Econometrica, Vol. 89, No. 6 (November, 2021), 2601–2635

LEARNING DYNAMICS IN SOCIAL NETWORKS

SIMON BOARD
UCLA

MORITZ MEYER-TER-VEHN
UCLA

This paper proposes a tractable model of Bayesian learning on large random networks where agents choose whether to adopt an innovation. We study the impact of the network structure on learning dynamics and product diffusion. In directed networks, all direct and indirect links contribute to agents' learning. In comparison, learning and welfare are lower in undirected networks and networks with cliques. In a rich class of networks, behavior is described by a small number of differential equations, making the model useful for empirical work.

KEYWORDS: Networks, diffusion, social learning.

1. INTRODUCTION

HOW DO COMMUNITIES, ORGANIZATIONS, OR ENTIRE SOCIETIES learn about innovations? Consider consumers learning about a new brand of electric car from friends, farmers learning about a novel crop from neighbors, or entrepreneurs learning about a source of finance from nearby businesses. In all these instances, agents learn from other's choices, so the diffusion of the innovation depends on the social network. One would like to know: Do agents learn more quickly in a highly connected network? Do products diffuse faster in a more clustered network?

This paper proposes a tractable Bayesian model to answer these questions, and characterizes the diffusion of innovation in the social network via a system of differential equations. In contrast to most papers in the literature (e.g., Acemoglu et al. (2011)), our results speak to learning dynamics at each point in time, rather than focusing on long-run behavior. We thus recover the tractability of the reduced-form models of diffusion (e.g., Bass (1969)) in a model of Bayesian learning. Understanding the entire dynamics is important because empirical researchers must identify economic models from finite data, and because in practice, governments and firms care about *when* innovations take off, not just *if* they take off.

Our paper has two contributions. First, we describe how learning dynamics depend on the network structure. For large random networks, we show that an agent typically benefits when her neighbors have more links. However, additional links that correlate the information of an agent's neighbors or create feedback loops can muddle her learning.

Simon Board: sboard@econ.ucla.edu
Moritz Meyer-ter-Vehn: mtv@econ.ucla.edu

We have received useful comments from Ben Golub, Niccolo Lomys, Alessandro Pavan, Andy Postlewaite, Klaus Schmidt, Bruno Strulovici, Fernando Vega-Redondo, Chen Zhao, and especially Evan Sadler. Hersh Chopra provided excellent research assistance. We also thank seminar audiences at Bar-Ilan, Bocconi, Bonn, CEMFI, Chapman, Chicago Theory Conference, Columbia Theory Conference, Conference on Network Science and Economics, Duke, Hebrew, HKU, Indiana, Mannheim, Michigan, Munich, NASM, Northwestern, NYU, Oxford, Penn State, Pittsburgh, Princeton, Purdue, Seoul National, Tel-Aviv, Toronto, Toulouse, UC3M, UCLA, UCSD, Utah, Warwick, Yale. Moritz Meyer-ter-Vehn would like to thank the CESifo and ECONtribute, whose hospitality he enjoyed while working on this research.

© 2021 The Econometric Society https://doi.org/10.3982/ECTA18659

FIGURE 1.—**Illustrative networks.** The **left** panel shows an Erdős–Rényi network; social learning is described by a one-dimensional ODE (13). The **middle** panel shows a random network with triangles; social learning is described by a two-dimensional ODE (18)–(19). The **right** panel shows a random network with homophily; social learning is described by a four-dimensional ODE (42).

For example, welfare is lower in a "clustered" network than in a random "bilateral" network with the same degree distribution. These results can help us understand how the diffusion of products and ideas changes with the introduction of social media, differs between cities and villages, and is affected by government programs that form new social links (e.g., Cai and Szeidl (2018)).

Our second contribution is methodological. The complexity of Bayesian updating means that applied and empirical papers typically study heuristic behavior on the exact network (e.g., Golub and Jackson (2012), Banerjee et al. (2013)). In comparison, we take a "macroeconomic approach" by studying equilibrium behavior on an approximate network. Figure 1 illustrates random networks exhibiting cliques and homophily that we can analyze with low-dimensional ODEs. Given a "real life" network, one can then study diffusion and learning on an approximate network with the same network statistics (e.g., agents' types, degree distributions, cluster coefficients).

In the model, agents are connected via an exogenous network. They may know the entire network (our "deterministic" networks) or only their local neighborhood (our "random" networks). An agent "enters" at a random time and considers a product whose quality is high or low. For example, a driver's car breaks down, and she contemplates buying a new brand of electric car. The agent observes which of her neighbors has adopted the product and chooses whether to inspect the product at a cost (e.g., via a test drive). Inspection perfectly reveals the common quality, and the agent adopts the product if its quality is high.

The agent learns directly from her neighbors via their adoption decisions; she also learns indirectly from further removed agents as their adoption decisions influence her neighbors' inspection (and adoption) decisions. The agent's own inspection decision is thus based on the hypothesized inspection decisions of her neighbors, which collectively generate her *social learning curve* (formally, the probability that at least one of her neighbors adopts a high-quality product as a function of time). In turn, her adoption decision feeds into the social learning curves of her neighbors.

In Section 2, we characterize agents' adoption decisions via a system of ordinary differential equations (ODEs). We start with some simple deterministic examples (e.g., chains, complete networks) that can be characterized by one-dimensional ODEs. These provide intuition and serve as building blocks for our large random networks. For general networks, the dimension of this system is exponential in the number of agents, since one must keep track of the joint adoption probabilities; for example, if two agents have a neighbor in common, their adoption decisions are correlated.

In Section 3, we turn to large random networks, where agents know their neighbors but not their neighbors' neighbors. Such incomplete information is both realistic and simplifies the analysis: the less agents know, the less they can condition on, and the simpler their behavior. Formally, we model network formation via the configuration model: Agents draw link-stubs that we randomly connect in pairs or triples. In the limit economy with infinitely many agents, we characterize adoption behavior both for directed networks with multiple types of agents (e.g., Twitter) and undirected networks with cliques (e.g., Facebook) in terms of a low-dimensional system of ODEs.[1] Intuitively, large networks locally resemble trees of elemental networks (e.g., links, cliques), where information outside an element is independent. In such trees, it suffices to keep track of adoption within each element, ignoring any correlation. We validate our analysis by showing that equilibrium behavior in large finite networks converges to the solution of these ODEs.

The ODEs allow for sharp comparative statics of social learning as a function of the network structure. Given a mild condition on the hazard rate of inspection costs, an agent's adoption rate rises in her social information. Therefore, more neighbors lead to more adoption, which leads to more information, which leads to more adoption, and so on. Thus, an agent benefits from both direct and indirect links. However, not all links are equally beneficial. We show that learning is superior in a bilateral network than in a clustered network with the same degree distribution. Intuitively, if i's neighbors j and k observe one another, then j's lack of adoption makes k more pessimistic and raises the probability that neither of them adopts the product. We also show that learning is superior in a directed network than in an undirected network with the same degree distribution. Intuitively, i's neighbor j cannot see i adopt prior to the time i enters; thus the backward link $j \to i$ makes j more pessimistic, lowering his adoption and i's information, precisely when i needs to make a decision.

Finally, we connect our theory to prominent themes in the literature on learning in networks. First, we extend the model to allow for correlation neglect (e.g., Eyster and Rabin (2014)) and show that it reduces learning and welfare. Intuitively, agent i's misspecification causes her to overestimate the chance of observing an adoption, and means she grows overly pessimistic when none of her neighbors adopt; this reduces i's adoption and other agents' social information. Second, we reconsider the classic question of information aggregation (e.g., Smith and Sørensen (1996), Acemoglu et al. (2011)) by letting the network and average degree grow large. When the network remains sparse, agents aggregate information perfectly; yet when it becomes clustered, information aggregation may fail. Thus, adding links may lower social welfare.

1.1. *Literature*

The literature on observational learning originates with the classic papers of Banerjee (1992) and Bikhchandani et al. (1992). In these models, agents observe both a private signal and the actions of all prior agents before making their decision. Smith and Sørensen (2000) show that "asymptotic learning" arises if the likelihood ratio of signals is unbounded. Smith and Sørensen (1996) and Acemoglu et al. (2011) dispense with the assumption that agents observe all prior agents' actions, and interpret the resulting observation structure as a social network. The latter paper generalizes Smith and Sørensen's (2000) asymptotic learning result to the case where agents are (indirectly) connected to

[1]The directed network with multiple types has one ODE per type. The undirected network with cliques has one ODE per type of link (i.e., bilateral links and triangles).

an unbounded number of other agents. Subsequent papers quantify the amount of information aggregation when signals are bounded (e.g., Monzón and Rapp (2014), Lobel and Sadler (2015)).

Our model departs from these papers in two ways. First, we study agents' choice of whether to *inspect* the good given their social information. In many applications, such as buying an electric car, it seems natural that agents have to acquire information about the good (via test drives and reviews) rather than being born with it. A few recent papers consider models of this flavor. Assuming agents observe all predecessors, Mueller-Frank and Pai (2016) and Ali (2018) show asymptotic learning is perfect if experimentation costs are unbounded below. Lomys (2020) reaches the same conclusion in a network setting if the network is sufficiently connected. Like Mueller-Frank and Pai (2016) and Lomys (2020), we assume that inspection reveals quality perfectly.

Second, we study product *adoption* and assume that agents observe only the adoption decisions of their neighbors, but not their entry times or inspection decisions. This assumption is consistent with the classic observational learning model, and is reasonable in many applications. For example, the above driver observes whether any neighbors drive the new electric car, but is unlikely to interrogate neighbors who drive alternative cars why they did not adopt the new innovation. A number of papers have analyzed related problems in complete networks. Guarino et al. (2011) suppose an agent sees how many others have adopted the product, but not the timing of others' actions or even her own action. Herrera and Hörner (2013) suppose an agent observes who adopted and when they did so, but not who refrained from adopting. Hendricks et al. (2012) suppose an agent knows the order in which others move, but sees only the total number of adoptions; as in our model, the agent then uses this public information to acquire information before making her purchasing decision. These papers characterize asymptotic behavior and find an asymmetry in social learning: good products may fail but bad products cannot succeed. In Section 4.1, we show a similar result applies in our setting.[2]

These two complementary modeling choices allow us to characterize equilibria in terms of simple social learning curves. Agents enter without private information, but learn quality perfectly upon inspection. This cleanly separates the role of social information (which determines the inspection decision) and private information (which determines the adoption decision). Additionally, agents do not observe when their neighbors move and never see neighbors adopting low-quality goods. This further simplifies the agent's problem to a binary decision at a single information set: Should an agent inspect when no neighbor has yet adopted?

Our contribution over the prior literature then lies in the questions we address. Traditionally, herding papers ask whether society correctly aggregates information as the number of agents grows. In their survey of observational learning models, Golub and Sadler (2016) write: "A significant gap in our knowledge concerns short-run dynamics [...] The complexity of Bayesian updating in a network makes this difficult, but even limited results would offer a valuable contribution to the literature." In this paper we characterize such

[2]There is a wider literature on product adoption without learning. There are "awareness" models where agents become aware of the product when their neighbors adopt it. One can view Bass (1969) as such a model with random matching; Campbell (2013) studies diffusion on a fixed network. There are also models of "local network goods" where agents want to adopt the product if enough of their neighbors adopt. Morris (2000) characterizes stable points in such a game. Sadler (2020) puts these forces together and studies diffusion of a network good where agents become aware of it from their neighbors. Banerjee (1993) and McAdams and Song (2020) integrate awareness and social learning, allowing people to infer a good's quality from the time at which they become aware of it.

"short-run" learning dynamics in rich classes of social networks that allow for homophily, clustering, and arbitrary degree distributions. We then study how an agent's information and welfare vary with the network structure.[3]

2. MODEL

The Network. A finite set of I agents is connected via an exogenous, directed network $G \subseteq I \times I$ that represents who observes whom. If i (she) observes j (he), we write $i \to j$ or $(i, j) \in G$, say i is *linked to* j, and call j a *neighbor* of i. The set of i's neighbors is $N_i(G)$. Agents may have incomplete information about the network. We capture such information via finite *signals* $\xi_i \in \Xi_i$ and a joint prior distribution over networks and signal profiles $\mu(G, \xi)$. A random network is given by $\mathcal{G} = (I, \Xi, \mu)$. To be more concrete, we consider several special cases:

- Deterministic network G. Signal spaces are degenerate, $|\Xi_i| = 1$, and the prior μ assigns probability one to G. While complete information might seem to simplify matters, in fact learning dynamics become very complicated once we move beyond the simplest networks G; this motivates us to study random networks with incomplete information.
- Directed configuration model with finite *types* $\theta \in \Theta$. Agents draw types θ and random stubs for each type θ'. We then randomly connect the type θ' stubs to type θ' agents. Agents know how many outlinks of each type they have. For example, Twitter users know what kind of other users they follow. We study this model in Section 3.1.
- Undirected configuration model with binary links and triangles. Agents draw $\bar{d}$ binary stubs and $\hat{d}$ pairs of triangle stubs. We then randomly connect binary stubs in pairs, and pairs of triangle stubs in triples. Agents know how many binary and triangle links they have. For example, consider groups of friends linked on Facebook. We study this model in Sections 3.2–3.5.

The Game. The agents seek to learn about the quality of a single product of quality $q \in \{L, H\} = \{0, 1\}$. Time is continuous, $t \in [0, 1]$. At time $t = 0$, agents share a common prior $\Pr(q = H) = \pi_0 \in (0, 1)$, independent of network G and signals ξ.

Agent i develops a need for the product, or *enters*, at time $t_i \sim U[0, 1]$.[4] She observes which of her neighbors have adopted the product and updates her belief about product quality to π_i. The agent can then inspect the product at cost $\kappa_i \sim F[\underline{\kappa}, \bar{\kappa}]$, with bounded pdf f. If she inspects the product, she observes its quality and adopts it iff $q = H$. If the agent does not inspect, she can either pass on the product or adopt it blindly, without inspection. Entry times t_i and inspection costs κ_i are private information, independent within agents, and i.i.d. across agents. All of these are independent of product quality q, the network G, and agents' signals ξ.

Adopting the product yields utility 1 if quality is high and $-M$ if quality is low, whereas nonadoption yields utility 0. Behavior is then as follows: If agent i sees a neighbor adopt, her posterior is $\pi_i = 1$ and she adopts blindly. If she sees no adoption, her posterior is $\pi_i \leq \pi_0$. We assume $M \geq \pi_0/(1 - \pi_0)$, so adopting blindly is dominated in this case;

[3] A different approach is to look at the rate at which agents' beliefs converge. For example, Hann-Caruthers, Martynov, and Tamuz (2018) compare the cases of "observable signals" and "observable actions" in the classic herding model of Bikhchandani et al. (1992).

[4] The uniform distribution is a normalization: t_i should not be interpreted as calendar time, but rather as time-quantile in the product life cycle.

the agent thus inspects if $\kappa_i \leq \pi_i$ and otherwise passes on the product. We assume that some cost-types inspect at the prior belief, $\underline{\kappa} < \pi_0$. Our solution concept is Bayesian Nash equilibrium.

Remarks. As discussed in Section 1.1, the two salient aspects of our social learning model are inspection and adoption. Having agents enter without private information and learn quality perfectly upon inspection cleanly separates the role of social and private information. The adoption aspect makes learning asymmetric. If agent i sees that j has adopted, she knows that quality is high. Conversely, if j has not adopted, this may be because (i) he has yet to develop a need for the product, (ii) he developed a need but chose not to inspect, or (iii) he inspected and quality is low.

