



DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES

Inducing Herding with Capacity Constraints

Alexei Parakhonyaky and Nick Vikander

Number 808
October, 2016

Manor Road Building, Oxford OX1 3UQ

Inducing Herding with Capacity Constraints*

Alexei Parakhonyak[†] and Nick Vikander[‡]

October 2016

Abstract

This paper shows that a firm may benefit from restricting capacity so as to trigger herding behavior from consumers, in situations where such behavior is otherwise unlikely. We consider a setting with social learning, where consumers observe sales from previous cohorts and update beliefs about product quality before making their purchase. A capacity constraint directly limits sales but also results in coarser information: upon observing a sellout, consumers attach positive probability to all levels of demand that exceed the constraint. The resulting discrete jump in beliefs following a sellout benefits the firm, and can make it optimal to restrict capacity.

JEL-codes: D82, D83, L15

Key Words: Capacity Constraints, Herding, Informational Cascades

*This paper has benefited from discussions with Itai Arieli, Vincent Crawford, Ignacio Monzón, Gosia Poplawska, Larry Samuelson, Peter Norman Sørensen, Marta Troya-Martinez, and participants of the Oxford Economic Theory Workshop, Game Theory Society World Congress, EARIE 2016, and Transatlantic Theory Workshop.

[†]University of Oxford, Department of Economics and Lincoln College. E-mail: aleksei.parakhonyak@economics.ox.ac.uk

[‡]University of Copenhagen, Department of Economics. E-mail: nick.vikander@econ.ku.dk

1 Introduction

The purpose of this paper is to show that firms may want to strategically restrict capacity in order to influence consumer learning, specifically in a way that triggers positive purchase cascades. As is standard in the literature on social learning (see foundational papers by Banerjee (1992) and Bikhchandani et al. (1992)), each consumer receives a noisy private signal about the quality of the product or service, but also infers some information from observing sales from other consumers. To illustrate the mechanism by which capacity affects consumer beliefs, consider ten consumers who visit a firm. Suppose, as in SgROI (2002), that each consumer follows her own private signal, and hence buys if and only if that signal was good. Another consumer who arrives in a later round and observes initial sales of three may well refuse to buy, even if her own signal was good, because she infers that only three out of the ten others had a good signal. But if the firm had initially limited its capacity to three sales per period, then the consumer would observe a sellout, and infer that *at least* three out of the ten other signals were good. That may well be sufficiently good news for the consumer to buy, even if her own signal was bad.

This example demonstrates how restricting capacity can influence consumer learning in a way that benefits the firm. For exactly the same history, a firm without the capacity constraint ends up in a negative cascade (no new arrivals want to buy), whereas a firm with the capacity constraint ends up in a positive cascade (all new arrivals want to buy). In the latter case consumers want to buy because the sellout makes them more optimistic about product quality. However, this potential positive change in beliefs comes at a cost, since the capacity-constrained firm must turn some consumers away, because it cannot serve the whole market. Our analysis focuses on this trade-off. In particular, we are interested in whether a firm might find it optimal to restrict capacity, even though limited capacity may lead to rationing.

There is certainly evidence of a variety of products where demand can far outstrip supply.

Debo et al. (2012) mention that consumers faced significant waiting times for Cabbage Patch Kids in 1983 and Beanie Babies in the 1990s, when these products were launched, and suggest this was a result of strategic firm decisions. In the United Kingdom, Boots anti-ageing No.7 serum attracted such long queues that it needed to be rationed to consumers.¹ High-end restaurants also often have very long waiting lists. “Noma” in Copenhagen opens up bookings for dates three months in advance, and tables are still notoriously difficult to come by. Reported waiting times to dine at “Damon Baehrel” or “Club 33” in the United States range from ten to fourteen years. Music festivals, such as those in Roskilde and Reading, also often sell out well in advance, with tickets for Glastonbury 2016 selling out in just 30 minutes.² The analysis in this paper suggests that these type of situations, with restricted capacity and excess demand, can arise from seller profit-maximizing behavior.

