs for all agents searching only once. Some of these agents, however, search only once because they exploit their social information and not because of high search costs. Thus, the researcher underestimates the share $F(V_{\mathrm{I}}(\underline{u}))$ of agents with low search costs or, equivalently, overestimates the search frictions in the environment.

3.4 Discussion and Generalizations

The mechanism driving our lack of identification and biased estimation results is that social information changes the distribution of observable outcomes that identify the search model in a specific way. In particular, social information makes agents more likely to choose alternatives with higher utility and search less than when they are isolated.

These observations do not depend on the parsimonious model specification that we adopt. They are general features that would emerge in any bandit model in which agents trade off the exploitation of social information with independent exploration (i.e., search) for the best alternative. Therefore, although deriving the same results in more general specifications of the search model may become algebraically more complicated (hence, less transparent), and the precise quantitative effect may differ from one specification to another, the conceptual insights we uncover with our parsimonious model are robust and remain generally applicable. In particular, qualitatively analogous insights would emerge with more than two available alternatives, non-perfectly correlated utilities across the agent and the peer, ex-ante differentiated alternatives, utilities of the agent and the peer that are not perfectly correlated, alternative-specific search costs, utilities and search costs of the agent and the peer that are correlated, if each agent observes more than one peer or a random number of them, data on optimal stopping, and in a simultaneous (as opposed to sequential) search model, among many other cases.

Below, we discuss in some detail three generalizations of our previous results.[6]

Data on Optimal Stopping. To begin, we consider a different dataset. Suppose that, for some sample size N, the researcher observes $d_N^{\underline{u}}$, the empirical distribution of agents who discontinue the search after searching for a first alternative with utility $\underline{u}$:

$$d_N^{\underline{u}} := \frac{\sum_{n=1}^N \mathbb{1}_{\{s_n^2=d\}} \mathbb{1}_{\left\{u_n^{s_n^1}=\underline{u}\right\}}}{\sum_{n=1}^N \mathbb{1}_{\left\{u_n^{s_n^1}=\underline{u}\right\}}}.$$

Data on optimal stopping are readily available for price search models. The researcher needs to observe the shares of agents who discontinue the search (and purchase) when the price of the first searched alternative is high. Data on optimal stopping, instead, may not always be available for match-value search models, as match values may not be observable.

If the researcher assumes all agents search in isolation, she infers high search costs for all agents who discontinue the search after searching for a first alternative with low utility. However, some agents observe a peer choosing that specific alternative. Since the peers have possibly searched twice, agents infer that the other alternative is not likely to have a higher utility than the one they searched first. Therefore, they discontinue the search despite the first searched alternative having low utility because they exploit social information and not because they have a high search cost. Formally, *ceteris paribus* (and, in particular, given the same search cost distribution), the probability that agents in the population discontinue the search after searching for a first alternative with low utility is greater with social information than without. As a result, by neglecting social information, the search cost distribution is not identified, and the researcher overestimates search costs.

Simultaneous Search. All insights remain valid if the researcher assumes agents collect information about the utility of the available alternatives via simultaneous search à la Stigler (1961), the other workhorse model in the empirical search literature. Under simultaneous search, each agent commits to searching a fixed set of alternatives before she begins searching. If the agent has social information, she observes the choices of some of her peers before committing, but neither the peers' searches nor their search cost.

For analogous reasons to those at play in a sequential search model, social information changes the distribution that generates the observable outcomes identifying the search

[6] The formal derivations, which we omit, are available upon request and are part of an earlier version of this paper that we circulated as CEPR Discussion Paper # DP17740.

model. In particular, agents with social information are more likely to choose alternatives with higher utility and commit to searching fewer alternatives than isolated agents. As a result, both with data on choice and the number of searches, neglecting social information results in the non-identification of the search cost distribution.[7] Since social information plays an analogous role in the two models, how the resulting estimation bias depends on the dataset's content is the same as under sequential search.

Non-Binary Utility Distribution. Suppose the utility distribution is non-binary, i.e., $U := \{u_1, u_2, \dots, u_I\}$ for some $I > 2$. If so, and the researcher knows or can consistently estimate the utility distribution, she can identify and estimate $F(V_{\mathrm{I}}(u_i))$ for all $i = 1, \dots, I$, where $V_{\mathrm{I}}(u_i)$ denotes the expected gain from the second search of an isolated agent after searching for an alternative with utility u_i first. Thus, by enriching the model and bringing it closer to real empirical applications with more than two utility levels, the researcher can recover more information (i.e., more quantiles) about the search cost distribution.