Additionally, we assume that agents adopt the product iff quality is high. A single adoption is thus proof of high quality and induces agents to adopt blindly; conversely, agents who observe no adoption get more pessimistic and either inspect the product or pass on it. Jointly, these assumptions reduce agent i's problem to a binary decision at a single information set: Whether or not to inspect if none of her neighbors have adopted by t_i. The analysis is unchanged if, due to idiosyncratic preferences, agents adopt high-quality products with probability $\alpha^H < 1$. In Section 4.2, we discuss a model variant where agents also adopt low-quality products with probability $\alpha^L > 0$.

2.1. *Examples: Directed Networks*

The next two examples of directed trees (i.e., networks where any two agents i, j are connected by at most one path $i \to \cdots \to j$) illustrate agents' inference problem.

EXAMPLE 1—Directed Pair $i \to j$: Suppose there are two agents, Iris and John. John has no social information, while Iris observes John. Let $x_{j,t}$ be the probability that John adopts product H by time t.[5] He enters uniformly over $t \in [0, 1]$, and so the time-derivative $\dot{x}_{j,t}$ equals the probability he adopts conditional on entering at time t. Since he inspects iff $\kappa_j \leq \pi_0$ and then always adopts product H, we have $\dot{x}_j = \Pr(j \text{ adopt}) = \Pr(j \text{ inspect}) = F(\pi_0)$, where we drop the time subscript.

Iris, in turn, learns from John's adoption. We thus interpret John's adoption curve x_j as Iris's social learning curve. If John has adopted, Iris infers that quality is high and also adopts. Conversely, if John has not adopted, Iris's posterior that quality is high is given by Bayes' rule,

$$\pi(x_j) := \frac{(1 - x_j)\pi_0}{(1 - x_j)\pi_0 + (1 - \pi_0)}. \tag{1}$$

Iris inspects if $\kappa_i \leq \pi(x_j)$. As x_j rises over time, Iris becomes more pessimistic when John does not adopt. All told, Iris's adoption rate equals

$$\begin{aligned}
\dot{x}_i &= 1 - \Pr(i \text{ not adopt}) \\
&= 1 - \Pr(j \text{ not adopt}) \times \Pr(i \text{ not inspect} | j \text{ not adopt}) \\
&= 1 - (1 - x_j)\big(1 - F\big(\pi(x_j)\big)\big) =: \Phi(x_j).
\end{aligned} \tag{2}$$

[5]Since no agent adopts when $q = L$, it suffices to keep track of the adoption probability conditional on $q = H$.

The function Φ, which maps i's social information x_j to her own adoption $\dot{x}_i$, plays a central role throughout the paper.

EXAMPLE 2—Directed Chain: Suppose there is an infinite chain of agents, so Kata observes Lili, who observes Moritz, and so on ad infinitum.[6] Analogous to equation (2), adoption in the symmetric equilibrium is governed by the ODE

$$\dot{x} = \Phi(x). \tag{3}$$

This captures the idea that Kata's decision takes into account Lili's decision, which takes into account Moritz's decision, and so on. The simplicity of the adoption curve is in stark contrast to the cyclical behavior seen in traditional herding models when agents observe only their immediate predecessors (Celen and Kariv (2004)).

2.2. *General Networks*

We now turn to the analysis of general random networks $\mathcal{G} = (I, \Xi, \mu)$. We first study agents' adoption rate and their social learning curves, and show that adoption rises with social information under a bounded hazard rate assumption. We then close the model and establish that it admits a unique equilibrium.

We start with some definitions. As in the examples of Section 2.1, we generally denote agent i's probability of adopting product H by x_i. Agent i may not know the network, and so we keep track of agents' adoption across realizations of G and signals ξ. Let $x_{i,G,\xi}$ be agent i's *realized adoption curve*, given (G, ξ) after taking expectations over others' entry times t_j and cost draws κ_j. Taking expectation over (G, ξ_{-i}), let $x_{i,\xi_i} := \sum_{G,\xi_{-i}} \mu(G, \xi_{-i}|\xi_i) x_{i,G,\xi}$ be i's *interim adoption curve* given her signal ξ_i. Throughout, we drop the time subscript t for these and all other curves.

Bayesian agents form beliefs over their neighbors' adoption decisions. Since a single adoption by one of i's neighbors perfectly reveals high quality, she only keeps track of this event. Specifically, let $y_{i,G,\xi}$ be the probability that *at least one* of i's neighbors adopts product H by time $t \le t_i$ in network G given signals ξ, and $y_{i,\xi_i} := \sum_{G,\xi_{-i}} \mu(G, \xi_{-i}|\xi_i) y_{i,G,\xi}$ be the expectation conditional on ξ_i.

To solve for i's realized adoption curve $x_{i,G,\xi}$, consider two cases. If she sees one of her neighbors adopt, she updates her belief to $\pi_i = 1$ and adopts blindly. Conversely, if she sees no adoption, she updates her belief to $\pi_i = \pi(y_{i,\xi_i}) \le \pi_0$ and inspects and adopts iff her inspection cost is below this cutoff, $\kappa_i \le c_{i,\xi_i} := \pi_i$. Analogous to equation (2), i's realized adoption curve follows:

$$\dot{x}_{i,G,\xi} = 1 - (1 - y_{i,G,\xi})\big(1 - F\big(\pi(y_{i,\xi_i})\big)\big) =: \phi(y_{i,G,\xi}, y_{i,\xi_i}). \tag{4}$$

Note that equation (4) depends on both the realized and the interim adoption probability of i's neighbors, $y_{i,G,\xi}$ and y_{i,ξ_i}, respectively. The former determines whether i actually observes an adoption, given (G, ξ); the latter determines i's posterior belief when none of her neighbors adopt, which depends only on i's coarser information ξ_i. Taking expectations over (G, ξ_{-i}) given ξ_i, agent i's interim adoption curve is then

$$\dot{x}_{i,\xi_i} = 1 - (1 - y_{i,\xi_i})\big(1 - F\big(\pi(y_{i,\xi_i})\big)\big) = \phi(y_{i,\xi_i}, y_{i,\xi_i}) = \Phi(y_{i,\xi_i}). \tag{5}$$

[6]In Section 3.1, we interpret this infinite network as a limit of the finite networks in our model.

Equation (5) captures our positive implications for the diffusion of new products.

Our primary results concern normative implications, quantifying the value of social learning, as captured by i's *social learning curve* y_{i,ξ_i}. To see that this curve indeed measures i's learning, observe that the probability that i sees an adoption is given by

	≥ 1 adopt	0 adopt
$q = H$	y_{i,ξ_i}	$1 - y_{i,\xi_i}$
$q = L$	0	1

If $y_{i,\xi_i} = 1$, then agent i has perfect information about the state; if $y_{i,\xi_i} < 1$, she has effectively lost the signal with probability $1 - y_{i,\xi_i}$. A rise in y_{i,ξ_i} thus Blackwell-improves her social information, and thereby her expected utility.[7]

Clearly, i's social information improves over time: Since adoption is irreversible, $y_{i,G,\xi}$ rises in t for every G, and hence also in expectation. Much of our paper compares social learning curves across networks. For networks $\tilde{\mathcal{G}}$ and $\mathcal{G}$ (with overlapping agents i and types ξ_i), we write $\tilde{y}_{i,\xi_i} \geq y_{i,\xi_i}$ if social information is greater in $\tilde{\mathcal{G}}$ for all t, and $\tilde{y}_{i,\xi_i} > y_{i,\xi_i}$ if it is strictly greater for all $t > 0$.

Social learning and adoption are linked by (5). One would think that as i collects more information, her adoption of product H increases. Indeed, with perfect information she always adopts. More generally, monotonicity requires an assumption.

ASSUMPTION 1: *The distribution of costs has a* bounded hazard rate (*BHR*) *if*

$$\frac{f(\kappa)}{1 - F(\kappa)} \leq \frac{1}{\kappa(1-\kappa)} \quad \textit{for } \kappa \in [0, \pi_0]. \tag{6}$$

LEMMA 1: *If F has a bounded hazard rate*, (6), *then i's interim adoption probability x_{i,ξ_i} increases in her information y_{i,ξ_i}.*

PROOF: Differentiating (5),

$$\begin{aligned}\Phi'(y) &= \partial_1 \phi(y, y) + \partial_2 \phi(y, y) = 1 - F\big(\pi(y)\big) + (1-y) \cdot \pi'(y) \cdot f\big(\pi(y)\big) \\ &= 1 - F\big(\pi(y)\big) - \pi(y) \cdot \big(1 - \pi(y)\big) \cdot f\big(\pi(y)\big),\end{aligned} \tag{7}$$

where the second equality uses Bayes' rule (1) to show

$$(1-y) \cdot \pi'(y) = -(1-y) \frac{\pi_0(1-\pi_0)}{(1-y\pi_0)^2} = -\frac{(1-y)\pi_0}{1-y\pi_0} \frac{1-\pi_0}{1-y\pi_0} = -\pi(y) \cdot \big(1-\pi(y)\big).$$

Equation (7) captures two countervailing effects: Its first term is positive because i adopts blindly if she observes an adoption. The second term is negative because an increase in y makes i is more pessimistic when she sees no adoption. The aggregate effect is positive iff BHR holds at $\kappa = \pi(y)$. Thus if BHR holds for all $\kappa \in [0, \pi_0]$, then $\Phi'(y) \geq 0$ for all $y \in [0, 1]$, and better information y_{i,ξ_i} means higher slope $\dot{x}_{i,\xi_i}$ and level x_{i,ξ_i}. *Q.E.D.*

[7] Since agent i never adopts the low-quality product, her welfare cost consists of failing to adopt the high-quality product and paying to inspect the low-quality product.

For an intuition, recall that adoption probabilities x condition on high quality, $q = H$. Agent i's expected posterior belief $E[\pi_i|H]$ thus exceeds the prior π_0 and increases with her information. Whether i's adoption probability $E[F(\pi_i)|H]$ also increases in her information depends on the curvature of the cost distribution F; this is guaranteed by BHR. In turn, BHR is satisfied if f is increasing on $[0, 1]$, which includes $\kappa \sim U[0, 1]$ as a special case.[8]

To see how Lemma 1 relies on BHR, assume the distribution $F(\kappa)$ has support $[0, \pi_0]$; this violates BHR since the denominator in the left-hand side of (6), $1 - F(\pi_0)$, vanishes. Without social information, agent i adopts product H with probability 1. With social information, agent i adopts with probability 1 if some neighbor adopts, and below 1 if no neighbor adopts, since $\pi(y_{i,\xi_i}) < \pi_0$. In expectation, social information lowers agent i's adoption rate, contradicting Lemma 1.

Our main results, Theorems 1–4, compare social learning curves (and thus welfare) across networks. Lemma 1 means that if we assume BHR, then the same comparisons apply to adoption rates.[9]

Lastly, we close the model in equilibrium.

PROPOSITION 1: *In any random network $\mathcal{G}$, there exists a unique equilibrium.*

The challenge with proving Proposition 1 in Appendix A.1 is to keep track of the adoption probability y_i of i's neighbors on the right-hand side of (4). There are two issues. First, the *self-reflection* problem: If i and j observe each other, then when i enters, she knows that j cannot have seen her adopt, which j interprets as bad news. Second, the *correlation* problem: when i's neighbors j, k observe each other or share sources of information, their adoption is correlated. The directed tree networks in Examples 1 and 2 abstract from these problems. The next section considers complete networks where we see the effects of self-reflection and correlation. Beyond such simple examples, the differential equation governing equilibrium must keep track of agents' joint adoption probabilities, whose dimensionality grows exponentially with the number of agents. For this reason, Section 3 considers large, random networks, where we recover simple formulas for diffusion curves akin to the examples.

2.3. *Examples: Undirected Networks*

Examples 1 and 2 illustrated social learning in directed tree networks. The next two examples prepare our analysis of undirected networks and networks with cliques in Sections 3.2 and 3.3, respectively.

EXAMPLE 3—Undirected Pair $i \leftrightarrow j$: Agent i's social learning curve equals i's expectation of j's adoption curve at $t \le t_i$; for convenience we denote this by $\bar{x}_j$. To solve for $\bar{x}_j$, we must distinguish between two probability assessments of the event that j observes i adopt. From i's "objective" perspective, this probability equals 0 since i knows she has not entered at $t \le t_i$. From j's "subjective" perspective, the probability equals $\bar{x}_i$, since j

[8]An increasing density guarantees $f(\kappa) \le \frac{\int_\kappa^1 f(z)dz}{1-\kappa} \le \frac{1-F(\kappa)}{\kappa(1-\kappa)}$. For other densities f, BHR is automatically satisfied when $\kappa \approx 0$ since the RHS increases to infinity. For higher costs, BHR states that the density does not decrease too quickly, $d \log f(\kappa)/d\kappa \ge -2/\kappa$. In particular, BHR holds with equality at all κ if $f(\kappa) \propto 1/\kappa^2$.

[9]As a corollary, BHR implies that interim adoption curves x_{i,ξ_i} are convex in time since social information y_{i,ξ_i} increases.

thinks he is learning from agent i given $t \leq t_j$. These objective and subjective probabilities play the role of the "realized" and "interim" probabilities in equation (4), respectively. Thus,

$$\dot{\bar{x}}_j = \phi(0, \bar{x}_i) = F\big(\pi(\bar{x}_i)\big). \tag{8}$$

By symmetry, $\bar{x}_i = \bar{x}_j =: \bar{x}$, reducing (8) to a one-dimensional ODE. The actual (unconditional) adoption probability follows $\dot{x} = \Phi(\bar{x})$.

EXAMPLE 4—Complete Network: More generally, consider the complete network of $I + 1$ agents. When $I = 1$, this is equivalent to the undirected pair. With more agents, agent j's adoption before i enters depends on agent k's adoption before both i and j enter. One might worry about higher-order beliefs as I gets large. Fortunately, we can side-step this complication by thinking about the game from the first mover's perspective, before anyone else has adopted.

To be specific, let the *first adopter probability* $\hat{x}$ be the probability an agent adopts given that no one else has yet adopted. Since everyone is symmetric, intuition suggests that the first adopter attaches subjective probability $(1 - \hat{x})^I$ to the event that none of the other potential first adopters has adopted. By definition, the first adopter observes no adoption herself, and so we define $\hat{x}$ as the solution of

$$\dot{\hat{x}} = \phi\big(0, 1 - (1 - \hat{x})^I\big) = F\big(\pi\big(1 - (1 - \hat{x})^I\big)\big), \tag{9}$$

generalizing equation (8). We prove the following in Appendix A.2.

LEMMA 2: *In the complete network with $I + 1$ agents, any agent's social learning curve is $1 - (1 - \hat{x})^I$; the adoption probability follows $\dot{x} = \Phi(1 - (1 - \hat{x})^I)$.*

3. LARGE RANDOM NETWORKS

In this section, we characterize equilibria in large random networks, such as those in Figure 1. This allows us to compare social learning curves across networks, and assess the impact of more links and clustering. Moreover, these networks are sufficiently rich to resemble reality and be used for empirical applications.

Formally, we introduce two network configuration models. In Section 3.1, we study directed networks with multiple types; this nests deterministic directed trees (Examples 1 and 2) and directed stochastic block networks as special cases. We study their limit equilibria as the networks grow large, ensuring the network locally resembles a tree. In such a tree, neighbors' adoption decisions are mutually independent, thus eliminating the correlation and reflection effects. We characterize equilibrium via simple ODEs and show that more direct or indirect links improve learning.

Next, we study undirected networks with cliques. To aid intuition, we build the model up in steps. In Section 3.2, we consider undirected random networks, such as Erdős–Rényi, where neighbors' decisions are mutually independent. We show that agents learn less than in a directed network with the same degree distribution. In Section 3.3, we study networks of cliques, where i's neighbors' decisions are correlated within cliques but independent across cliques. We show that agents learn less than in an undirected random network with the same degree distribution. In Section 3.4, we study correlation neglect, where agents are connected via a network of cliques but believe they are in a network of independent, bilateral links. Such correlation neglect reduces social learning. Finally, in Section 3.5 we justify the heuristic approach taken in the prior sections by nesting these models in a general model and confirming that the identified strategies indeed constitute limit equilibria.