Our paper is closely related to a growing literature on social learning. In our paper, like in Arieli (2016), we assume that consumers arrive in cohorts, or generations, and observe only the actions of their predecessors.³ The set of observable actions (i.e. whether earlier consumers demanded the product) is neither exogenous, as in most of the literature, nor stochastic. It instead depends in part on whether the firm chose to restrict capacity. The use of a capacity constraint effectively coarsens the information available to consumers, since they cannot infer the exact number of earlier consumers who demanded the product if that number exceeds capacity.⁴

We show that the firm can sometimes profit from imposing a capacity constraint, even though this potentially limits total sales, because it can also increase the probability of a

¹See <http://www.nottinghampost.com/miracle-cream/story-21088010-detail/story.html>.

²See <http://www.glastonburyfestivals.co.uk/glastonbury-2016-tickets-sell-out-in-30-minutes/>

³The idea that agents arrive in cohorts can also be found in Banerjee and Fudenberg (2004) and in Monzón (forthcoming)

⁴An implication is that the set of observed consumer actions and their signals are not independent in our setting, as they are in Acemoglu et al. (2011) where the network topology does not depend on the actions of consumers.

positive purchase cascade. That is a situation where consumers all buy regardless of their own private signal. Such cascades can occur in our setting because consumers have bounded private beliefs, i.e. a single private signal can never lead consumers to always ignore public information, no matter how strong that public information is. As is known from Smith and Sørensen (2000) and Acemoglu et al. (2011), bounded private beliefs present the possibility for “pathological” outcomes, where all consumers might herd on the wrong action.⁵ This is precisely the mechanism which the firm exploits by limiting capacity. By selling out, the firm is interested to start a cascade which it finds favourable, but is not necessarily favorable for consumers, who may all buy even though product quality is low.

This strategic aspect of the firm’s capacity choice is helpful in understanding our results. In particular, we show that the firm typically benefits from setting a capacity constraint in situations which a priori look grim, i.e. when prior beliefs about quality are quite low, and signals are quite accurate. The flip side is that if the firm expects that on average more than half of consumers will receive positive signals, then it will never want to restrict capacity. The reason is quite straightforward: selling to more than half of consumers is already sufficient to trigger a positive cascade, so there is no reason to further improve this probability by restricting capacity, given that doing so also limits total sales. We show that the decision to set up the constraint is monotone in the discount factor, so a more patient firm has a large incentive to restrict capacity. This result follows naturally from the fact that a capacity constraint always reduces immediate expected profits, and all potential gains from a cascade come in future periods.

Unfortunately, due to the combinatorial nature of the problem, a precise characterization of the optimal capacity is very challenging. Two conditions must be satisfied for a constraint to be optimal: it should deliver higher profits than any other constraint, and it

⁵Smith and Sørensen (2000) and Acemoglu et al. (2011) both consider settings where individual agents arrive sequentially. Monzón (forthcoming) shows agents may herd on the wrong action when they arrive in successive cohorts, in a setting with aggregate uncertainty.

must be more profitable than remaining unconstrained. Depending on the circumstances, the optimal constraint may induce a sufficient shift in beliefs to trigger a positive cascade following a single sellout, or such a cascade may only occur after the firm sells out multiple times.⁶ Nonetheless, we show that for any capacity constraint, there are situations where setting precisely this constraint yields higher profits than remaining unconstrained, and also generates a positive cascade after a single sell out. We also show by means of numerical analysis that profits can be non-monotone in the level of capacity.