All our insights, however, apply to this case. By neglecting social information, the researcher: (i) underestimates $F(V_{\mathrm{I}}(u_i))$ for all $i = 1, \dots, I$ (i.e., overestimates search frictions) with data on choice, and (ii) overestimates $F(V_{\mathrm{I}}(u_i))$ for all $i = 1, \dots, I$ (i.e., underestimates search frictions) with data on the number of searches. Again, the reasons are the ones we highlighted with our parsimonious model: agents with social information are more likely to choose alternatives with higher utility and search less than isolated agents.

4 Remedies

In this section, we develop approaches that allow the researcher to restore identification and consistent estimation.

First, in Section 4.1, we develop a robust approach that allows the researcher to recover informative bounds on the search cost distribution while imposing only minimal assumptions on the environment, namely, only those required by the General Social Information setting. This partial identification approach is motivated by the observation that real social networks, communication channels, and informational externalities among agents are complex. Hence, it may be hard for a researcher to specify the exact content and type of social information agents have access to. As we will argue, under this approach, the researcher can recover robust bounds on search cost distributions when agents' amount and type of social information are unobserved.

Next, using the Simple Social Information setting as the baseline environment, we show how additional data requirements or stronger assumptions (as opposed to the minimal ones under General Social Information) allow the researcher to restore identification and obtain more informative estimates. We consider two cases. In Section 4.2, we assume that

[7] As optimal stopping is not well-defined with simultaneous search, such data play no role in this setting.

the researcher has available agent-level data that help distinguish between isolated agents and agents with social information (Section 4.2). In Section 4.3, we present two partial identification approaches, one that requires estimating offline the share of agents with social information and the other accounting for the lack of any information on such a share.

4.1 General Social Information

Let θ_n denote agent n's "amount" of social information. Under the General Social Information setting, for any search problems n, such an amount θ_n is bounded by the minimal and the maximal social information.

Agent n's social information is minimal, $\theta_n = \underline{\theta}$, if agent n is isolated. In this case, by equation (9),

$$\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \underline{\theta}) = \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})),$$

and, by equation (18),

$$\mathbb{P}\Big(s_n^2 = d \mid \theta_n = \underline{\theta}\Big) = \alpha + (1-\alpha)[1 - F(V_{\mathrm{I}}(\underline{u}))].$$

Agent n's social information is maximal, $\theta_n = \overline{\theta}$, if agent n chooses the alternative with the highest realized utility at the first search just by exploiting her social information. That is, agent n's social information is so abundant that it suffices her to identify the best alternative. If so, agent n: (i) chooses an alternative with high utility if and only if $\max\{u_n^0, u_n^1\} = \overline{u}$, which occurs with probability $\alpha(2-\alpha)$; and (ii) always discontinues her search. Thus,

$$\mathbb{P}\Big(u_n^{a_n} = \overline{u} \mid \theta_n = \overline{\theta}\Big) = \alpha(2-\alpha),$$

and

$$\mathbb{P}\Big(s_n^2 = d \mid \theta_n = \overline{\theta}\Big) = 1.$$

No matter the exact amount of agents' social information or where it comes from, it will always be between the minimal and the maximal ones. Hence, without further assumptions on social information, the probability that any agent n in the population chooses a high-utility alternative must satisfy

$$\alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})) \leq \mathbb{P}(u_n^{a_n} = \overline{u}) \leq \alpha(2-\alpha), \tag{26}$$

and the probability that any agent n in the population discontinues her search must satisfy

$$\alpha + (1-\alpha)[1 - F(V_{\mathrm{I}}(\underline{u}))] \leq \mathbb{P}\Big(s_n^2 = d\Big) \leq 1. \tag{27}$$

The bounds on the probability that agents choose a high-utility alternative and discontinue the search described by conditions (26) and (27) hold with no assumption on social information. First, to derive such bounds, there is no need to make assumptions about the agents' type of social information, i.e., whether agents learn from others' experiences via observation, communication, advertising, ratings, reviews, etc. Second, there is no need to make assumptions about the agents' amount of social information, i.e., how many peers they observe or talk to, how many other peers, in turn, these peers have observed or talked to, the information content of the peers' choices or words, or the information in the advertisements, ratings, or reviews agents have access to. Finally, such bounds hold independently of how agents break any indifference they may face in their search and choice problems. As a result, with the set estimator that we construct in the remaining part of this section, the researcher can derive robust bounds on the search cost distribution when agents' amount and type of social information are unobserved by remaining agnostic about most features of the search environment.