Taken together, this analysis suggests a "macroeconomic" approach to studying diffusion empirically: First, calibrate a random network to the real-life network by matching the pertinent network parameters (agents' types, degree distributions, cluster coefficients). Then solve for equilibrium behavior on this approximate network. This contrasts with the "microeconomic" approach typically used, where one studies a behavioral heuristic on the actual network.

3.1. *Directed Networks With Multiple Types*

We first consider diffusion in large directed networks with different types of agents. For example, think of Twitter users as celebrities, posters, and lurkers: Celebrities only follow celebrities, while posters and lurkers follow posters and celebrities. Agents know who they follow and know the distribution over networks, but do not know exactly who others follow.

To formalize this idea, we generate the random network $\mathcal{G}_I$ via the configuration model (e.g., Jackson (2010), Sadler (2020)). For any agent i, independently draw a finite type $\theta \in \Theta$ according to some distribution with full support. For any agent with type θ, independently draw a vector of labeled outlinks $d = (d_{\theta'})_{\theta'} \in \mathbb{N}^{\Theta}$; these are realizations of a random vector $D_\theta = (D_{\theta,\theta'})_{\theta'}$ with finite expectation $E[D_{\theta,\theta'}]$. We call $D = (D_{\theta,\theta'})_{\theta,\theta'}$ the *degree distribution*. To generate the network, connect type-θ' outlinks to type-θ' agents independently across outlinks. Finally, prune self-links from i to i, multilinks from i to j, and—in the unlikely case that no type-θ' agent was realized—all of the unconnectable type-θ' outlinks. Agent i's signal ξ_i consists of her degree $d \in \mathbb{N}^{\Theta}$ after the pruning.[10]

Since $\mathcal{G}_I$ is symmetric across agents, we drop their identities i from the notation of Section 2.2, and write the adoption probabilities, learning curves, and cost thresholds of a degree-d agent as x_d, y_d, and $c_d = \pi(y_d)$, respectively. Taking expectation over the degree of a type-θ agent, we write $x_\theta = E[x_{D_\theta}]$.

When solving for equilibrium, we consider the limit as the number of agents I grows large. The model then nests many natural special cases.

- Directed Finite Trees. In the case of Example 1, set $\Theta = \{i, j\}$ with deterministic degrees $D_{i,j} \equiv 1$ and $D_{j,j} \equiv D_{j,i} \equiv D_{i,i} \equiv 0$. Every Iris-type thus observes one John-type, and equilibrium adoption probabilities are as in Example 1.[11] The following analysis generalizes this example to any finite directed tree.
- Directed Regular Trees. In the case of Example 2, the type space is trivial and the degree is deterministic, $D \equiv 1$. For any finite I, the realized network G gives rise to cycles, but as I grows large the length of these cycles grows large, and the network approximates an infinite line. More generally, setting $D \equiv d > 1$ gives rise to regular directed trees.

[10]This definition differs from the literature on configuration models, such as Sadler (2020), in three ways. (a) Sadler considers undirected networks, to which we turn in Section 3.2. (b) We model agent i's degree as a random variable $D_{\theta'}$, while Sadler fixes the realized degrees d and imposes conditions on the empirical distribution of degrees as I grows large. (c) When a realized network G has self-links or multilinks, we prune these links from G, while Sadler discards G by conditioning the random network on realizing no such links. We view (b) and (c) as technicalities, and deviate from the literature because doing so simplifies our analysis.

[11]Conversely, a John-type is observed by a random number of Iris-types, but that does not matter for individual adoption probabilities.

- Directed Stochastic Block Networks. A prominent instance of networks with random degree are Erdős–Rényi networks, which correspond to a single type and Poisson-distributed D. More generally, stochastic block networks (which are useful for capturing homophily) correspond to a multitype generalization with Poisson-distributed $D_{\theta,\theta'}$.

For large I, the random network locally resembles a tree where the adoption probabilities of an agent's neighbors are approximately independent. The probability that an agent with degree $d = (d_{\theta'})$ observes an adoption is then approximated by $y_d \approx 1 - \prod_{\theta'}(1 - x_{\theta'})^{d_{\theta'}}$. Substituting this approximation into (5), we define (x^*_θ) to be the solution of

$$\dot{x}_\theta = E\left[\Phi\left(1 - \prod_{\theta'}(1 - x_{\theta'})^{D_{\theta,\theta'}}\right)\right]. \tag{10}$$

This is a Θ-dimensional ODE, which is easy to compute. Note that while the number of possible degrees is *infinite*, agents cannot observe their neighbors' degrees and so we solve the learning problem at the level of neighbors' *finite* types. Thus, in the Twitter example, we get one ODE each for celebrities, posters, and lurkers. In a regular, single-type network with degree d, (10) simplifies to

$$\dot{x} = \Phi\big(1 - (1 - x)^d\big), \tag{11}$$

and for $d = 1$ we recover (3).

We now show that this simple, Θ-dimensional ODE is a good description of adoption behavior for large I. Formally, say that a vector of cutoff costs (c_d) is a *limit equilibrium* of the large directed random network with degree distribution D if it is an ϵ_I-equilibrium in $\mathcal{G}_I$ for some sequence (ϵ_I) with $\lim_{I\to\infty} \epsilon_I = 0$. Specifically, let $c^*_d := \pi(1 - \prod_{\theta'}(1 - x^*_{\theta'})^{d_{\theta'}})$ be the cutoff costs associated with (x^*_θ).

PROPOSITION 2: *The cutoffs (c^*_d) are the unique limit equilibrium of the large directed random network with degree distribution D.*

PROOF: See Appendix B.1. *Q.E.D.*

The notion of a limit equilibrium extends our "macroeconomic perspective" from the modeler to the agents. While the real network is finite, agents treat it as infinite; in large networks, the resulting behavior is approximately optimal. For completeness, Appendix D.1 in the Online Supplementary Material (Board and Meyer-ter-Vehn (2021)) provides a "microeconomic perspective" by showing that the equilibria of the finite models $\mathcal{G}_I$ converge to (c^*_d).

Turning to the substantive question of our paper, we now argue that both direct and indirect links improve agents' social learning. Thus lurkers are better off if celebrities or posters increase their number of links. While not surprising, such simple comparative static results have eluded traditional herding models. Figure 2 illustrates the social learning curves as we add links to a directed tree. The left panel compares a lone agent (John in Example 1), an agent with one link (Iris in Example 1), and an infinite chain (Kata in Example 2). The social learning curves shift up as neighbors add more links; the Blackwell-ranking implies that Kata is better off than Iris, who is better off than John. The right panel shows the social learning curves in regular networks with $d = 1$ (i.e., an infinite chain), $d = 5$, and $d = 20$. Again, these social learning curves shift up, so agents benefit from denser trees.

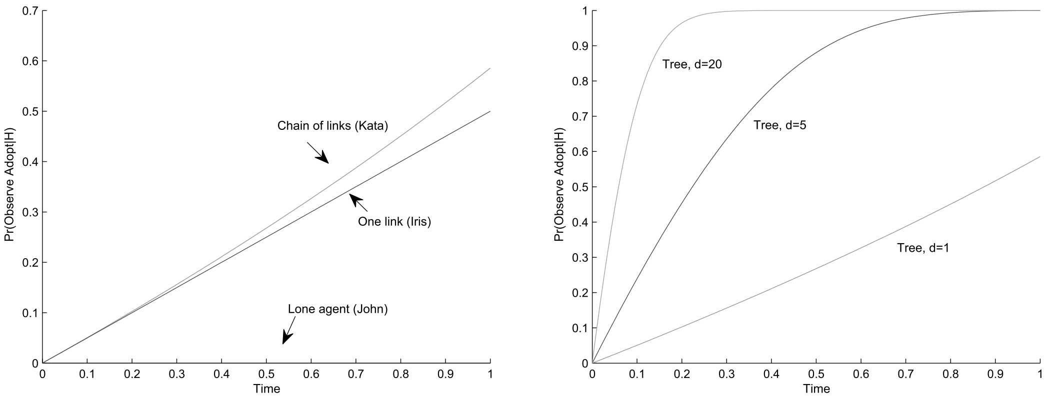

FIGURE 2.—**Social Learning Curves in Directed Tree Networks.** The **left** panel illustrates Examples 1 and 2. The **right** panel shows regular directed trees with degree d. This figure assumes uniform inspection costs, $\kappa \sim U[0, 1]$, and an even prior, $\pi_0 = 1/2$.

Formally, consider two degree distributions such that $\tilde{D} \succeq_{\text{FOSD}} D$ in the usual multivariate stochastic order. Let x^*_θ, $\tilde{x}^*_\theta$ be the associated adoption probabilities derived from (10), and y^*_d and $\tilde{y}^*_d$ be the corresponding social learning curves.

THEOREM 1: *Assume F has a bounded hazard rate* (6). *Social learning and welfare improve with links*: *If $\tilde{D} \succeq_{\text{FOSD}} D$,*
(a) *For any degree d, $\tilde{y}^*_d \geq y^*_d$.*
(b) *For any type θ, $E[\tilde{y}^*_{\tilde{D}_\theta}] \geq E[y^*_{D_\theta}]$.*

PROOF: Recalling Lemma 1, assumption (6) means that $\Phi' \geq 0$. Thus, the right-hand side of (10) FOSD-rises in both the exogenous degree D and the endogenous adoption probabilities x_θ. Thus, the solution (x^*_θ) rises in D, $\tilde{x}^*_\theta \geq x^*_\theta$, and so $\tilde{y}^*_d \geq y^*_d$. Taking expectations over d then yields part (b).[12] *Q.E.D.*

Part (a) says that social information rises if we fix the degree d, and thus speaks to the value of additional indirect links. Obviously, the additional direct links also help, as shown in part (b). This result confirms the intuition that social information is more valuable for "visible" products (e.g., laptops) that are represented by a dense network than for "hidden" products (e.g., PCs).

Theorem 1 is silent about the quantitative impact of direct and indirect links. The next example emphasizes the importance of direct links.

EXAMPLE 5—Two Links vs. Infinite Chain: In Appendix B.2, we show that agent i learns more if she has two uninformed neighbors than if she learns from the infinite directed chain from Example 2. Intuitively, if $i \to j \to k$, then agent k affects i's action only if k enters first, then j enters, and then i enters. Thus, the chance of learning information from the nth removed neighbor in the chain is $\frac{1}{n!}$, suggesting that an infinite chain of signals is worth less than two direct signals, as $\sum_{n=1}^{\infty} 1/n! = e - 1 < 2$. Moreover, these

[12]If we additionally assume that the inequality $\tilde{D} \succeq_{\text{FOSD}} D$ is strict and that the Markov chain on Θ induced by $\tilde{D}$ is irreducible, we more strongly get the strict inequality $\tilde{y}^*_d > y^*_d$.

indirect signals are intermediated (i.e., k's signal must pass through j), which further reduces their information value.

3.2. *Undirected Networks*

We now consider undirected random networks, representing friends on Facebook, or drivers who learn about cars from observing other drivers' choices. To formalize this, we use a single-type, undirected variant of the configuration model. Each agent independently draws $d \in \mathbb{N}$ *link-stubs* generated by a random variable D. We then independently connect these stubs in pairs, and prune self-links, multilinks, and residual unconnected stubs (if the total number of stubs is odd). Appendix D.2 in the Online Supplementary Material extends the model to multiple types.

An important feature of random undirected networks is the *friendship paradox*. Namely, i's neighbors typically have more neighbors than i herself. Formally, we define the neighbor's degree distribution D' by

$$\Pr(D' = d) := \frac{d}{E[D]} \Pr(D = d). \tag{12}$$

For example, in an Erdős–Rényi network, D is Poisson and $D' = D + 1$, whereas in a regular network, $D' = D \equiv d$.

We now study the behavior of the limit economy as the number of agents grows large. This allows us to treat neighbors' adoptions as independent; Proposition 2′ in Section 3.5 justifies this approach. The simplest such network, corresponding to $D \equiv 1$, is the undirected pair in Example 3. Following this example, we write $\bar{x}$ for the probability that i's neighbor j has adopted at $t \leq t_i$. With general degree distribution D, neighbor j additionally learns from another $D' - 1$ independent links, from which he observes no adoption with probability $(1 - \bar{x})^{(D'-1)}$. All told, agent i expects j to observe an adoption with objective probability $1 - (1 - \bar{x})^{(D'-1)}$, while j expects to observe an adoption with the higher, subjective probability $1 - (1 - \bar{x})^{D'}$. So motivated, define $\bar{x}^*$ as the solution of

$$\dot{\bar{x}} = E\big[\phi\big(1 - (1 - \bar{x})^{D'-1}, 1 - (1 - \bar{x})^{D'}\big)\big]. \tag{13}$$

Agent i's actual, unconditional adoption rate then equals $E[\Phi(1 - (1 - \bar{x}^*)^D)]$. Equation (13) implies that, as in Theorem 1, social learning increases with the number of links D. We prove this in Appendix D.3 in the Online Supplementary Material.

We now show how backward links harm social learning.

EXAMPLE 3—Undirected Pair $i \leftrightarrow j$, continued: Start with $i \to j$, and consider the effect of the backward link $j \to i$ on i's social information, as illustrated in the left-hand panel of Figure 3. Equation (8) implies that $\dot{\bar{x}}_j \leq F(\pi_0)$, and so $\bar{x}_j \leq F(\pi_0)t$, which is j's adoption curve if he does not observe i. Thus, the link $j \to i$ lowers i's social information and her utility. Intuitively, when agent i enters the market at t_i, she knows that j cannot have seen her adopt; however, j does not know the reason for i's failure to adopt. This makes j more pessimistic, reduces his adoption probability, and lowers i's social learning curve and utility. Of course, the backward link also makes j better off.

To address the overall welfare effect of backward links, we compare a network where agents have D directed links to one with D undirected links, as illustrated in Figure 4. Recalling neighbors' limit adoption probabilities in directed and undirected networks x^*,

FIGURE 3.—**Networks from Examples 3 and 6.** The **left** panel adds a backward link. The **right** panel adds a correlating link.

$\bar{x}^*$ from equations (10) and (13), respectively, we write $y_d^* = 1-(1-x^*)^d$ and $\bar{y}_d = 1-(1-\bar{x}^*)^d$ for the respective social learning curves.