Turning to the literature, other work on social learning has touched on related concerns about imperfect information over the history of play (Acemoglu et al. (2011) and Monzón and Rapp (2014)) or congestion in learning (Eyster et al. (2014)). The closest paper in this literature, and one of the few that looks at strategic manipulation of the information available to consumers, is Debo et al. (2012). In their model, the firm knows its own quality, and chooses a service speed for consumers who arrive according to a Poisson process. Consumers are either informed and know product quality or are uninformed and do not. A long queue suggests to uninformed customers that informed consumer demand is high (and hence so is quality), giving the firm an incentive to slow down service to influence these beliefs. However, if the service is too slow, the queue becomes too long and consumers are driven away.

The mechanism in Debo et al. (2012) is similar to ours in its broad lines; restricting capacity and slowing down service both reduce immediate profits, but also influence consumer beliefs in a way that can increase profits in the future. That being said, there are important differences. First, the focus of their analysis is on quality signaling, whereas we assume the firm does not know its quality when setting the constraint. Second, consumers in our setting are not fully informed or uninformed, but instead receive boundedly informative private signals. Third, the mechanism of increasing profits in their paper is just the opposite to

⁶Suppose for example that one hundred consumers visit a firm that has set a capacity of one. A single sellout then only implies that at least one of the consumers had a positive signal. This will have little impact on other consumers' beliefs, and is unlikely to make them ignore their private signals.

ours: they show how longer queues provide consumers with more information, while in our case restricting capacity provides consumers with less.⁷ Finally, consumer decision-making in our paper is solely based on inferred quality and does not depend on any cost of waiting.

Bose et al. (2006) and Bose et al. (2008) both analyze how dynamic pricing, rather than restricting capacity, can influence consumer learning, in a setting where individual consumers are served sequentially. They show how the firm’s choice of price affects what consumers can infer about product quality from observing each others’ purchases. In our setting, a capacity constraint also affects consumer inference, but in a very different way: it explicitly hides information from consumers in precisely those situations where demand exceeds capacity, so that consumers become more optimistic after observing a sellout. We do not explicitly consider pricing in our analysis, effectively assuming that any fixed price for the service is reflected in the value of the consumers’ outside option. Allowing the seller to optimize over both price and capacity would complicate the analysis, but would not change the basic mechanism by which sellouts can increase subsequent demand.

Our analysis also shares certain features with the literature on Bayesian Persuasion that has followed from Rayo and Segal (2010) and Kamenica and Gentzkow (2011). Typically, in these models, a Sender faces a Receiver, where the Sender’s payoff depends on the Receiver’s action, and the Receiver’s optimal action depends on the state. The Sender does not know the state but commits to a signal structure that determines the information available to the Receiver. Translated to our setting, the firm always wants consumers to buy, consumers only want to buy if quality is high, and the firm is uninformed about product quality but can influence consumer inference through its choice of capacity constraint. The twist in our setting is that the relevant information for consumers pertains to other consumers’ actions. It is the interaction between these endogenous actions (how many demand the product) and the capacity constraint (how many at most can buy the product) that determine what

⁷In Section 4, we elaborate on this discussion of revealing information vs. hiding information, and briefly relate our analysis to other work.

consumers observe (number of sales).

The rest of the paper is organised as follows. Section 2 provides the description of the model. Section 3 contains the analysis, where we derive our main results. Section 4 follows with discussion and then concludes. Technical proofs can be found in the appendix.

2 Model

Suppose there is a service of unknown quality and two possible states of the world, $\Omega = \{G, B\}$. In state G , the service is of good quality and delivers to each consumer who uses it a value of $u_G = 1$. In state B , quality is bad and the consumer who uses the service obtains $u_B = 0$. A consumer who does not use the service gets reservation utility r .

The actual state of the world is known neither to the service provider nor to the consumers at the moment when they make their decisions. A priory beliefs are that $P(G) \equiv \beta$ and $P(B) = 1 - \beta$. Before making her decision, each consumer receives a noisy private signal about the state of the world $s \in \{g, b\}$. We assume that signal accuracy does not depend on the state, which simplifies some of our derivations.