Now let G denote any search cost distribution with full support on $[0, \bar{c}]$ (not necessarily the true distribution F). By the previous observations, we have that, under General Social Information, the interval of model predictions about choice with search cost distribution G, denoted by $\mathrm{P_c}(G)$, is

$$\mathrm{P_c}(G) \coloneqq [\alpha + \alpha(1-\alpha)G(V_{\mathrm{I}}(\underline{u})), \alpha(2-\alpha)], \tag{28}$$

and the interval of model predictions about the number of searches with search cost distribution G, denoted by $\mathrm{P_{ns}}(G)$, is

$$\mathrm{P_{ns}}(G) \coloneqq [\alpha + (1-\alpha)[1 - G(V_{\mathrm{I}}(\underline{u}))], 1]. \tag{29}$$

Suppose that amounts θ_n are i.i.d. across search problems according to some distribution H_θ unknown to the researcher. By the strong law of large numbers, the empirical distribution of agents who choose an alternative with high utility, $\overline{u}_N$, converges almost surely to some probability distribution $\mathbb{P}(u_n^{a_n} = \overline{u})$ corresponding to the probability that an agent in the population chooses an alternative with high utility. Such a probability distribution must be an element of the set of model predictions about choice $\mathrm{P_c}(F)$ (i.e., the one corresponding to the true search cost distribution F). That is,

$$\overline{u}_N \xrightarrow{a.s.} \mathbb{E}[\overline{u}_N] = \mathbb{P}(u_n^{a_n} = \overline{u}) \in \mathrm{P_c}(F). \tag{30}$$

Similarly, again by the strong law of large numbers, the empirical distribution of agents who conduct only one search, d_N, converges almost surely to some probability distribution $\mathbb{P}(s_n^2 = d)$ corresponding to the probability that an agent in the population discontinues the search. Such a probability distribution must be an element of the set of model predictions about the number of searches $\mathrm{P_{ns}}(F)$ (i.e., the one corresponding to the true search cost distribution F). That is,

$$d_N \xrightarrow{a.s.} \mathbb{E}[d_N] = \mathbb{P}\Big(s_n^2 = d\Big) \in \mathrm{P_{ns}}(F). \tag{31}$$

Suppose the researcher has access to both data on choice and the number of searches. By inequalities (26) and (27) and the convergences in (30) and (31), the identified set for

$F(V_{\mathrm{I}}(\underline{u}))$, denoted by Λ^{gsi}, consists of all $G(V_{\mathrm{I}}(\underline{u})) \in [0,1]$ compatible with $\mathbb{P}(u_n^{a_n} = \overline{u})$ and $\mathbb{P}(s_n^2 = d)$ as model predictions about choice and the number of searches for some level of social information between the minimal and the maximal ones. That is, the identified set for $F(V_{\mathrm{I}}(\underline{u}))$ is

$$\Lambda^{\mathrm{gsi}} := \Big\{ G(V_{\mathrm{I}}(\underline{u})) \in [0,1] : \mathbb{P}(u_n^{a_n} = \overline{u}) \in \mathrm{P_c}(G) \ \text{ and } \ \mathbb{P}\Big(s_n^2 = d\Big) \in \mathrm{P_{ns}}(G) \Big\}. \tag{32}$$

By the definition of $\mathrm{P_c}(G)$ and $\mathrm{P_{ns}}(G)$ in (28) and (29), the set Λ^{gsi} defined by (32) is equivalent to the set of all $G(V_{\mathrm{I}}(\underline{u})) \in [0,1]$ for which equations (26) and (27) hold true (when distribution F is replaced by distribution G), and so

$$\Lambda^{\mathrm{gsi}} = \left\{ G(V_{\mathrm{I}}(\underline{u})) \in [0,1] : \frac{1 - \mathbb{P}\Big(s_n^2 = d\Big)}{1-\alpha} \leq G(V_{\mathrm{I}}(\underline{u})) \leq \frac{\mathbb{P}(u_n^{a_n} = \overline{u}) - \alpha}{\alpha(1-\alpha)} \right\}. \tag{33}$$