THEOREM 2: *Assume $D'-1 \preceq D$ in the FOSD-order. Social learning and welfare are higher when the network is directed rather than undirected: For any degree d, $y_d^* > \bar{y}_d^*$.*

PROOF: Rewriting (10) for a single type, the adoption probability of any given neighbor in the directed network follows:

$$\dot{x} = E\big[\phi\big(1-(1-x)^D, 1-(1-x)^D\big)\big]. \tag{14}$$

Using $D'-1 \preceq D \preceq D'$, $\partial_1\phi > 0$, and $\partial_2\phi < 0$, the right-hand side of (14) exceeds the right-hand side of (13) for $x = \bar{x}$. By the Single-Crossing Lemma (Appendix B.3), the solution of (14) exceeds the solution of (13), $x^* > \bar{x}^*$, and so $y_d^* > \bar{y}_d^*$. *Q.E.D.*

Theorem 2 says that fixing the degree distribution, directed networks generate better information than undirected networks. Intuitively, in an undirected network an agent's neighbors cannot have seen her adopt when she enters; this makes them more pessimistic

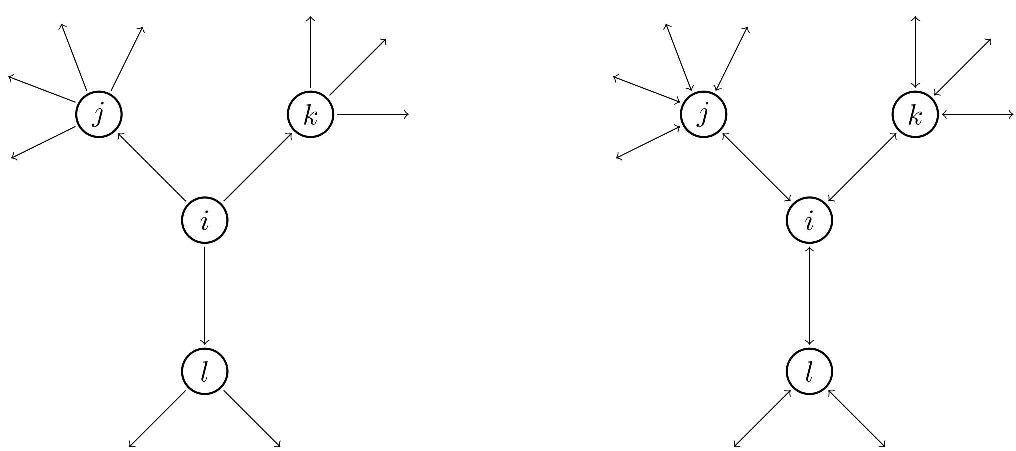

FIGURE 4.—**Comparing Directed and Undirected Networks.** This figure illustrates two Erdős-Rényi networks from agent i's perspective, showing i's neighbors and their outlinks. For simplicity, the picture of the directed network does not show inlinks to i or her neighbors; in a large network, these inlinks do not affect i's learning. Observe that i's neighbors have one more outlink in the undirected network, namely the link to i; this reflects the friendship paradox.

and reduces social learning. Countervailing this effect is the fact that neighbors have a higher degree in the undirected network because of the friendship paradox, $D \preceq D'$. The assumption $D' - 1 \preceq D$ limits this countervailing effect.[13]

3.3. *Clustering*

One prominent feature of real social networks is clustering, whereby i's neighbors j, k are also linked to each other. For example, consider an agent who gets information from her family, her geographic neighbors, and her colleagues; we think of information as independent across groups but correlated within them.

The bilateral configuration models in the previous subsections do not give rise to such clustering since the chance that any two neighbors of i are linked vanishes for large I. To capture clustering, we consider the following variant of the configuration model. Each agent independently draws D *pairs of link-stubs*, which are then randomly connected to two other pairs of link-stubs to form a triangle. As in Section 3.2, we then prune self-links and multilinks, as well as leftover pairs if the total number of pairs is not divisible by three. Also, recall the weighted distribution D' from (12) that captures the number of link-pairs of a typical neighbor by accounting for the friendship paradox.

We now study the behavior of the limit economy as the number of agents grows large; Proposition 2′ in Section 3.5 justifies this approach. Adoption is independent across neighboring triangles but correlated within them. The simplest such network, corresponding to $D \equiv 1$, is the triangle from Example 4 with $I = 2$. There we argued that the learning curve is determined by the adoption probability $\hat{x}$ of the first adopter. Since the first adopter expects to see no adoption with subjective probability $(1 - \hat{x})^2$ but objectively never observes an adoption, we concluded that $\dot{\hat{x}} = \phi(0, 1 - (1 - \hat{x})^2)$. For a general distribution D, agent i's neighbors additionally learn from another $D' - 1$ independent triangles, from which they observe no adoption with probability $(1 - \hat{x})^{2(D'-1)}$. All told, define $\hat{x}^*$ as the solution of

$$\dot{\hat{x}} = E\big[\phi\big(1 - (1 - \hat{x})^{2(D'-1)}, 1 - (1 - \hat{x})^{2D'}\big)\big]. \tag{15}$$

Agent i's actual, unconditional adoption rate then equals $E[\Phi(1 - (1 - \hat{x})^{2D})]$.

We now show how clustering can be harmful to social information and welfare.

EXAMPLE 6—Correlating Link: Assume agent i initially observes two uninformed agents j and k, as in the right panel of Figure 3. The probability that neither adopts is $(1 - F(\pi_0)t)^2$. Now, suppose we add a link from j to k, correlating their adoption outcomes. Agent k's behavior is unchanged, but the probability that agent i sees an adoption decreases. This is because the probability $x_{j|\neg k}$ that j adopts conditional on k not adopting follows $\dot{x}_{j|\neg k} = F(\pi(x_k)) < F(\pi_0)$. Intuitively, agent i just needs one of her neighbors to adopt. Adding the link $j \to k$ makes j more pessimistic and lowers his adoption probability exactly in the event when his adoption would be informative for i, namely, when k has

[13]To see how the friendship paradox can overturn Theorem 2, suppose that agents are equally likely informed, $d = 100$, or uninformed, $d = 0$. In the directed network, agents are equally likely to be looking at an informed or uninformed neighbor. In contrast, in the undirected network, agents are only looking at informed neighbors.

The condition $D' - 1 \preceq D$ is tight: Assume a degenerate, binary cost distribution F with atoms at $\underline{\kappa} = 0$ and $\bar{\kappa} > \pi_0$; such an F is approximated by distributions that satisfy our bounded pdf assumption. Agents who do not observe an adoption inspect iff $\kappa = \underline{\kappa}$, irrespective of y, and so $\partial_2 \phi = 0$. If the degree distribution is Poisson, $D' - 1 = D$, the right-hand sides of (13) and (14) coincide, and so $y_d^* = \bar{y}_d^*$.

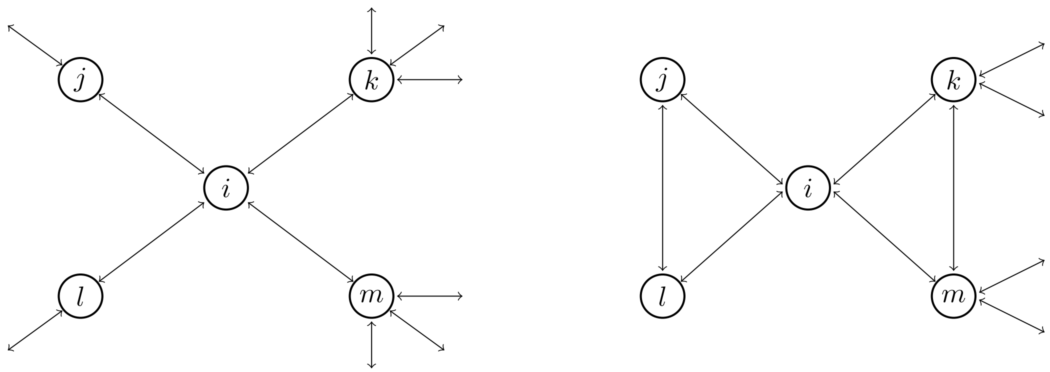

FIGURE 5.—**Bilateral Links vs. Triangles.** This figure illustrates two networks from agent i's perspective. In the left network, everyone has $2D$ bilateral links, where $D = 2$ for $\{i, k, m\}$ and $D = 1$ for $\{j, l\}$. In the right network, everyone is part of D triangles.

not adopted. Thus, the correlating link makes agent i worse off. Of course, this link also makes agent j better off.

To address the overall welfare effect of correlating links, we compare an undirected network with D pairs of link-stubs to one with $2D$ bilateral link-stubs, as illustrated in Figure 5. The social learning curve equals $\hat{y}^*_{2d} = 1 - (1 - \hat{x}^*)^{2d}$ in the former network, and $\bar{y}^*_{2d} = 1 - (1 - \bar{x}^*)^{2d}$ in the latter, where $\hat{x}^*$ solves (15) and $\bar{x}^*$ solves (13) with D replaced by $2D$.

THEOREM 3: *Clustering reduces social learning and welfare*: *For any degree* d, $\hat{y}^*_{2d} < \bar{y}^*_{2d}$.

PROOF: Equation (12) implies that with $2D$ bilateral links, the link distribution of a neighbor equals $(2D)' = 2D'$.[14] Thus, the conditional adoption probability of one's neighbor follows:

$$\dot{\bar{x}} = E\left[\phi\left(1 - (1 - \bar{x})^{2D'-1}, 1 - (1 - \bar{x})^{2D'}\right)\right]. \tag{16}$$

Since $\partial_1 \phi > 0$, the right-hand side of (16) exceeds the right-hand side of (15) when $\bar{x} = \hat{x}$. Thus, the Single-Crossing Lemma implies $\hat{x}^* < \bar{x}^*$ and so $\hat{y}^*_{2d} < \bar{y}^*_{2d}$. *Q.E.D.*

Agents learn slower in cliques than with an equivalent number of independent links. Intuitively, agent i needs one of her neighbors to be sufficiently optimistic that they are willing to experiment. Cliques correlate the decisions of i's neighbors, making them more pessimistic in exactly the event when i most wants them to experiment.

3.4. *Correlation Neglect*

Bayesian updating on networks can be very complex as agents try to distinguish new and old information. For example, Eyster and Rabin's (2014) "shield" example shows that a Bayesian agent i should counterintuitively "anti-imitate" a neighbor j if j's action is also encoded in the actions of i's other neighbors. Instead of anti-imitating, agents may adopt

[14] Indeed, $\Pr((2D)' = 2d) = \Pr(2D = 2d)\frac{2d}{E[2D]} = \Pr(D = d)\frac{d}{E[D]} = \Pr(D' = d) = \Pr(2D' = 2d)$.

heuristics, such as ignoring the correlations between neighbors' actions (e.g., Eyster et al. (2018), Enke and Zimmermann (2019), Chandrasekhar et al. (2020)). Our model can be adapted to capture such misspecifications, and predicts that correlation neglect reduces social learning.[15]

To model correlation neglect, we consider a configuration model where agents draw D pairs of undirected triangular stubs, but agents believe all their information is independent. That is, i believes that her neighbors are not connected, believes her neighbors think their neighbors are not connected, and so on. All told, agents think that the network is as in Section 3.2, while in reality it is as in Section 3.3.[16]

Consider the limit as I grows large. Since agent i believes that links are generated bilaterally, her subjective probability assessment that any of her neighbors has adopted, $\bar{x}^*$, solves (16). An agent with $2d$ links thus uses cutoff $\pi(1-(1-\bar{x}^*)^{2d})$ when choosing whether to inspect. In reality, agent i's neighbors form triangles (i, j, k), and so the objective probability $\check{x}^*$ that the first adopter in a triangle adopts follows a variant of the usual first adopter triangle formula (15),

$$\dot{\check{x}} = E\left[\phi\left(1-(1-\check{x})^{2(D'-1)}, 1-\left(1-\bar{x}^*\right)^{2D'}\right)\right]. \tag{17}$$

Intuitively, the first adopter in (i, j, k) expects to see an adoption with probability $\bar{y}^*_{2d} = 1-(1-\bar{x}^*)^{2D'}$, but the objective adoption probability is $\check{y}^*_{2d} = 1-(1-\check{x}^*)^{2(D'-1)}$.[17]

Turning to the effect of correlation neglect on social learning, consider an agent with $2d$ links in a network of triangles. Her social information is $\hat{y}^*_{2d} = 1-(1-\hat{x}^*)^{2d}$ in equilibrium and $\check{y}^*_{2d}$ under correlation neglect.

THEOREM 4: *Correlation neglect reduces social learning*: *For any degree* d, $\check{y}^*_{2d} < \hat{y}^*_{2d}$.

PROOF: By the proof of Theorem 2, clustering decreases neighbors' adoption rates, $\bar{x}^* > \hat{x}^*$. Since $\partial_2\phi < 0$, the RHS of (17) is smaller than the RHS of (15) when $\check{x} = \hat{x}$. Then the Single-Crossing Lemma implies that $\check{x}^* < \hat{x}^*$ and so $\check{y}^*_{2d} < \hat{y}^*_{2d}$. *Q.E.D.*

Intuitively, when agent j believes all his sources of information are independent, he overestimates the chance of observing at least one adoption, and grows overly pessimistic when he observes none. This reduces j's adoption probability and reduces agent i's social information.

As a corollary, correlation neglect reduces welfare: By Theorem 4, it reduces i's social information; additionally, it causes her to react suboptimally to her information. Ironically, while correlation neglect lowers i's objective expected utility, it raises her subjective expected utility: Formally, $\bar{y}^*_{2d} > \hat{y}^*_{2d}$, so her subjective social information is higher than in equilibrium. Intuitively, correlation neglect makes i overly optimistic about the chance

[15]There is a growing literature studying herding models with misspecified beliefs. Eyster and Rabin (2010) and Bohren and Hauser (2021) study complete networks, while Eyster and Rabin (2014) and Dasaratha and He (2020) consider the role of the network structure.

[16]To formally capture misspecification in the general model of Section 2, we drop the assumption that agents' beliefs $\mu(G, \xi_{-i}|\xi_i)$ are deduced from a common prior $\mu(G, \xi)$.

[17]Formally, (17) is an instance of equations (33)–(34) in Appendix B.4. All of these equations leverage the fact that adoption dynamics in our model are modular in (i) the true network (the first argument of ϕ), and (ii) agents' behavior (the second argument of ϕ), which in turn depends on their beliefs about the network. It is this modularity, which allows us to easily extend our analysis to non-equilibrium behavior, such as correlation neglect in this section.

of observing at least one adoption. It is precisely this over-optimism that reduces actual adoption probabilities by flipping into over-pessimism when the adoption fails to materialize.

3.5. *General Undirected Networks: A Limit Result*

We now define random networks that encompass the undirected links and cliques from the last three sections and show that limit equilibria of this model are described by a simple, two-dimensional ODE (18)–(19).

To define these networks $\hat{\mathcal{G}}_I$, suppose every agent independently draws $\bar{D}$ bilateral stubs and $\hat{D}$ pairs of triangle stubs with finite expectations. We connect pairs of bilateral stubs and triples of triangular stubs at random, and then prune self-links, multilinks, and left-over stubs (if $\sum \bar{d}_i$ is odd or $\sum \hat{d}_i$ is not divisible by three). Agents know their number of bilateral links and triangle links after the pruning.[18]

Assume that I is large, and define neighbors' link distributions $\bar{D}'$ and $\hat{D}'$ as in (12). Since $\bar{D}$ and $\hat{D}$ are independent, a neighbor on a bilateral link has $\bar{D}'$ bilateral links and $\hat{D}$ triangle link pairs, whereas a neighbor on a triangular link has $\bar{D}$ bilateral links and $\hat{D}'$ triangle link pairs. As in Sections 3.2 and 3.3, agents condition on the fact that their neighbors cannot have seen them adopt. So motivated, define $(\bar{x}^*, \hat{x}^*)$ as the solution to the two-dimensional ODE,

$$\dot{\bar{x}} = E\big[\phi\big(1-(1-\bar{x})^{\bar{D}'-1}(1-\hat{x})^{2\hat{D}}, 1-(1-\bar{x})^{\bar{D}'}(1-\hat{x})^{2\hat{D}}\big)\big], \tag{18}$$

$$\dot{\hat{x}} = E\big[\phi\big(1-(1-\bar{x})^{\bar{D}}(1-\hat{x})^{2(\hat{D}'-1)}, 1-(1-\bar{x})^{\bar{D}}(1-\hat{x})^{2\hat{D}'}\big)\big]. \tag{19}$$

As in Section 3.2, $\bar{x}$ is the probability that i's bilateral neighbor j adopts before t_i. Agent j's subjective probability of observing no adoption conditions on $\bar{D}'$ bilateral links and $\hat{D}$ triangle link pairs; but from i's objective perspective, the number of bilateral links on which j could observe an adoption drops to $\bar{D}'-1$. Similarly, as in Section 3.3, $\hat{x}$ is the probability that the first adopter j in one of i's triangles adopts before t_i. Agent j's subjective probability of observing no adoption conditions on $\bar{D}$ bilateral links and $\hat{D}'$ triangle link pairs; but from i's objective perspective, the number of triangle link pairs on which j could observe an adoption drops to $\hat{D}'-1$.