Assumption 1. $P(g|G) = P(b|B) \equiv \alpha \in (1/2, 1)$.

As our aim is to show the *existence* of situations where the service provider can benefit from restricting its capacity, imposing symmetry from the outset can only make our task more difficult. We also assume that α is bounded away from one, so our signals are *boundedly informative*. Thus there is no asymptotic learning (see Smith and Sørensen (2000) and Acemoglu et al. (2011)) and there is a possibility of both positive and negative cascades, where consumers ignore their private signal when deciding whether to use the service.

We consider a setting where in each period, $2n$ consumers arrive at the provider and decide whether they want to use the service.⁸ More specifically, the timing of the model is

⁸The assumption of an even number of consumers allows us to avoid tie-breaking, but is not crucial for

as follows. At the start of the game, $t = -1$, the seller can set a *capacity constraint* $K \leq 2n$. This capacity choice is irreversible and limits potential sales in each period (i.e. how many consumers can use the service), which cannot exceed capacity. The state of the world is realized at $t = 0$, which in particular implies that the constraint itself does not contain any information about service quality.

In each period $t \in [1, \infty)$, $2n$ consumers arrive and observe both capacity and total sales from the consumers who arrived in period $t - 1$. That is, consumers do not directly observe how many wanted to use the service (i.e. quantity demanded) in the previous period, but only how many actually used the service (i.e. quantity sold). Notice that a sellout in period $t - 1$, where sales equal capacity, need not imply that demand precisely equaled capacity. Thus, a capacity constraint can potentially limit the information available to consumers about each others' behavior.

Each of the $2n$ consumers who arrive in period t choose whether they want to use the service based on (i) prior beliefs about quality, β , (ii) a private signal, s , and (iii) observed sales from the previous cohort.⁹ If more than K consumers want to use the service in period t , a random selection of them are served, and the remaining consumers have to use their outside option. All period- t consumers then leave the market forever, and a new cohort of $2n$ consumers arrives in period $t + 1$.

The service provider receives a fixed profit per consumer who uses the service, which we normalize to 1. He discounts profits in future periods with a factor of δ , and sets K so as to maximize expected discounted future profits.

our analysis.

⁹In our setting it does not matter whether a consumer observes only sales in the previous cohort, or in all previous cohorts. This assumption might matter in other social learning models, as we discuss in Section 4.

3 Analysis

We look at the basic trade-off faced by the service provider. On the one hand, setting a capacity constraint limits potential sales. On the other hand, the constraint can potentially prevent consumers from learning information that the provider finds unfavorable. This trade-off is the main focus of this section of the paper.

Let $P(i, N)$ denote the probability of i signals being good, out of a total of $N \in \mathbb{N}$ signals:

$$P(i, N) = \binom{N}{i} [\beta \alpha^i (1 - \alpha)^{N-i} + (1 - \beta) \alpha^{N-i} (1 - \alpha)^i]. \quad (1)$$

We can now write the conditional probability of quality being good, if consumers know there were precisely i good signals:

$$\begin{aligned} P(G|i, N) &= \frac{P(G \cap i, N)}{P(i, N)} \\ &= \frac{\binom{N}{i} P(g|G)^i P(b|G)^{N-i} P(G)}{\binom{N}{i} P(g)^i P(b)^{N-i}} \\ &= \frac{P(g|G)^i P(b|G)^{N-i} P(G)}{P(g|G)^i P(b|G)^{N-i} P(G) + P(g|B)^i P(b|B)^{N-i} P(B)}, \end{aligned}$$

since terms $\binom{N}{i}$ cancel. Using our notation, we can rewrite this conditional probability as

$$P(G|i, N) = \frac{\beta \alpha^i (1 - \alpha)^{N-i}}{\beta \alpha^i (1 - \alpha)^{N-i} + (1 - \beta) \alpha^{N-i} (1 - \alpha)^i}. \quad (2)$$

We assume that in the absence of additional information, a consumer wants to buy if her signal was good but not if it was bad. This assumption is necessary for our analysis to be non-trivial, and it amounts to the following.