Replacing $\mathbb{P}(u_n^{a_n} = \overline{u})$ and $\mathbb{P}(s_n^2 = d)$ with their sample analogs $\overline{u}_N$ and d_N in the set on the right-hand side of equality (33), we obtain the set estimator

$$\widehat{\Lambda}_N^{\mathrm{gsi}} := \left\{ G(V_{\mathrm{I}}(\underline{u})) \in [0,1] : \frac{1 - d_N}{1-\alpha} \leq G(V_{\mathrm{I}}(\underline{u})) \leq \frac{\overline{u}_N - \alpha}{\alpha(1-\alpha)} \right\}.$$

The next proposition, which follows from the convergences in (30) and (31), establishes that the set estimator $\widehat{\Lambda}_N^{\mathrm{gsi}}$ almost surely contains the true parameter of interest as $N \to \infty$.

Proposition 3. *As $N \to \infty$, $F(V_{\mathrm{I}}(\underline{u})) \in \widehat{\Lambda}_N^{\mathrm{gsi}}$ almost surely.*

It is worth making a few remarks about the partial-identification approach under general social information.

Empirical Implementation. The set estimator $\widehat{\Lambda}_N^{\mathrm{gsi}}$ is easy to construct in practice. Its implementation only requires estimating $F(V_{\mathrm{I}}(\underline{u}))$ twice, once with data on the number of searches and once with data on choice, under the assumption that all agents are isolated. The former estimate corresponds to the minimum of set $\widehat{\Lambda}_N^{\mathrm{gsi}}$, the latter estimate to the maximum of set $\widehat{\Lambda}_N^{\mathrm{gsi}}$. Hence, the researcher can construct the set estimator $\widehat{\Lambda}_N^{\mathrm{gsi}}$ by estimating the model under the standard assumption in the empirical search literature that all agents act in isolation.

The distance between the minimum and the maximum of set $\widehat{\Lambda}_N^{\mathrm{gsi}}$ provides a measure of how misguided conclusions can be when neglecting social information. The true search cost distribution or, better, the true value of $F(V_{\mathrm{I}}(\underline{u}))$, will almost surely (as $N \to \infty$) be between these two bounds, i.e., between the two estimates of $F(V_{\mathrm{I}}(\underline{u}))$ with data on the number of searches and data on choice under the assumption that all agents are isolated.

Other Features of the Search Cost Distribution. For any amount of social information θ in the support of the unknown distribution H_θ, let $F(V_\theta(\underline{u}))$ denote the expected gain from the second search of an agent with the amount of social information θ after searching for a first alternative with utility $\underline{u}$. Under our general social information approach, the researcher can recover only $F(V_{\mathrm{I}}(\underline{u}))$ (i.e., $F(V_\theta(\underline{u}))$ for $\theta = \underline{\theta}$). Ideally, the researcher would want to recover more information about the search cost distribution, namely, the value of $F(V_\theta(\underline{u}))$ for all θ in the support of H_θ. With no assumption on social information, this is not possible.

However, we note that $F(V_{\mathrm{I}}(\underline{u}))$ already provides an informative measure of search frictions in the environment. The reason is that the threshold $V_{\mathrm{I}}(\underline{u})$ separates searchers from non-searchers. Namely, agents whose search costs are smaller than $V_{\mathrm{I}}(\underline{u})$ search for the second alternative with positive probability (depending on their amount of social information) after searching for a first action with low utility; hence, they are searchers. In contrast, agents whose search costs are greater than $V_{\mathrm{I}}(\underline{u})$ never search for the second alternative after searching for a first action with low utility, independently of their amount of social information; hence, they are non-searchers.

4.2 Agent-Level Data on Social Information

In this and the remaining sections, we assume the Simple Social Information setting is the data-generating process, and we focus on data on choice. Analogous arguments apply to different datasets, e.g., data on the number of searches or optimal stopping.