Given beliefs $(\bar{x}^*, \hat{x}^*)$, an agent with $\bar{d}$ bilateral neighbors and $\hat{d}$ pairs of triangular neighbors adopts cutoffs $c^*_{\bar{d},\hat{d}} = \pi(1-(1-\bar{x}^*)^{\bar{d}}(1-\hat{x}^*)^{2\hat{d}})$, and her unconditional adoption probability follows $\dot{x}_{\bar{d},\hat{d}} = \Phi(1-(1-\bar{x})^{\bar{d}}(1-\hat{x})^{2\hat{d}})$. We can now extend the limit result, Proposition 2, to undirected networks with cliques.

PROPOSITION 2′: *The cutoffs $(c^*_{\bar{d},\hat{d}})$ are the unique limit equilibrium of $\hat{\mathcal{G}}_I$.*

PROOF: See Appendix B.4. *Q.E.D.*

[18] Chandrasekhar and Jackson (2021) propose an alternative, closely related "Subgraph generation model" (SUGM) of large random networks. SUGMs avoid the notion of link stubs and rather connect any set of nodes into subgraphs (e.g., triangles with a specific combination of node types), independently across sets. That model accommodates rich subgraphs more easily than our configuration model; on the other hand, it does not allow the flexibility of specifying the degree distribution directly. But these differences are not crucial, and we conjecture that one can solve for our social learning equilibria on SUGMs based on simple subgraphs.

Social learning in this complex network is thus characterized by a simple, two-dimensional ODE, (18)–(19). The model in this section is already rich enough to match important network statistics, such as the degree distribution or the clustering coefficient. But the logic behind equations (18)–(19) and Proposition 2′ easily accommodates additional features and alternative modeling assumptions. Allowing for larger $(n+1)$-cliques amounts to replacing the "2" in the exponents of (18)–(19) by "n". Allowing for multiple types θ of agents is slightly more complicated since it requires keeping track of the conditional adoption probability of type θ's neighbor θ' for all pairs (θ, θ'); we spell this out in Appendix D.2 in the Online Supplementary Material. Our analysis also extends to correlation of $\bar{D}$ and $\hat{D}$, and to the alternative assumption that agents know only their total number of links but cannot distinguish bilateral from triangle links.

4. DISCUSSION

We round the paper off by studying the model's implications for information aggregation (Section 4.1) and extending the analysis to imperfect social learning (Section 4.2).

4.1. *Information Aggregation and the Value of Links*

We now reconsider one of the most central issues in social learning: Does society correctly aggregate dispersed information? The configuration model provides a novel perspective on this question. Consider a regular network as both the degree d and the total number of agents I grow large. We show that if I grows sufficiently faster than d, then agents have access to a large number of almost independent signals and society correctly aggregates information. However, if d grows too quickly, the network becomes "clustered" and information aggregation can fail. As a corollary, in a large society with a fixed number of agents, more links can introduce excessive correlation and lower everyone's utility.

To see the problem with clustering, recall the complete network with $I+1$ agents from Example 4 and the first adopter probability $\hat{x}_I$. When seeing no adoption, inspection stops once the social information y reaches the "choke point" $\bar{y} = \pi^{-1}(\underline{\kappa})$, recalling the lowest cost type $\underline{\kappa} < \pi_0$. As $I \to \infty$, social information $y_I = 1 - (1 - \hat{x}_I)^I$ immediately rises to $\bar{y}$ and stays there for all $t > 0$.[19] Intuitively, observing no adoption from an exploding number of agents I makes agents grow pessimistic so fast that they are willing to inspect only at the very first instant. Learning is perfect when $\bar{y} = 1$, which is the case if and only if $\underline{\kappa} = 0$. Thus, unboundedly low costs are necessary and sufficient for information aggregation, as in Mueller-Frank and Pai (2016) and Ali (2018). In particular, when $\underline{\kappa} > 0$, high-quality products fail to take off with probability $1 - \bar{y} > 0$, in which case they fizzle out immediately.[20]

This failure of information aggregation does not arise when the network remains sparse as the degree grows. Specifically, consider the limit $d \to \infty$ of our large directed networks,

[19] *Proof*: To see $\lim_{I\to\infty} y_I = \bar{y}$ for all $t > 0$, first note that the first adopter stops experimenting, $\dot{\hat{x}}_I = 0$, when $y_I = 1 - (1 - \hat{x}_I)^I$ rises above $\bar{y}$, so $y_I \leq \bar{y}$ for any I, t. For the opposite inequality, if $\limsup_{I\to\infty} y_I < \bar{y}$ for some $t > 0$, then $\dot{\hat{x}}_I = F(\pi(y_I))$ is bounded away from 0 on $[0, t]$, hence $\hat{x}_I$ is bounded away from 0 at t. Then $\limsup_{I\to\infty}(1 - \hat{x}_I)^I \to 0$ at t, contradicting the initial assumption that $\limsup_{I\to\infty} 1 - (1 - \hat{x}_I)^I < \bar{y} \leq 1$.

[20] As $I \to \infty$, agents always stop inspecting low-quality products. The asymmetry between good and bad products is seen elsewhere in the literature (e.g., Guarino et al. (2011), Hendricks et al. (2012), and Herrera and Hörner (2013)).

where adoption and social information are given by $\dot{x}_d^* = \Phi(y_d^*)$ and $y_d^* = 1 - (1 - x_d^*)^d$, respectively, as defined in (11). In other words, this is the double limit of a regular random network where first $I \to \infty$ and then $d \to \infty$. Since $\Phi(y)$ is bounded away from 0,[21] we have $y_d^* \to 1$ for all $t > 0$ as $d \to \infty$; that is, information becomes perfect instantaneously, irrespective of $\underline{\kappa}$. Intuitively, agents' signals on a sparse network are independent and their joint adoption decisions become perfectly informative as the degree grows. This contrasts with Acemoglu et al. (2011) where "for many common deterministic and stochastic networks, bounded private beliefs are incompatible with asymptotic learning."[22]

4.2. *Imperfect Social Learning*

The baseline model assumes that agents adopt the product if and only if quality is high. Observing an adoption is thus "perfect good news," and the agent's belief $\pi = \Pr(q = H)$ jumps to 1. This simplifies the analysis by allowing us to summarize i's social information (at time-t) by a single number, y_i. This analysis immediately extends to the case when, for idiosyncratic reasons, agents adopt only the high-quality product with probability $\alpha^H < 1$. In this section, we show how to extend the analysis to more general imperfect social learning where agents adopt the low-quality product with probability $\alpha^L > 0$. We first show that in large directed random networks, agents' learning improves with additional direct and indirect links (as in Theorem 1). We then discuss the self-reflection and correlation effects, arguing that backward links still inhibit learning (as in Example 3) but that correlating links may improve learning (unlike in Example 6). Proofs are in Appendix C.

As in the baseline model, agent i enters at time $t_i \sim U[0, 1]$, observes which neighbors have adopted by time t_i, and updates her belief about quality, π_i. Inspection costs $\kappa_i \sim F[\underline{\kappa}, \overline{\kappa}]$ and perfectly reveals agent i's utility, which is determined by the quality $q \in \{L, H\}$ and the agent's idiosyncratic preference.[23] Specifically, agent i's utility from adopting a product of quality q is random (and i.i.d. across agents), equal to v^q with probability α^q, and equal to $-M$ with probability $1 - \alpha^q$. As before, nonadoption yields zero utility. Upon inspection, agent i thus adopts good q with probability α^q and has expected utility $\alpha^q v^q$; naturally, we assume $0 < \alpha^L < \alpha^H < 1$ and $0 < \alpha^L v^L < \alpha^H v^H$. Given belief π, the cutoff cost at which the agent is indifferent between inspecting and passing is given by $c(\pi) = \pi(\alpha^H v^H - \alpha^L v^L) + \alpha^L v^L$, which is linear and increasing in π. For simplicity, we assume an even prior, $\pi_0 = 1/2$, and M large enough so that agents never adopt blindly.[24]

[21]*Proof*: Since $F(\pi(0)) > 0$, there exists $\epsilon > 0$ with $F(\pi(y)) \geq \epsilon$ for all $y \leq \epsilon$. Then $\Phi(y) = 1 - (1 - y)(1 - F(\pi(y))) \geq \max\{y, F(\pi(y))\} \geq \epsilon$ for all y.

[22]The distinguishing feature of our configuration model $\mathcal{G}_I$ with exploding I is that each agent's neighborhood naturally resembles a tree. This feature does not arise naturally in Acemoglu et al's model with one infinite sequence of agents. Indeed, Acemoglu et al.'s positive results (e.g., their Theorem 4) rely on a carefully crafted, slowly exploding number of "guinea pig" agents, who have little social information, and whose actions are thus very informative about their private signals. In our model, such guinea pigs arise naturally, by virtue of being the first to enter the market. Other differences between the models, such as the fact that their agents enter the game with private information whereas ours choose to acquire private information, are not important for this contrast.

[23]In this model, the social signals from other agents' adoption are imperfect, while private signals from inspection are perfect. It would be natural to consider a model with imperfect private signals (e.g., Ali (2018)), but this would undermine the clean separation between the role of social information (which determines the inspection decision) and private information (which determines the adoption decision), that enhances the tractability of our analysis.

[24]We could alternatively adopt Hendricks et al.'s (2012) model where quality and idiosyncratic preferences are additively separable. Our approach is useful for the interpretation of Proposition 3.

We first study adoption on the (multitype) directed random network described in Section 3.1. An agent with degree $d \in \mathbb{N}^\Theta$ observes that a_θ neighbors of type θ have adopted and $d_\theta - a_\theta$ have not. Let $y^q_{d,a}$ be the probability of observing $a \in \mathbb{N}^\Theta$ neighbors adopt given quality q (as always, omitting the time subscript). Given $\pi_0 = 1/2$, the posterior probability of high quality in this event equals

$$\pi_{d,a} = \frac{y^H_{d,a}}{y^H_{d,a} + y^L_{d,a}}. \tag{20}$$

We interpret $Y_d = (y^q_{d,a})_{a,q}$ as the agent's information structure, and rank such structures by the Blackwell-order $\succeq_{\mathrm{BW}}$; this order is characterized by a mean-preserving spread of the random posteriors Π_d (e.g., Börgers (2021)).[25]

We start with best responses. Analogous to equation (5), the adoption curve of a degree-d agent follows:

$$\dot{x}^q_d = \alpha^q \sum_{a \le d} y^q_{d,a} F\big(c(\pi_{d,a})\big). \tag{21}$$

Unpacking (21), $y^q_{d,a}$ is the probability that a neighbors adopt; then the agent inspects with probability $F(c(\pi_{d,a}))$ and adopts with probability α^q.

We now show that the informativeness of agent d's adoption improves in her information (as in Lemma 1). To do this, we strengthen our bounded hazard rate assumption (6) and assume that

$$\pi F\big(c(\pi)\big) \text{ is convex and } (1-\pi) F\big(c(\pi)\big) \text{ is concave in } \pi. \tag{22}$$

This is satisfied if $\kappa \sim U[\alpha^L v^L, \alpha^H v^H]$, so the lowest-cost agent always inspects, while the highest-cost agent never does.

LEMMA 1′: *Assume $\pi_0 = 1/2$ and F satisfies* (22). *In a directed random network, when an agent's social information Y_d Blackwell-improves, her adoption rate $\dot{x}^q_d$ rises for $q = H$ and falls for $q = L$.*[26]

We now suppose that the network becomes large, and show that more links improve learning. In large networks, an agent's neighbors adopt independently with probability $x^q_\theta := E[x^q_{D_\theta}]$, where the expectation is taken over the random degree D_θ. Thus, we can close (21) by computing the probability that $a \le d$ adopt as

$$y^q_{d,a} = \prod_\theta \big(x^q_\theta\big)^{a_\theta} \big(1 - x^q_\theta\big)^{d_\theta - a_\theta}. \tag{23}$$

Substituting (23) into (20) and (21) and taking expectations over $D_\theta = d$ yields a 2Θ-dimensional ODE for adoption probabilities x^q_θ, generalizing equation (10).

[25] With perfect good news, we have $y^H_{d,0} = 1 - y_d$ and $y^L_{d,0} = 1$. Thus, the random posterior Π_d equals $\pi_{d,a} = 1$ for $a > 0$, which happens with probability $\pi_0 y_d$, and $\pi_{d,0} = \pi(y_d)$ with the residual probability. Since $\pi(\cdot)$ is a decreasing function, Π_d is increasing in y_d in the $\succeq_{\mathrm{BW}}$ order.

[26] In the baseline model, Lemma 1 showed that information raises adoption when $q = H$. Here, Lemma 1′ further shows that information lowers adoption when $q = L$. Thus, adoption becomes more dependent on quality and thereby more informative.

THEOREM 1′: *Assume* $\pi_0 = 1/2$ *and* F *satisfies* (22). *In a large directed random network, social learning and welfare improve with links*: *If* $\tilde{D} \succeq_{\text{FOSD}} D$,

(a) *For any degree* d, $\tilde{Y}_d \succeq_{\text{BW}} Y_d$,

(b) *For any type* θ, $\tilde{Y}_{\tilde{D}_\theta} \succeq_{\text{BW}} Y_{D_\theta}$.

We can now examine how social information depends on agents' idiosyncratic preferences for the two goods. We say there is a *broadening* of good q if α^q rises and v^q declines so expected utility $\alpha^q v^q$ and thus the cutoff cost $c(\pi)$ remain constant. That is, the good becomes twice as popular, but the fans are half as enthusiastic. In the opposite case, we speak of a *narrowing*.

PROPOSITION 3: *Assume* $\pi_0 = 1/2$ *and* F *satisfies* (22). *In a large directed random network*, *social information* Y_d *rises for all* d *if*:

(a) *There is a broadening of good* H.

(b) *There is a narrowing of good* L.

(c) *There is a broadening of both* L, H *such that* α^H/α^L *stays constant*.

Parts (a,b) reflect the idea that there is more social information if more people adopt H and fewer people adopt L. That is, social learning can be used to learn about popularity but not passion. In the limit, when $\alpha^L = \alpha^H$, there is no social learning, even if v^H is far higher than v^L. Part (c) states that if both goods become twice as popular, then social information also increases. Intuitively, lowering (α^L, α^H) while fixing α^H/α^L amounts to losing the signal that neighbor j adopted with some probability, leading to a Blackwell-decrease in information.

Finally, we return to the self-reflection and correlation effects. Appendix C.4 reconsiders Example 3 under imperfect social learning and confirms our finding that adding the backward link $j \to i$ to the directed pair $i \to j$ harms i's learning. As before, the backward link makes j more pessimistic, lowering his adoption of both L and H goods by equal amounts. This is analogous to losing j's signal with positive probability, and Blackwell-decreases i's information.

Appendix C.5 reconsiders Example 6, where i observes j and k, and shows that adding the correlating link $j \to k$ may benefit agent i. In the baseline model, the correlating link lowers the probability that at least one of i's neighbors adopts, and thus lowers i's social information. Central to this argument is the assumption that one adoption is enough. With imperfect social learning, we present a cost distribution $F(\kappa)$ where high-cost agents require two adoptions to inspect. The correlating link then raises the probability that both adopt, and thus raises i's social information. This is consistent with Centola's (2010) experiment where clustering raises social learning.