Assumption 2. *We assume that parameters α , β , and r are such that*

$$P(G|0, 1) < r < P(G|1, 1)$$

By Assumption 2, consumers in the first cohort follow their own signal. Suppose that there is no capacity constraint, so the seller is unconstrained, and that consumers in the

second cohort observe that more than n consumers used the service in period 1. In this case, each consumer in the second cohort will use the service regardless of her own signal (i.e there is a positive cascade). This is because $P(G|i, 2n + 1)$ is strictly increasing in i and

$$P(G|n + 1, 2n + 1) = \frac{\beta\alpha^{n+1}(1 - \alpha)^n}{\beta\alpha^{n+1}(1 - \alpha)^n + (1 - \beta)\alpha^n(1 - \alpha)^{n+1}} = \frac{\beta\alpha}{\beta\alpha + (1 - \beta)(1 - \alpha)} = P(G|1, 1) > r,$$

by Assumption 2. In the similar way, if fewer than n consumers used the service in period 1, then consumers in the second cohort will not use the service regardless of their private signal (i.e there is a negative cascade), since

$$P(G|n, 2n + 1) = \frac{\beta\alpha^n(1 - \alpha)^{n+1}}{\beta\alpha^n(1 - \alpha)^{n+1} + (1 - \beta)\alpha^{n+1}(1 - \alpha)^n} = \frac{\beta(1 - \alpha)}{\beta(1 - \alpha) + (1 - \beta)\alpha} = P(G|0, 1) < r.$$

Finally, if exactly n out of $2n$ consumers used the service in period 1, then each consumer in the second cohort will follow her own private signal. This is because observed sales then provide no information about product quality.

Once a cascade has started it will continue indefinitely, since consumers in cohort t who observe that all those in cohort $t - 1$ used the service will infer that one of the following occurred: either a positive cascade was already triggered prior to period $t - 1$, possibly preceeded by multiple periods where exactly n consumers used the service; or exactly n consumers used the service in each of these previous periods. Either way, all consumers in cohort t find it optimal to use the service, regardless of their own signal. The same reasoning applies if consumers in cohort t observe sales of zero, in which case they will all choose not to use the service.¹⁰

Combining these observations allows us to write down the profit of a seller that remains

¹⁰The fact that cascades last forever is a standard result in the literature (see, e.g. Proposition 3 in Sgroi (2002)).

unconstrained:

$$\pi_U = \sum_{i=1}^{2n} iP(i, 2n) + \delta P(n, 2n)\pi_U + \frac{\delta}{1-\delta} \sum_{i=n+1}^{2n} 2nP(i, 2n). \quad (3)$$

Looking at (3), the first term gives expected sales in period 1, the second term gives expected sales as of period 2 if period-1-sales are equal to n , and the third term corresponds to discounted sales as of period 2 if period-1-sales exceed n and generate a positive cascade. Period-1-sales of less than n will generate a negative cascade, with zero profits in all subsequent periods.

To show that the seller may sometimes want to restrict capacity, we start by specifying the evolution of consumer beliefs and the seller profit function, for a given capacity constraint $K \leq n$.¹¹ Consider a consumer in cohort $l + 1$ who receives a bad signal, realizes there were sellouts in all l previous periods, and believes that all consumers in these cohorts followed their own private signals. Let $\gamma(l, K)$ denote this consumer's belief that the state is good. Then,

$$\gamma(l, K) = \frac{P(G \cap b, l \text{ sellouts})}{P(G \cap b, l \text{ sellouts}) + P(B \cap b, l \text{ sellouts})} = \frac{(1-\alpha)\beta \left(\sum_{j=K}^{2n} \binom{2n}{j} \alpha^j (1-\alpha)^{2n-j} \right)^l}{(1-\alpha)\beta \left(\sum_{j=K}^{2n} \binom{2n}{j} \alpha^j (1-\alpha)^{2n-j} \right)^l + \alpha(1-\beta) \left(\sum_{j=K}^{2n} \binom{2n}{j} \alpha^{2n-j} (1-\alpha)^j \right)^l}. \quad (4)$$