Suppose the researcher can distinguish isolated agents from agents with social information. Formally, for some sample size N, the researcher observes

$$\overline{u}_N(\mathrm{I}) := \frac{\sum_{n=1}^{N} \mathbb{1}_{\{u_n^{a_n}=\overline{u}\}} \mathbb{1}_{\{\theta_n=\mathrm{I}\}}}{\sum_{n=1}^{N} \mathbb{1}_{\{\theta_n=\mathrm{I}\}}} \quad \text{and} \quad \overline{u}_N(\mathrm{S}) := \frac{\sum_{n=1}^{N} \mathbb{1}_{\{u_n^{a_n}=\overline{u}\}} \mathbb{1}_{\{\theta_n=\mathrm{S}\}}}{\sum_{n=1}^{N} \mathbb{1}_{\{\theta_n=\mathrm{S}\}}}.$$

In this case, the researcher can implement a two-step procedure to identify and estimate, consistently and without bias, both $F(V_{\mathrm{I}}(\underline{u}))$ and $F(V_{\mathrm{S}}(\underline{u}))$, as we next show.

Step 1. Consider first isolated agents. By the strong law of large numbers,

$$\overline{u}_N(\mathrm{I}) \xrightarrow{a.s.} \mathbb{E}[\overline{u}_N(\mathrm{I})] = \mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{I}). \tag{34}$$

That is, $\overline{u}_N(\mathrm{I})$ is an unbiased and strongly consistent estimator of $\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{I})$, the probability that an isolated agent chooses an alternative with high utility.

By equation (9), $F(V_{\mathrm{I}}(\underline{u}))$ is identified by

$$F(V_{\mathrm{I}}(\underline{u})) = \frac{\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{I}) - \alpha}{\alpha(1-\alpha)}. \tag{35}$$

Replacing $\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{I})$ with its sample analog $\overline{u}_N(\mathrm{I})$ in equation (35), we have

$$\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N := \frac{\overline{u}_N(\mathrm{I}) - \alpha}{\alpha(1-\alpha)}, \tag{36}$$

which, by the convergence and the equality in (34), is an unbiased and strongly consistent estimator of $F(V_{\mathrm{I}}(\underline{u}))$.

Step 2. Consider now agents with simple social information. Since the researcher observes $\overline{u}_N(\mathrm{S})$, once $F(V_{\mathrm{I}}(\underline{u}))$ is identified and consistently estimated, it is also possible to identify and consistently estimate $F(V_{\mathrm{S}}(\underline{u}))$. By the strong law of large numbers,

$$\overline{u}_N(\mathrm{S}) \xrightarrow{a.s.} \mathbb{E}[\overline{u}_N(\mathrm{S})] = \mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{S}). \tag{37}$$

That is, $\overline{u}_N(\mathrm{S})$ is an unbiased and strongly consistent estimator of $\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{S})$, the probability that an agent with simple social information chooses an alternative with high utility.

By equation (10), $F(V_{\mathrm{S}}(\underline{u}))$ is identified by

$$F(V_{\mathrm{S}}(\underline{u})) = \frac{\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{S}) - \alpha - \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u}))}{(1-\alpha)^2}. \tag{38}$$

Replacing $\mathbb{P}(u_n^{a_n} = \overline{u} \mid \theta_n = \mathrm{S})$ with its sample analog $\overline{u}_N(\mathrm{S})$ and $F(V_{\mathrm{I}}(\underline{u}))$ with its estimator $\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N$ in equation (38), we have

$$\widehat{F(V_{\mathrm{S}}(\underline{u}))}_N := \frac{\overline{u}_N(\mathrm{S}) - \alpha - \alpha(1-\alpha)\widehat{F(V_{\mathrm{I}}(\underline{u}))}_N}{(1-\alpha)^2},$$

which, by the convergences and the equalities in (34) and (37), is an unbiased and strongly consistent estimator of $F(V_{\mathrm{S}}(\underline{u}))$.

4.3 Partial Identification Approaches

Agent-level data on social information are hardly available. Thus, we consider alternative partial-identification approaches that allow for the possibility that the researcher has no access to such a detailed dataset.

4.3.1 Estimating γ Offline

Suppose the researcher can identify and consistently estimate offline the share of agents with simple social information, e.g., by using detailed network data on the agents with access to social media or survey evidence on agents' reliance on the choices of a peer for purchase decisions. Hence, hereafter, we assume the researcher knows γ.