5. CONCLUSION

Social learning plays a crucial role in the diffusion of new products (e.g., Moretti (2011)), financial innovations (e.g., Banerjee et al. (2013)), and new production techniques (e.g., Conley and Udry (2010)). This paper proposes a tractable model of social learning on large random networks, characterizes equilibrium in terms of simple differential equations, and studies the effect of network structure on learning dynamics. The model can be used to structurally estimate diffusion in real-world networks while maintaining Bayesian rationality.

We started the paper by asking about the effect of clustering and connectedness on learning and adoption. In our baseline model, we showed that clustering unambiguously slows learning by correlating neighbors' adoption decisions. Connectedness thus improves social information as long as the network remains sparse, but eventually harms learning as the network becomes clustered. And these results on information directly apply to adoption under our bounded hazard rate assumption (6).

Our analysis goes beyond traditional contagion models of behavior, such as Morris (2000). Indeed, consider a binary distribution F with atoms at $\underline{\kappa} = 0$ and $\bar{\kappa} > \pi_0$ as in footnote 13. Behavior is then mechanical, with agents adopting with probability $F(0)$ if they see no adoption, and with certainty otherwise. In this limit case of our model, network density still enhances learning (Theorem 1 applies), but we lose the adverse effects of the backward link (Example 3) and the correlating link (Example 6). Moreover, in contrast to Section 4.1, a clique maximizes social learning.[27]

Moving forward, one can take the model in a number of different directions. One could study the effect of policies, such as changing the price of the product (e.g., Campbell (2013)) or seeding the network (e.g., Akbarpour et al. (2020)). While we studied correlation neglect, one could allow for other misspecifications of beliefs (e.g., Bohren and Hauser (2021)). Finally, one could endogenize the timing of moves by allowing skeptical agents to delay their decision (e.g., Chamley and Gale (1994)).

APPENDIX A: PROOFS FROM SECTION 2

A.1. *Proof of Proposition 1*

We will characterize equilibrium adoption in a general random network $\mathcal{G} = (I, \Xi, \mu)$ via a system of ODEs, albeit in a large state space. Denote the *state* of the network by $\lambda = \{\lambda_i\}_{i\in I}$, where $\lambda_i \in \{\emptyset, a, b\}$. Let $\lambda_i = \emptyset$ if i has yet to enter, $t \le t_i$; $\lambda_i = a$ if i has entered and adopted; and $\lambda_i = b$ if i has entered and not adopted. Given state λ, let λ^{-i} denote the same state with $\lambda_i = \emptyset$.

Fix a network G and agents' signals ξ, and condition on a high-quality product, $q = H$. We can then describe the distribution over states by $z_{\lambda,G,\xi}$ (as always omitting the dependence on time t). Figure 6 illustrates the evolution of the state via a Markov chain in a three-agent example. Probability mass flows into state $\lambda = (\lambda_i, \lambda_j, \lambda_k) = (\emptyset, a, b)$ from state λ^{-j} as agent j enters and adopts, and from λ^{-k} as agent k enters and doesn't adopt.

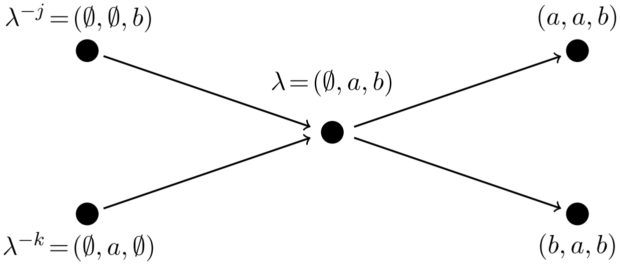

FIGURE 6.—Illustrative Markov Transitions underlying Proposition 1.

[27]*Proof*: In the I-clique, i observes an adoption at time t_i iff $t_j \le t_i$ and $\kappa_j = 0$ for some $j \ne i$. This event is also necessary for i to observe an adoption in any other network.

Similarly, probability flows out of state λ, and into states (a, a, b) and (b, a, b), as agent i enters.

Quantifying these transition rates, we now argue that the equilibrium distribution over the states λ evolves according to the following ODE:

$$(1-t)\dot{z}_{\lambda,G,\xi} = -\sum_{i:\lambda_i=\emptyset} z_{\lambda,G,\xi} + \sum_{\substack{i:\lambda_i=a,\\ \exists j\in N_i(G):\lambda_j=a}} z_{\lambda^{-i},G,\xi} + \sum_{\substack{i:\lambda_i=a,\\ \forall j\in N_i(G):\lambda_j\neq a}} z_{\lambda^{-i},G,\xi} F\big(\pi(y_{i,\xi_i})\big) + \sum_{\substack{i:\lambda_i=b,\\ \forall j\in N_i(G):\lambda_j\neq a}} z_{\lambda^{-i},G,\xi}\big(1-F\big(\pi(y_{i,\xi_i})\big)\big). \tag{24}$$

To close (24), the probability that i observes an adoption at time $t_i = t$ equals $y_{i,\xi_i} = E[y_{i,G,\xi}]$ with $y_{i,G,\xi} = \Pr(\exists j \in N_i(G) : \lambda_j = a|\lambda_i = \emptyset) = \frac{1}{1-t}\sum z_{\lambda,G,\xi}$ where the sum is over all λ with $\lambda_i = \emptyset$ and $\lambda_j = a$ for at least one $j \in N_i(G)$.

To derive (24), fix a state λ. Agents i that have not yet entered, $\lambda_i = \emptyset$, enter uniformly over time $[t, 1]$, and so probability escapes at rate $z_{\lambda,G,\xi}/(1-t)$ for each such agent. This outflow is the first term in (24). Turning to inflows, if $\lambda_i = a$ then in state λ^{-i} agent i enters uniformly over time $[t, 1]$ and adopts blindly if one of her neighbors $j \in N_i(G)$ has adopted (the second term in (24)), and after inspecting with probability $F(\pi(y_{i,\xi_i}))$ if none of her neighbors has adopted (the third term in (24)). If $\lambda_i = b$, inflows from λ^{-i} are similarly captured by the fourth term in (24).

Given (24), the existence of a unique equilibrium follows from the Picard–Lindelöf theorem since the boundedness of f implies the system is Lipschitz.[28]

REMARK: The system of ODEs (24) implies equilibrium existence and uniqueness. But it is less useful as a tool to compute equilibrium numerically since there are $3^I \times 2^{I\times I} \times |\Xi|$ triples (λ, G, ξ), making it impossible to evaluate (24) numerically. Even if the network G is common knowledge, there are still 3^I states λ.

A.2. *Proof of Lemma 2*

We wish to study agent i's adoption decision given that no other agent in the $(I+1)$-clique has yet adopted. Using the notation of Appendix A.1, the probability that i observes at least one adoption when she enters at t_i equals $y = 1 - \Pr(\lambda_j \neq a \text{ for all } j \neq i|\lambda_i = \emptyset)$. When agent i enters and no other agent has yet adopted, she thus adopts with probability $F(\pi(y)) = \phi(0, y)$. Agents who have not adopted may either not have entered ($\lambda_j = \emptyset$) or decided not to adopt ($\lambda_j = b$). Let

$$z_\nu := \Pr(\lambda_j = \emptyset \text{ for } \nu \text{ others } j \neq i, \text{ and } \lambda_j = b \text{ for the other } I-\nu|\lambda_i = \emptyset).$$

Clearly, $y = 1 - \sum_{\nu=0}^{I} z_\nu$.

We now characterize $\{z_\nu\}$ recursively. For $\nu = I$, the probability that no one has entered is $z_I = (1-t)^I$. For states $\nu < I$, probability flows in from state $\nu+1$ as one of these agents enters and chooses not to adopt given that no one else had adopted; this rate is given by

[28]In fact, if we assume that time t is discrete rather than continuous, this proof shows more strongly that our game—much like most herding models—is dominance-solvable.

$\frac{\nu+1}{1-t}z_{\nu+1}(1-\phi(0,y))$. There is also outflow at rate $\frac{\nu}{1-t}z_\nu$ as one of the ν agents enters; this scales down the time-s inflow by a factor $(\frac{1-t}{1-s})^\nu$ at time-t. All told,

$$\begin{aligned} z_{\nu,t} &= \int_0^t \left(\frac{1-t}{1-s}\right)^\nu z_{\nu+1,s}\frac{\nu+1}{1-s}\big(1-\phi(0,y_s)\big)\,ds \\ &= (\nu+1)(1-t)^\nu \int_0^t z_{\nu+1,s}\frac{1-\phi(0,y_s)}{(1-s)^{\nu+1}}\,ds \end{aligned} \tag{25}$$

We claim that this can be inductively written as

$$z_{\nu,t} = \frac{I!}{\nu!}(1-t)^\nu \underbrace{\int_0^t\int_0^s\cdots\int_0^r}_{I-\nu \text{ integrals}}\big[\big(1-\phi(0,y_q)\big)\ldots\big(1-\phi(0,y_s)\big)\big]\,dq\ldots dr\,ds. \tag{26}$$

For $\nu = I$, this reduces to $z_{I,t} = (1-t)^I$. For the induction step, assume (26) holds for $\nu+1$ and substitute into (25). This becomes

$$\begin{aligned} z_{\nu,t} = (\nu+1)(1-t)^\nu \int_0^t \bigg[&\frac{I!}{(\nu+1)!}(1-s)^{\nu+1} \\ &\times \underbrace{\int_0^s\cdots\int_0^r}_{I-\nu-1 \text{ integrals}}\big[\big(1-\phi(0,y_q)\big)\ldots\big(1-\phi(0,y_r)\big)\big]\,dq\ldots dr\bigg]\frac{1-\phi(0,y_s)}{(1-s)^{\nu+1}}\,ds \end{aligned}$$

which collapses to (26).

The integration domain of (26) consists of all $(I-\nu)$-tuples $(q,\ldots,r,s)$ with $0\le q\le \cdots \le r\le s\le t$ and the integrand is symmetric in $(q,\ldots,r,s)$. Since there are $(I-\nu)!$ permutations of the integration variables, (26) equals

$$\begin{aligned} z_{\nu,t} &= \frac{I!}{\nu!(I-\nu)!}(1-t)^\nu \underbrace{\int_0^t\int_0^t\cdots\int_0^t}_{I-\nu \text{ integrals}}\big[\big(1-\phi(0,y_q)\big)\ldots\big(1-\phi(0,y_s)\big)\big]\,dq\,dr\ldots ds \\ &= \binom{I}{\nu}(1-t)^\nu\left(\int_0^t\big(1-\phi(0,y_s)\big)\,ds\right)^{I-\nu}. \end{aligned}$$

Summing over ν and using the binomial formula finishes the proof

$$1-y_t = \sum_{\nu=0}^I z_{\nu,t} = \left((1-t)+\int_0^t\big(1-\phi(0,y_s)\big)\,ds\right)^I = \left(1-\int_0^t\phi(0,y_s)\,ds\right)^I.$$

APPENDIX B: PROOFS FROM SECTION 3

B.1. *Proof of Proposition 2*

We break the proof into four steps:

(1) Define the branching process.

(2) Characterize the limit adoption $X_\theta(c)$ and the equilibrium limit adoption x^*_θ associated with the branching process.

(3) Show that the induced cutoffs (c_d^*) are a limit equilibrium by showing that learning curves $y_{I,d}^*$ in $\mathcal{G}_I$ converge to those associated with the branching process.
(4) Show that the limit equilibrium is unique.

While steps 1 and 2 may seem clear enough for the directed networks studied here, we spell them out in formal detail to prepare the ground for the more involved case of undirected networks with cliques in Proposition 2′.

B.1.1. *Directed Branching Process*

Here, we formalize the idea that the random network $\mathcal{G}_I$ locally approximates a tree for large I. Following Sadler (2020, Section 8.2), for any degree $d \in \mathbb{N}^\Theta$ we consider a multitype branching process $\mathcal{T}_d$ where offspring equals d in the first step, and is distributed according to D in all subsequent steps.[29] For any radius $r \in \mathbb{N}$, let $\mathcal{T}_{d,r}$ be the random rooted graph generated by $\mathcal{T}_d$, truncated after step r.

Turning to our finite networks, for agent i with degree d in network G, define *i's r-neighborhood* $G_{i,r}$ as the subgraph consisting of all nodes and edges in G that can be reached from i via paths of length at most r; e.g., for $r = 1$, $G_{i,r}$ consists of i, her outlinks, and her neighbors. Let $\mathcal{G}_{I,d,r}$ be the random rooted graph generated by realizing a network G from $\mathcal{G}_I$, choosing an agent i with degree d at random, and truncating G to i's r-neighborhood $G_{i,r}$.

We can now state our formal result, which mirrors Sadler's Lemma 1.

LEMMA 3: *$\mathcal{G}_{I,d,r}$ can be coupled to $\mathcal{T}_{d,r}$ with probability $\tilde{\rho}_{I,d,r}$, where $\lim_{I\to\infty} \tilde{\rho}_{I,d,r} = 1$ for all d, r.*

PROOF: We uncover the rooted graph $G_{i,r}$ following a breadth-first search: Start by connecting the d_θ outlinks of root i to randomly chosen type-θ nodes for all θ; then connect the outlinks of these neighbors, and so on until $G_{i,r}$ is realized. The coupling with the truncated branching process $T_{d,r}$ succeeds if at every step in this process, the respective type-θ outlink connects to a previously unvisited type-θ node; this could fail (i) if type-θ outlinks cannot be connected because no type-θ node was realized, or (ii) by realizing a self-link or multi-link (which we then prune from the network), or (iii) by realizing a link to a node that has already been visited (then $G_{i,r}$ is not a tree). Since the expected number of nodes in $G_{i,r}$ is finite, the chance of either of these three causes of failure (aggregated over all $|G_{i,r}|$ nodes) vanishes for large I. *Q.E.D.*

B.1.2. *Limit Adoption*

Here, we compute limit adoption probabilities of agents on an infinite random tree, generated by the branching process. Specifically, for any cost-cutoffs $c = (c_d)$ define the limit adoption probabilities as the solution $X_\theta(c)$ of the ODE

$$\dot{x}_\theta = E\left[\phi\left(1 - \prod_{\theta'}(1 - x_{\theta'})^{D_{\theta,\theta'}}, \pi^{-1}(c_{D_\theta})\right)\right]. \tag{27}$$

That is, when all agents in the branching process employ cost-cutoffs c, an agent with degree $d = (d_{\theta'})$ sees an adoption with probability $1 - \prod_{\theta'}(1 - x_{\theta'})^{d_{\theta'}}$, in which case she

[29] In contrast, Sadler (2020) uses the forward distribution $D' - 1$ to account for the friendship paradox in his undirected networks. We follow that approach in Appendix B.4.1.

adopts if $\kappa \leq c_d$. Taking expectations over D_θ yields (27). This nests as a special case the solution $x^*_\theta = X_\theta(c^*)$ of (10) for cost cutoffs $c^*_d := \pi(1 - \prod_{\theta'}(1 - x^*_{\theta'})^{d_{\theta'}})$.

B.1.3. *Limit Equilibrium*

We now turn to the proof of Proposition 2 proper, and show that $c^* = (c^*_d)$ is a limit equilibrium. In analogy to the limit probabilities $X_\theta(c)$ and x^*_θ, we write $Y_{I,d}(c)$ for the social learning curve in $\mathcal{G}_I$ when agents use cutoffs $c = (c_d)$, and $y^*_{I,d} = Y_{I,d}(c^*)$. Then c^* is a limit equilibrium iff $\lim_{I\to\infty} \pi(y^*_{I,d}) = c^*_d$ or equivalently, iff

$$\lim_{I\to\infty} y^*_{I,d} = 1 - \prod_\theta (1 - x^*_\theta)^{d_\theta}. \tag{28}$$

Given a finite number of agents I, equation (28) fails for the usual two reasons: correlating and backward links. We now show that these concerns vanish as I grows large. Lemma 3 showed that any agent i's neighborhood in G resembles a tree. We complement this argument by showing that i's social learning in our model only depends on i's neighborhood in G.

To formalize this statement, say that a path $i = i_0 \to i_1 \to \cdots \to i_r$ of length r is a *learning path* of agent i if $t_{i_{\nu-1}} > t_{i_\nu}$ for all $\nu = 1, \dots, r$; the chance of this is $1/(r+1)!$. Let $p_{\theta,r}$ be the probability that a type-θ node has no learning path of length r in the infinite random tree generated by the branching process. The expected number of length-r paths is bounded above by $(\max_\theta \sum_{\theta'} E[D_{\theta,\theta'}])^r$. Thus, $p_{\theta,r} \geq 1 - (\max_\theta \sum_{\theta'} E[D_{\theta,\theta'}])^r/(r+1)!$ and so $\lim_{r\to\infty} p_{\theta,r} = 1$. For an agent i with degree $d = (d_\theta)$, the probability of the event $\mathcal{E}$ that "all of i's neighbors have no learning path of length $r-1$" equals $p_{d,r} := \prod_\theta p^{d_\theta}_{\theta,r-1}$, with limit $\lim_{r\to\infty} p_{d,r} = 1$.

Turning from the branching process to the random network $\mathcal{G}_I$, note that the probability of event $\mathcal{E}$ depends on the network G only via i's r-neighborhood $G_{i,r}$. Thus, conditional on the coupling of $\mathcal{G}_{I,d,r}$ and $\mathcal{T}_{d,r}$ in Lemma 3, $p_{d,r}$ also equals the probability of $\mathcal{E}$ in $\mathcal{G}_I$. All told, write $\rho_{I,d,r} = \tilde{\rho}_{I,d,r} p_{d,r}$ for the joint probability that the coupling succeeds and of event $\mathcal{E}$. Then $\lim_{r\to\infty} \lim_{I\to\infty} \rho_{I,d,r} = \lim_{r\to\infty} p_{d,r} \lim_{I\to\infty} \tilde{\rho}_{I,d,r} = 1$.

We can now study adoption probabilities on i's neighborhood. Write $y^*_{I,d,r}$ for i's probability of observing an adoption, conditional on the intersection of three events: i's r-neighborhood being coupled to the branching process, i having d neighbors, and none of these neighbors having a learning path of length r. Similarly, write $x^*_{\theta,r}$ for the adoption probability of a type-θ agent in the branching process, conditional on her not having a learning path of length r. By construction, $y^*_{I,d,r} = 1 - \prod_\theta (1 - x^*_{\theta,r})^{d_\theta}$.

We now return to equation (28), which states that the social learning curve on $\mathcal{G}_I$ converges to the learning curve on the branching process. The triangle inequality implies

$$\begin{aligned}
\left| y^*_{I,d} - \left(1 - \prod_\theta (1 - x^*_\theta)^{d_\theta}\right)\right| &\leq \left|y^*_{I,d} - y^*_{I,d,r}\right| + \left| y^*_{I,d,r} - \left(1 - \prod_\theta (1 - x^*_{\theta,r})^{d_\theta}\right)\right| \\
&\quad + \left| \prod_\theta (1 - x^*_{\theta,r})^{d_\theta} - \prod_\theta (1 - x^*_\theta)^{d_\theta}\right| \\
&\leq (1 - \rho_{I,d,r}) + 0 + \sum_\theta d_\theta (1 - p_{\theta,r})
\end{aligned} \tag{29}$$

for any r. Since the LHS does not depend on r, we get

$$\limsup_{I\to\infty}\left|y^*_{I,d}-\left(1-\prod_\theta\left(1-x^*_\theta\right)^{d_\theta}\right)\right|\le\limsup_{r\to\infty}\limsup_{I\to\infty}(1-\rho_{I,d,r})+\sum_\theta d_\theta(1-p_{\theta,r})=0.$$

This implies (28) and thereby establishes that (c^*_d) is indeed a limit equilibrium.

B.1.4. *Uniqueness*

Uniqueness of the limit equilibrium follows immediately: Since the asymptotic independence of adoptions (28) does not rely on the optimality of the cutoffs (c^*_d), the same argument implies that for any cutoffs $c=(c_d)\neq c^*$,

$$\lim_{I\to\infty}Y_{I,d}(c)=1-\prod_\theta\left(1-X_\theta(c)\right)^{d_\theta}. \tag{30}$$

But since the solution to (10) is unique, we have $\pi(1-\prod_\theta(1-X_\theta(c))^{d_\theta})\neq c_d$. Thus, $\lim_{I\to\infty}\pi(Y_{I,d}(c))\neq c_d$, and so c is not a limit equilibrium.

B.2. *Example 5: Two Links vs. Infinite Chain*

When agent i has two uninformed neighbors j, k, each neighbor's adoption curve equals $x_j=x_k=F(\pi_0)t$. Hence the probability that at least one adopts is

$$y(t):=1-\left(1-F(\pi_0)t\right)^2. \tag{31}$$

With an infinite chain, agent i's social learning curve is given by $\dot{x}=\Phi(x)\le 1-(1-x)(1-F(\pi_0))$. Solving this ODE,

$$x_t\le\zeta(t):=1-\frac{1-F(\pi_0)e^{(1-F(\pi_0))t}}{1-F(\pi_0)}. \tag{32}$$

We wish to show that $y(t)-\zeta(t)\ge 0$ for all t. It suffices to show this inequality for $t=1$; this follows since $y(t)-\zeta(t)$ is concave and $y(0)-\zeta(0)=0$.

Setting $t=1$, abbreviating $\delta:=1-F(\pi_0)\in(0,1]$, and multiplying by δ, we thus wish to show that $\xi(\delta):=1-(1-\delta)e^\delta-\delta^3\ge 0$. Differentiating, one can see that ξ has a unique local extremum δ^* on $[0,1]$ and that $\xi''(\delta^*)\le 0$. Thus, it is quasi-concave with $\xi(0)=\xi(1)=0$, and hence is positive everywhere.

B.3. *The Single-Crossing Lemma*

We use the following version of Milgrom and Weber (1982, Lemma 2) that allows for weak inequality at the initial condition.

LEMMA 4: *Let (x_t), $(\tilde{x}_t)$ solve $\dot{x}=\psi(x)$ and $\dot{\tilde{x}}=\tilde{\psi}(\tilde{x})$ with $x_0=\tilde{x}_0=0$, where $\psi(x)>\tilde{\psi}(x)>0$ for all $x\in(0,1]$ and $\psi(0)=\tilde{\psi}(0)>0$. Then $x_t>\tilde{x}_t$ for all $t>0$.*

PROOF: For any $\epsilon>0$, define $x^\epsilon:[\epsilon,1]\to\mathbb{R}$ as the solution of $\dot{x}^\epsilon=\psi(x^\epsilon)$ with initial condition $x^\epsilon_\epsilon=\tilde{x}_\epsilon$. Then $x^\epsilon_t>\tilde{x}_t$ for all $t>\epsilon$ by Milgrom and Weber (1982, Lemma 2).

Since the solution of a differential equation is continuous in its initial conditions, we have $\lim_{\epsilon\to 0}x^\epsilon_t=x_t$ and so $x_t\ge\tilde{x}_t$ for all $t>0$. But $x_t\ge\tilde{x}_t>0$ implies $x_{t'}>\tilde{x}_{t'}$ for all $t'>t$, and so we get $x_t>\tilde{x}_t$ for all $t>0$. *Q.E.D.*

B.4. *Proof of Proposition 2′*

The proof of Proposition 2′ mirrors the proof Proposition 2. Instead of repeating all of the arguments, we only discuss where they need to be adapted.

B.4.1. *Undirected Branching Process*

Define a two-type branching process with bilateral and triangle types. In the first step, the number of offspring are given by some fixed degree $(\bar{d}, 2\hat{d})$. In every subsequent step, the (forward) degree is drawn from $(\bar{D}' - 1, 2\hat{D})$ for bilateral offspring, and from $(\bar{D}, 2(\hat{D}' - 1))$ for triangle offspring; all draws are independent and generate distinct nodes, including the draws from two triangle offspring on the same triangle. The resulting undirected network consists of the tree generated by the branching process and the links connecting neighbors on any given triangle.

Next, we couple the r-neighborhoods[30] of agent i with degree $(\bar{d}, 2\hat{d})$ in the finite network and the branching process, $\mathcal{G}_{I,(\bar{d},2\hat{d}),r}$and $\mathcal{T}_{(\bar{d},2\hat{d}),r}$. This is where we have to account for the friendship paradox: When uncovering i's neighbor j on a bilateral link, the probability distribution of j's bilateral degree $\Pr(\bar{D}_j = d)$ must be reweighted by $d/E[\bar{D}]$; that is, it is drawn from D', defined in (12). Also, one of j's D' bilateral links goes back to i, and so only $\bar{D}' - 1$ go forward to additional nodes. Since j's bilateral and triangle links $\bar{D}_j$, $\hat{D}_j$ are independent, $\hat{D}_j$ simply follows $\hat{D}$. All told, conditional on a successful coupling, the "forward-degree" of a bilateral neighbor follows $(\bar{D}' - 1, 2\hat{D})$. The argument that the degree of a triangle neighbor follows $(\bar{D}, 2(\hat{D}' - 1))$ is analogous.

B.4.2. *Limit Adoption*

Following Section B.1.2, we now characterize neighbors' adoption probabilities in the infinite network generated by the branching process for arbitrary strategies $c = (c_{\bar{d},\hat{d}})$. Indeed, let $\bar{X}(c)$, $\hat{X}(c)$ be the solution of

$$\dot{\bar{x}} = E\big[\phi\big(1 - (1 - \bar{x})^{\bar{D}'-1}(1 - \hat{x})^{2\hat{D}}, \pi^{-1}(c_{\bar{D}',\hat{D}})\big)\big], \tag{33}$$

$$\dot{\hat{x}} = E\big[\phi\big(1 - (1 - \bar{x})^{\bar{D}}(1 - \hat{x})^{2(\hat{D}'-1)}, \pi^{-1}(c_{\bar{D},\hat{D}'})\big)\big]. \tag{34}$$

We claim that (i) $\bar{X}(c)$ is the adoption probability of i's bilateral neighbor j at t_i, and (ii) $\hat{X}(c)$ is the first adopter probability in a triangle (i, j, k); more precisely, $(1 - \hat{X}(c))^2$ is the probability that neither j nor k have adopted at t_i.

Claim (i) follows by the standard argument that j has D' neighbors, but that includes i, who has not yet adopted at $t < t_i$. Claim (ii) is more subtle and follows by the proof of Lemma 2. As in that proof define z_ν as the probability that $\nu \in \{0, 1, 2\}$ of i's neighbors j, k in a given triangle have not yet entered, while the other $2 - \nu$ have entered but chose not to adopt. In the triangle, when one of the remaining ν neighbors enters she observes no adoption, and hence adopts herself with probability $\phi(0, y)$, where $y := 1 - \sum_{\nu=0}^{2} z_\nu$. Here, a triangular neighbor j has additional $\bar{D}$ bilateral links and $\hat{D}' - 1$ additional triangular link pairs, so observes an adoption with probability $1 - (1 - \bar{x})^{\bar{D}}(1 - \hat{x})^{2(\hat{D}'-1)}$, and

[30] In undirected networks, we define an agent's r-neighborhood of agent i as all agents and all undirected edges that can be reached from agent i via paths of length at most r.

is assumed to employ the exogenous cutoffs $c_{\bar{D},\hat{D}'}$ upon seeing no adoption. Thus, from i's perspective, j's adoption rate is given by the RHS of (34). Subject to substituting this term for $\phi(0, y)$, the proof of Lemma 2 applies as stated, yielding (34).

The solution $\bar{X}(c)$, $\hat{X}(c)$ of (33)–(34) nests as a special case the solution $\bar{x}^* = \bar{X}(c^*)$, $\hat{x}^* = \hat{X}(c^*)$ of (18)–(19) for cost cutoffs $c^*_{\bar{d},\hat{d}} = \pi(1-(1-\bar{x}^*)^{\bar{d}}(1-\hat{x}^*)^{2\hat{d}})$.

B.4.3. *Limit Equilibrium, Uniqueness, and Limit of Equilibria*

The arguments in Sections B.1.3, B.1.4, and D.1 only require adapting the notation. Indeed, write $Y_{I,\bar{d},\hat{d}}(c)$ for the social learning curve in $\hat{\mathcal{G}}_I$ when agents use cutoffs $c = (c_{\bar{d},\hat{d}})$, and $y^*_{I,\bar{d},\hat{d}} = Y_{I,\bar{d},\hat{d}}(c^*)$. Then, at the most general level, (30) generalizes to

$$\lim_{I\to\infty} \sup_c \left| Y_{I,\bar{d},\hat{d}}(c) - \left[1 - \left(1-\bar{X}(c)\right)^{\bar{d}}\left(1-\hat{X}(c)\right)^{2\hat{d}}\right]\right| = 0. \tag{35}$$

The fact that c^* is a limit equilibrium then follows by substituting c^* into (35) and recalling that $c^*_{\bar{d},\hat{d}} = \pi(1-(1-\bar{x}^*)^{\bar{d}}(1-\hat{x}^*)^{2\hat{d}})$.

Uniqueness follows by substituting any other cutoffs $c \neq c^*$ into (35) and noting that $c_{\bar{d},\hat{d}} \neq \pi(1-(1-\bar{X}(c))^{\bar{d}}(1-\hat{X}(c))^{2\hat{d}})$.

Finally, the exact equilibria c_I of $\hat{\mathcal{G}}_I$ converge to the limit equilibrium c^* by the same reasoning as in Appendix D.1 in the Online Supplementary Material, invoking Arzela–Ascoli to obtain a convergent subsequence of c_I and then using (35) to show that its limit must equal c^*.

APPENDIX C: IMPERFECT SOCIAL LEARNING

C.1. *Proof of Lemma 1′*

Given $\pi_0 = 1/2$, the unconditional probability of (d, a) is $y_{d,a} = (y^H_{d,a} + y^L_{d,a})/2$, and equation (20) implies that $y^H_{d,a} = 2\pi_{d,a}y_{d,a}$ and $y^L_{d,a} = 2(1-\pi_{d,a})y_{d,a}$. Thus, we can rewrite (21) as

$$\dot{x}^H_d = 2\alpha^H \sum_{a\le d} y_{d,a}\left[\pi_{d,a}F\left(c(\pi_{d,a})\right)\right] = 2\alpha^H E\left[\Pi_d F\left(c(\Pi_d)\right)\right], \tag{36}$$

$$\dot{x}^L_d = 2\alpha^L \sum_{a\le d} y_{d,a}\left[(1-\pi_{d,a})F\left(c(\pi_{d,a})\right)\right] = 2\alpha^L E\left[(1-\Pi_d)F\left(c(\Pi_d)\right)\right], \tag{37}$$

where the expectation $E[\cdot]$ is taken over the realizations $\pi_{d,a}$ of the random posterior Π_d. Given assumption (22), a mean-preserving spread of Π_d raises the RHS of (36) and lowers the RHS of (37).

C.2. *Proof of Theorem 1′*

For part (a), we discretize time $\{t\Delta\}$ for $t = 0, 1, 2, \ldots$ and argue by induction over t. That is, we interpret (21) as a finite difference equation, establish the Blackwell-rankings for any Δ, and then conclude by taking the continuous time limit $\Delta \to 0$.

At $t = 0$, there is no social information, so the Blackwell-ranking obtains (weakly). Now assume by induction that social information is ranked by $\tilde{Y}_{d,t} \succeq_{\text{BW}} Y_{d,t}$ at all times $t < s$.