Equation (11) in Section 3.3.1, which we here report,

$$\mathbb{P}(u_n^{a_n} = \overline{u}) = \alpha + \alpha(1-\alpha)F(V_{\mathrm{I}}(\underline{u})) + (1-\gamma)(1-\alpha)^2 F(V_{\mathrm{S}}(\underline{u})), \tag{39}$$

describes the probability that an agent in the population chooses an alternative with high utility. Equation (39) implies that neither $F(V_{\mathrm{I}}(\underline{u}))$ nor $F(V_{\mathrm{S}}(\underline{u}))$ can be point-identified even if the researcher knows γ. The researcher, however, can rely on a partial identification approach to recover bounds on these quantities.

Hereafter, let G denote any search cost distribution with full support on $[0, \overline{c}]$ (not necessarily the true distribution F). The joint identified set for $(F(V_{\mathrm{I}}(\underline{u})), F(V_{\mathrm{S}}(\underline{u})))$ given

γ, denoted by $\Lambda^{\text{off}}(\gamma)$, consists of all $(G(V_{\text{I}}(\underline{u})), G(V_{\text{S}}(\underline{u})))) \in [0,1] \times [0,1]$ compatible with $\mathbb{P}(u_n^{a_n} = \overline{u})$ as a model prediction for the given γ. That is,

$$\begin{aligned}\Lambda^{\text{off}}(\gamma) := \Big\{&(G(V_{\text{I}}(\underline{u})), G(V_{\text{S}}(\underline{u}))) \in [0,1] \times [0,1] : G(V_{\text{S}}(\underline{u})) < G(V_{\text{I}}(\underline{u}), \text{ and} \\ &\mathbb{P}(u_n^{a_n} = \overline{u}) = \alpha + \alpha(1-\alpha)G(V_{\text{I}}(\underline{u})) + (1-\gamma)(1-\alpha)^2 G(V_{\text{S}}(\underline{u}))\Big\}.\end{aligned} \tag{40}$$

The restriction $G(V_{\text{S}}(\underline{u})) < G(V_{\text{I}}(\underline{u}))$ in definition (40) captures that, conditional on the first searched alternative having low utility, the expected gain from the second search for an agent with social information is lower than that for an isolated agent, that is, $V_{\text{S}}(\underline{u}) < V_{\text{I}}(\underline{u})$.

The identified set for $F(V_{\text{I}}(\underline{u}))$ given γ, denoted by $\Lambda^{\text{off}}_{F_{\text{I}}}(\gamma)$, consists of all $G(V_{\text{I}}(\underline{u})) \in [0,1]$ compatible with $\mathbb{P}(u_n^{a_n} = \overline{u})$ as a model prediction for the given γ and some $G(V_{\text{S}}(\underline{u}))) < G(V_{\text{I}}(\underline{u})))$. The identified set for $F(V_{\text{S}}(\underline{u}))$ given γ, denoted by $\Lambda^{\text{off}}_{F_{\text{S}}}(\gamma)$, is analogously defined. Equivalently, the identified set $\Lambda^{\text{off}}_{F_{\text{I}}}(\gamma)$ (resp., $\Lambda^{\text{off}}_{F_{\text{S}}}(\gamma)$) is the projection of $\Lambda^{\text{off}}(\gamma)$ along its first (resp., second) dimension.

Replacing $\mathbb{P}(u_n^{a_n} = \overline{u})$ with its sample analog $\overline{u}_N$ in the definitions of identifies sets $\Lambda^{\text{off}}(\gamma)$, $\Lambda^{\text{off}}_{F_{\text{I}}}(\gamma)$, and $\Lambda^{\text{off}}_{F_{\text{S}}}(\gamma)$ above, we obtain the corresponding set estimators. By the convergence in (12), such set estimators almost surely contain the true parameters of interest as $N \to \infty$.

4.3.2 Unknown γ

Suppose the researcher knows nothing about the level of simple social information beyond that $\gamma \in (0,1]$. In this case, the researcher can aim at jointly identifying $(F(V_{\text{I}}(\underline{u})), F(V_{\text{S}}(\underline{u})), \gamma)$.