We first argue that adoption rates then obey the same ranking

$$\dot{\tilde{x}}^H_{\theta,t} = E\big[\dot{\tilde{x}}^H_{\tilde{D}_\theta,t}\big] \geq E\big[\dot{x}^H_{\tilde{D}_\theta,t}\big] \geq E\big[\dot{x}^H_{D_\theta,t}\big] = \dot{x}^H_{\theta,t}.$$

The first (weak) inequality follows by induction and (the discrete-time version of) Lemma 1′. The second inequality follows also by Lemma 1′, together with the inequality $\tilde{D}_\theta \succeq_{\text{FOSD}} D_\theta$ and the fact that observing (FOSD) fewer neighbors amounts to sometimes losing some adoption signals, which Blackwell-decreases social information; it is strict when $\tilde{D}_\theta \succ_{\text{FOSD}} D_\theta$. Integrating over $t < s$, we get $\tilde{x}^H_{\theta,s} \geq x^H_{\theta,s}$, with strict inequalities if $\tilde{D}_\theta \succ_{\text{FOSD}} D_\theta$. The analogous argument implies $\tilde{x}^L_{\theta,s} \leq x^L_{\theta,s}$.

We now show that observing type-θ's time-s adoption in the network generated by $\tilde{D}$ is Blackwell-sufficient for observing this under D. The above imply

$$\frac{\tilde{x}^H_{\theta,s}}{\tilde{x}^L_{\theta,s}} \geq \frac{x^H_{\theta,s}}{x^L_{\theta,s}} \quad \text{and} \quad \frac{1-\tilde{x}^H_{\theta,s}}{1-\tilde{x}^L_{\theta,s}} \leq \frac{1-x^H_{\theta,s}}{1-x^L_{\theta,s}}. \tag{38}$$

Denote by $\Pi_{\theta,s}$ the random posterior of an agent in network D with one type-θ neighbor and no other neighbors, and by $\pi_{\theta,1,s}$, $\pi_{\theta,0,s}$ the realized posteriors upon observing that type-θ neighbor adopt/not-adopt; analogously define $\tilde{\Pi}_{\theta,s}$ and $\tilde{\pi}_{\theta,1,s}$, $\tilde{\pi}_{\theta,0,s}$ for network $\tilde{D}$. Bayes' rule (20) and equation (38) imply $\tilde{\pi}_{\theta,1,s} \geq \pi_{\theta,1,s} \geq \pi_{\theta,0,s} \geq \tilde{\pi}_{\theta,0,s}$, so $\tilde{\Pi}_{\theta,s}$ is a mean-preserving spread of $\Pi_{\theta,s}$ and observing type-θ's adoption in the network generated by $\tilde{D}$ is Blackwell-sufficient for observing this under D. Since adoption is independent across neighbors, we get $\tilde{Y}_{d,s} \succeq_{\text{BW}} Y_{d,s}$, concluding the induction step.

Part (b) follows from $\tilde{Y}_{\tilde{D}_\theta,s} \succeq_{\text{BW}} Y_{\tilde{D}_\theta,s} \succeq_{\text{BW}} Y_{D_\theta,s}$. The first inequality follows by substituting the random degree $\tilde{D}_\theta$ into part (a). The second inequality follows by our above argument that observing (FOSD) fewer neighbors amounts to sometimes losing some adoption signals.

C.3. *Proof of Proposition 3*

(a) and (b). We follow the proof of Theorem 1′. Let $\tilde{\alpha}^H \geq \alpha^H$ and $\tilde{\alpha}^L \leq \alpha^L$. We must show that $\tilde{Y}_d \succeq_{\text{BW}} Y_d$. At $t = 0$, there is no social information, so the Blackwell-ranking obtains (weakly). Now assume by induction that $\tilde{Y}_{d,t} \succeq_{\text{BW}} Y_{d,t}$ for all $t < s$. Since there is more information, $\tilde{\alpha}^H \geq \alpha^H$, and $\tilde{\alpha}^L \leq \alpha^L$, equation (21) and the proof of Lemma 1′ imply that $\tilde{x}^H_{\theta,s} \geq x^H_{\theta,s}$ and $\tilde{x}^L_{\theta,s} \leq x^L_{\theta,s}$. Then, as in Theorem 1′, observing type-θ's adoption state is Blackwell-better for adoption rates $\tilde{\alpha}^q$ compared to α^q. By independence across neighbors, $\tilde{Y}_{d,s} \succeq_{\text{BW}} Y_{d,s}$, concluding the induction step.

(c) As in parts (a,b), suppose by induction that $\tilde{Y}_{d,t} \succeq_{\text{BW}} Y_{d,t}$ for all times $t < s$. Since $\tilde{\alpha}^H/\alpha^H = \tilde{\alpha}^L/\alpha^L =: \xi > 1$, the proof of Lemma 1′ implies $\dot{\tilde{x}}^H_{\theta,t} \geq \xi \dot{x}^H_{\theta,t}$ and $\dot{\tilde{x}}^L_{\theta,t} \leq \xi \dot{x}^L_{\theta,t}$, and so $\tilde{x}^H_{\theta,s} > x^H_{\theta,s}$ and $\tilde{x}^H_{\theta,s}/\tilde{x}^L_{\theta,s} \geq x^H_{\theta,s}/x^L_{\theta,s}$. Moreover, $(1 - x^H_\theta)/(1 - x^L_\theta) = (1/x^H_\theta - 1)/(1/x^H_\theta - x^L_\theta/x^H_\theta)$ falls in both x^H_θ and in x^H_θ/x^L_θ, and so $(1 - \tilde{x}^H_{\theta,s})/(1 - \tilde{x}^L_{\theta,s}) \leq (1 - x^H_{\theta,s})/(1 - x^L_{\theta,s})$. Thus, observing whether or not a type-θ neighbor has adopted is more informative under $\tilde{\alpha}^q$ compared to α^q.

C.4. *Backward Links*

Compare the directed and undirected pair in Example 3. With $i \to j$, agent j has no information, and his adoption follows $\dot{x}^q_j = \alpha^q F(c(\pi_0))$. With $i \leftrightarrow j$, define $\bar{x}^q$ to be j's adoption probability conditional on quality q and $t \le t_i$. His posterior upon observing no adoption becomes $\underline{\pi} := (1 - \bar{x}^H)\pi_0/((1 - \bar{x}^H)\pi_0 + (1 - \bar{x}^L)(1 - \pi_0)) < \pi_0$. Analogous to (13), $\dot{\bar{x}}^q = \alpha^q F(c(\underline{\pi}))$.

We claim that agent i has Blackwell-more information in the directed pair than in the undirected pair. Indeed, $\bar{x}^q = \zeta x_j^q$ for $\zeta := F(c(\underline{\pi}))/F(c(\pi_0)) < 1$. That is, observing j in the undirected pair is like observing him in the directed pair and losing the signal with probability $1 - \zeta$, and hence Blackwell-inferior.

C.5. *Correlating Links*

As in Example 6, suppose agent i initially observes agents j, k, and consider the effect of an additional link $j \to k$. Suppose agents' cost is equi-likely low or high, $\kappa \in \{0, \bar{\kappa}\}$, with $\bar{\kappa} = c(\alpha^H \pi_0/[\alpha^H \pi_0 + \alpha^L(1 - \pi_0)])$, and suppose the agent inspects if indifferent. The low-cost agent always inspects, and information has no value to her. The high-cost agent only benefits from social information if it pushes her posterior strictly above $\bar{\kappa}$, which only ever happens in either network when i observes both j and k adopt. Thus, we need to compare the (unconditional) probability of this event $x_{\{j,k\},i} = \pi_0 x^H_{\{j,k\},i} + (1 - \pi_0)x^L_{\{j,k\},i}$ and the induced posterior belief $\pi_{\{j,k\},i}$ across the two networks. Given $\pi_0 = 1/2$, the value of information then equals $\frac{1}{2}x_{\{j,k\},i}(\pi_{\{j,k\},i} - \bar{\kappa})$.

In either network, k has no information and only her low-cost type inspects; the probability that she adopts good q by time t thus equals $\frac{1}{2}\alpha^q t$. Without the correlating link, by symmetry and independence of j and k, the probability they both adopt product q equals $x^q_{\{j,k\},i} = (\frac{1}{2}\alpha^q t)^2$ and the posterior belief equals $\pi_{\{j,k\},i} = (\alpha^H)^2\pi_0/[(\alpha^H)^2\pi_0 + (\alpha^L)^2(1 - \pi_0)]$.

The correlating link raises j's adoption probability conditional on k having adopted: If j enters first, only low-cost j adopts and the probability is unchanged; but if k enters first, j inspects with certainty. Thus, the joint probability that both j and k adopt rises to $\hat{x}^q_{\{j,k\},i} = \frac{1}{2}\alpha^q t \cdot \frac{3}{4}\alpha^q t$, while the posterior belief is unchanged, $\hat{\pi}_{\{j,k\},i} = \pi_{\{j,k\},i}$. All told, the correlating link $j \to k$ is valuable to i because it increases the probability of observing j adopt in the event that k also adopts, which is precisely when this information is most valuable to i. Note the contrast to Example 6, where j's adoption was valuable to i in the complementary event, when k had not adopted.

REFERENCES

ACEMOGLU, D., M. A. DAHLEH, I. LOBEL, AND A. OZDAGLAR (2011): "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78 (4), 1201–1236. [2601,2603,2621]

AKBARPOUR, M., S. MALLADI, AND A. SABERI (2020): "Just a few Seeds More: Value of Network Information for Diffusion," Working paper, SSRN, https://ssrn.com/abstract=3062830. [2624]

ALI, S. N. (2018): "Herding With Costly Information," *Journal of Economic Theory*, 175, 713–729. [2604,2620, 2621]

BANERJEE (1993): "The Economics of Rumours," *Review of Economic Studies*, 107 (2), 309–327. [2604]

BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013): "The Diffusion of Microfinance," *Science*, 341 (6144), 1236498. [2602,2623]

BANERJEE, A. V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107 (3), 797–817. [2603]

BASS, F. M. (1969): "A New Product Growth for Model Consumer Durables," *Management Science*, 15 (5), 215–227. [2601,2604]

BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100 (5), 992–1026. [2603,2605]

BOARD, S., AND M. MEYER-TER-VEHN (2021): "Supplement to 'Learning Dynamics in Social Networks'," *Econometrica Supplemental Material*, 89, https://doi.org/10.3982/ECTA18659. [2612]

BOHREN, J. A., AND D. N. HAUSER (2021): "Learning With Heterogeneous Misspecified Models: Characterization and Robustness," Econometrica. (forthcoming). [2618,2624]

BÖRGERS, T. (2021): "Notes on Relations Among Experiments," Working paper, University of Michigan. [2622]

CAI, J., AND A. SZEIDL (2018): "Interfirm Relationships and Business Performance," *Quarterly Journal of Economics*, 3 (133), 1229–1282. [2602]

CAMPBELL, A. (2013): "Word-of-Mouth Communication and Percolation in Social Networks," *American Economic Review*, 103 (6), 2466–2498. [2604,2624]

CELEN, B., AND S. KARIV (2004): "Observational Learning Under Imperfect Information," *Games and Economic Behavior*, 47 (1), 72–86. [2607]

CENTOLA, D. (2010): "The Spread of Behavior in an Online Social Network Experiment," *Science*, 329 (5996), 1194–1197. [2623]

CHAMLEY, C., AND D. GALE (1994): "Information Revelation and Strategic Delay in a Model of Investment," Econometrica, 1065–1085. [2624]

CHANDRASEKHAR, A. G., AND M. JACKSON (2021): "A Network Formation Model Based on Subgraphs," Working paper, ArXiV, https://arxiv.org/abs/1611.07658. [2619]

CHANDRASEKHAR, A. G., H. LARREGUY, AND J. P. XANDRI (2020): "Testing Models of Social Learning on Networks: Evidence From Two Experiments," *Econometrica*, 88 (1), 1–32. [2618]

CONLEY, T. G., AND C. R. UDRY (2010): "Learning About a New Technology: Pineapple in Ghana," *American Economic Review*, 100 (1), 35–69. [2623]

DASARATHA, K., AND K. HE (2020): "Network Structure and Naive Sequential Learning," *Theoretical Economics*, 15 (2), 415–444. [2618]

ENKE, B., AND F. ZIMMERMANN (2019): "Correlation Neglect in Belief Formation," *Review of Economic Studies*, 86 (1), 313–332. [2618]

EYSTER, E., AND M. RABIN (2010): "Naive Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2 (4), 221–243. [2618]

——— (2014): "Extensive Imitation Is Irrational and Harmful," *Quarterly Journal of Economics*, 129 (4), 1861–1898. [2603,2617,2618]

EYSTER, E., M. RABIN, AND G. WEIZSÄCKER (2018): "An Experiment on Social Mislearning," Rationality and competition discussion paper series, no. 73, CRC TRR 190 Rationality and Competition. [2618]

GOLUB, B., AND M. O. JACKSON (2012): "How Homophily Affects the Speed of Learning and Best-Response Dynamics," *Quarterly Journal of Economics*, 127 (3), 1287–1338. [2602]

GOLUB, B., AND E. SADLER (2016): "Learning in Social Networks," in *The Oxford Handbook of the Economics of Networks*, Chapter 19, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers. Oxford: Oxford University Press, 504–542. [2604]

GUARINO, A., H. HARMGART, AND S. HUCK (2011): "Aggregate Information Cascades," *Games and Economic Behavior*, 73 (1), 167–185. [2604,2620]

HANN-CARUTHERS, W., V. V. MARTYNOV, AND O. TAMUZ (2018): "The Speed of Sequential Asymptotic Learning," *Journal of Economic Theory*, 173, 383–409. [2605]

HENDRICKS, K., A. SORENSEN, AND T. WISEMAN (2012): "Observational Learning and Demand for Search Goods," *American Economic Journal: Microeconomics*, 4 (1), 1–31. [2604,2620,2621]

HERRERA, H., AND J. HÖRNER (2013): "Biased Social Learning," *Games and Economic Behavior*, 80, 131–146. [2604,2620]

JACKSON, M. O. (2010): *Social and Economic Networks*. Princeton University Press. [2611]

LOBEL, I., AND E. SADLER (2015): "Information Diffusion in Networks Through Social Learning," *Theoretical Economics*, 10 (3), 807–851. [2604]

LOMYS, N. (2020): "Sequential Collective Search in Networks," Working paper, SSRN, https://ssrn.com/abstract=3197244. [2604]

MCADAMS, D., AND Y. SONG (2020): "Viral Social Learning," Working paper, SSRN, https://ssrn.com/abstract=3299129. [2604]

MILGROM, P., AND R. WEBER (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1089–1122. [2629]

MONZÓN, I., AND M. RAPP (2014): "Observational Learning With Position Uncertainty," *Journal of Economic Theory*, 154, 375–402. [2604]

MORETTI, E. (2011): "Social Learning and Peer Effects in Consumption: Evidence From Movie Sales," *Review of Economic Studies*, 78 (1), 356–393. [2623]

MORRIS, S. (2000): "Contagion," *Review of Economic Studies*, 67 (1), 57–78. [2604,2624]

MUELLER-FRANK, M., AND M. M. PAI (2016): "Social Learning With Costly Search," *American Economic Journal: Microeconomics*, 8 (1), 83–109. [2604,2620]

SADLER, E. (2020): "Diffusion Games," *American Economic Review*, 110 (1), 225–270. [2604,2611,2627]
SMITH, L., AND P. SØRENSEN (1996): "Rational Social Learning With Random Sampling," Working paper, SSRN, https://ssrn.com/abstract=1138095. [2603]
——— (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68 (2), 371–398. [2603]

Co-editor Bart Lipman handled this manuscript.

Manuscript received 6 July, 2020; final version accepted 30 May, 2021; available online 3 June, 2021.