Again, let G denote any search cost distribution with full support on $[0, \overline{c}]$ (not necessarily the true distribution F). The joint identified set for $(F(V_{\text{I}}(\underline{u})), F(V_{\text{S}}(\underline{u})), \gamma)$, denoted by Λ^{unk}, consists of all $(G(V_{\text{I}}(\underline{u})), G(V_{\text{S}}(\underline{u}))), \tilde{\gamma}) \in [0,1] \times [0,1] \times (0,1]$ compatible with $\mathbb{P}(u_n^{a_n} = \overline{u})$ as a model prediction. That is,

$$\begin{aligned}\Lambda^{\text{unk}} := \Big\{&(G(V_{\text{I}}(\underline{u})), G(V_{\text{S}}(\underline{u})), \tilde{\gamma}) \in [0,1] \times [0,1] : G(V_{\text{S}}(\underline{u})) < G(V_{\text{I}}(\underline{u}), \text{ and} \\ &\mathbb{P}(u_n^{a_n} = \overline{u}) = \alpha + \alpha(1-\alpha)G(V_{\text{I}}(\underline{u})) + (1-\tilde{\gamma})(1-\alpha)^2 G(V_{\text{S}}(\underline{u}))\Big\}.\end{aligned}$$

The joint identified set for $(F(V_{\text{I}}(\underline{u})), F(V_{\text{S}}(\underline{u})))$ when γ is unknown, denoted by $\Lambda^{\text{unk}}_{F_{\text{I}},F_{\text{S}}}$, consists of all $(G(V_{\text{I}}(\underline{u})), G(V_{\text{S}}(\underline{u})))) \in [0,1] \times [0,1]$ compatible with $\mathbb{P}(u_n^{a_n} = \overline{u})$ as a model prediction for some $\gamma \in (0,1]$. Equivalently, the identified set $\Lambda^{\text{unk}}_{F_{\text{I}},F_{\text{S}}}$ is the projection of Λ^{unk} along its fist and second dimensions. Moreover, note that the projection of Λ^{unk} along its first and second dimensions for a fixed value of γ, say $\hat{\gamma}$, corresponds to $\Lambda^{\text{off}}(\hat{\gamma})$.

Replacing $\mathbb{P}(u_n^{a_n} = \overline{u})$ with its sample analog $\overline{u}_N$ in the definition of identified sets Λ^{unk} and $\Lambda^{\text{unk}}_{F_{\text{I}},F_{\text{S}}}$ above, we obtain the corresponding set estimators. By the convergence in (12), such set estimators almost surely contain the true parameters of interest as $N \to \infty$.

5 Conclusion

Motivated by the overwhelming evidence that agents rely on others' choices and experiences to shape their behavior and beliefs, in this paper, we make a first step to understanding how social information affects the identification of search models.

We provide two main sets of results. First, we illustrate how neglecting social information leads to non-identification and biased estimation of search cost distributions. We connect the sign of the resulting estimation bias to how social information changes agents' optimal search and, hence, the distributions of the observable outcomes identifying the search model. These results highlight the non-trivial role that social information can play in empirical search models and its potential to undermine the quantification of search frictions under the common assumption that all agents act in isolation.

Second, we propose empirical strategies to restore identification and correct estimation. To begin with, we show how to construct robust bounds on the search cost distribution when agents' amount and type of social information remain unobserved and only minimal assumptions on this primitive are justifiable. Next, we illustrate how to recover more informative estimates by relying on additional data or stronger assumptions.

We present our results—both the negative and the positive ones—within a unified framework. To do so, we make some modeling choices that may not always reflect all relevant features of specific empirical applications. However, we discuss how our insights are robust to various generalizations of our leading model(s). Moreover, the remedies we outline give general insights into the data requirements or the (robust) assumptions that may help identify search models under social information. We provide empirical researchers with a coherent theoretical framework that guides how to account for social information in search models. We leave the task of tailoring the model to capture other relevant features of specific applications of interest to future empirical work.

References

Bailey, M., D. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2022): "Peer Effects in Product Adoption," *American Economic Journal: Applied Economics*, 14, 488–526.

Bergemann, D., B. Brooks, and S. Morris (2022): "Counterfactuals with Latent Information," *The American Economic Review*, 112, 343–368.

Bergemann, D. and S. Morris (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

Bikhchandani, S., D. Hirshleifer, O. Tamuz, and I. Welch (2022): "Information Cascades and Social Learning," *Working Paper.*

Cai, H., Y. Chen, and H. Fang (2009): "Observational Learning: Evidence From a Randomized Natural Field Experiment," *American Economic Review*, 99, 864–882.

Canen, N. and K. Song (2023): "A Decomposition Approach to Counterfactual Analysis in Game-Theoretic Models," *Working Paper.*

Chambers, C. P. and F. Echenique (2016): *Revealed Preference Theory*, Cambridge University Press.

De Oliveira, H. and R. Lamba (2023): "Rationalizing Dynamic Choices," *Working Paper.*

Deb, R. and L. Renou (2021): "Dynamic Choice and Common Learning," *Working Paper.*

Doval, L. and J. C. Ely (2020): "Sequential Information Design," *Econometrica*, 88, 2575–2608.

Forbes (2012): "Are Brands Wielding More Influence In Social Media Than We Thought?" .

——— (2022): "Studies Predict Social Media Will Be Biggest Influence On Shopping For Decades," .

Galeotti, A. (2010): "Talking, Searching, and Pricing," *International Economic Review*, 51, 1159–1174.

Garcia, D. and S. Shelegia (2018): "Consumer Search with Observational Learning," *The RAND Journal of Economics*, 49, 224–253.

Gergatsouli, E. and C. Tzamos (2023): "Weitzman's Rule for Pandora's Box with Correlations," *Working Paper.*

Golub, B. and E. Sadler (2016): "Learning in Social Networks," in *The Oxford Handbook of the Economics of Network*, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers, Oxford University Press.

Gualdani, C. and S. Sinha (2023): "Identification and Inference in Discrete Choice Models with Imperfect Information," *Working Paper.*

Heidhues, P. and P. Strack (2021): "Identifying Present Bias From the Timing of Choices," *The American Economic Review*, 111, 2594–2622.

Hendricks, K., A. Sorensen, and T. Wiseman (2012): "Observational Learning and Demand for Search Goods," *American Economic Journal: Microeconomics*, 4, 1–31.

Heumann, T. (2019): "Informationally Robust Comparative Statics in Incomplete Information Games," *Working Paper.*

Honka, E., A. Hortaçsu, and M. Wildenbeest (2019): "Empirical Search and Consideration Sets," in *Handbook of the Economics of Marketing*, Elsevier, vol. 1, 193–257.

Kang, Z. Y. and S. Vasserman (2022): "Robust Bounds for Welfare Analysis," Tech. rep., Working Paper.

Kircher, P. and A. Postlewaite (2008): "Strategic Firms and Endogenous Consumer Emulation," *The Quarterly Journal of Economics*, 123, 621–661.

Kline, B. and E. Tamer (2023): "Recent Developments in Partial Identification," *Annual Review of Economics*, 15, 125–150.

Libgober, J. (2023): "Identifying Wisdom (of the Crowd): A Regression Approach," *Working Paper.*

Liu, S. and N. Netzer (2023): "Happy Times: Measuring Happiness Using Response Times," *The American Economic Review (forthcoming).*

Lomys, N. (2023): "Collective Search in Networks," *Working Paper.*

Magnolfi, L. and C. Roncoroni (2022): "Estimation of Discrete Games with Weak Assumptions on Information," *The Review of Economic Studies (forthcoming)*, 90, 2006–2041.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*, vol. 5, Springer.

Mobius, M. and T. Rosenblat (2014): "Social Learning in Economics," *Annual Review of Economics*, 6, 827–847.

Molinari, F. (2020): "Microeconometrics with Partial Identification," *Handbook of Econometrics*, 7, 355–486.

Moretti, E. (2011): "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales," *The Review of Economic Studies*, 78, 356–393.

Mueller-Frank, M. and M. M. Pai (2016): "Social Learning with Costly Search," *American Economic Journal: Microeconomics*, 8, 83–109.

Shmaya, E. and L. Yariv (2016): "Experiments on Decisions under Uncertainty: A Theoretical Framework," *The American Economic Review*, 106, 1775–1801.

Stigler, G. J. (1961): "The Economics of Information," *Journal of Political Economy*, 69, 213–225.

Syrgkanis, V., E. Tamer, and J. Ziani (2021): "Inference on Auctions with Weak Assumptions on Information," *Working Paper.*

Ursu, R., S. Seiler, and E. Honka (2023): "The Sequential Search Model: A Framework for Empirical Research," *Working Paper.*

Weitzman, M. L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 43, 641–654.