This consumer will ignore her own bad signal and use the service if $\gamma(l, K) > r$. It is easy to spot that selling out in period 1 does not necessarily guarantee a cascade. Consumers in the second cohort expect that, on average, a fraction of $1 - \alpha$ consumers in the first cohort make the wrong decision, so that a single sellout may not be persuasive enough to shift consumer behavior, especially if capacity is low. In particular, holding K fixed, the condition

$$\gamma(1, K) = \frac{\sum_{i=K}^{2n} P(G|i, 2n+1)P(i, 2n)}{\sum_{i=K}^{2n} P(i, 2n)} > r > P(G|0, 1) \quad (5)$$

¹¹A capacity constraint of $K \geq n + 1$ will never be optimal, even though selling out at this capacity would immediately generate a positive cascade. This is because period-1 sales of at least $n + 1$ will also trigger a cascade if the seller is unconstrained.

might not be satisfied for a particular choice of parameters. However, with each subsequent sellout, consumers become more and more optimistic about the quality of service. Lemma 1 shows that this optimism grows without bound, in that consumers' belief that the quality is good tends to 1 as l approaches infinity.

Lemma 1. *For all $1 \leq K \leq n$, consumer beliefs $\gamma(l, K)$ are increasing in l , with $\lim_{l \rightarrow \infty} \gamma(l, K) = 1$.*

Proof. See Appendix. □

Let $\mathcal{L} = \{l : \gamma(l, K) > r\}$, which is non-empty by Assumption 2 and Lemma 1. Denote the smallest element of $\mathcal{L}$ by L . Given capacity constraint K , L is precisely the number of consecutive sellouts that are necessary as of period 1 in order to generate a positive cascade.

We can now write seller profits for a given constraint K and resulting value of L . Denote $\eta(K) \equiv \sum_{i=K}^{2n} P(i, 2n)$ – the probability of a sellout, and $S(K) \equiv \sum_{i=0}^{K-1} i P(i, 2n)$ – expected sales if there is no sell out. Not selling out in any of the first L periods not only fails to generate a positive cascade; it immediately triggers a negative cascade.¹² It follows that the profit function can be written as

$$\pi(K) = \frac{1 - [\delta\eta(K)]^L}{1 - \delta\eta(K)} [S(K) + K\eta(K)] + \frac{[\delta\eta(K)]^L}{1 - \delta} K. \quad (6)$$

To show that there exists a region of parameters where it is profitable to restrict capacity, we focus on the case where selling out once at capacity K immediately triggers a cascade, i.e. $L = 1$. Suppose the seller sets $K \leq n$ and sells out in the first period. A consumer in the

¹²The logic is as follows. Suppose that $L' \leq L$ is the first period where there is no sellout. Since the sellouts in all previous periods did not trigger a positive cascade, it must be that a consumer in cohort L' , who receives a bad signal, believes that the probability of the good state is less than r . These are the same beliefs held by a consumer in cohort $L' + 1$ who receives a good signal, having observed sales of exactly $n - 1$ in period L' . Thus, if there is no sellout in period L' (i.e. sales strictly less than $K \leq n$), then consumers in cohort $L' + 1$ have an incentive not to use the service, regardless of their own signal.

second cohort who receives a bad private signal, $s = b$, will use the service if her belief that the state is G is higher than the reservation value, so if condition (5) is satisfied. Observing a sellout makes her more optimistic about the state, compared to the counterfactual of having no information about period-1-sales and just relying on her private information: $\gamma(1, K) > \gamma(0, K) = P(G|0, 1)$ by Lemma 1. Thus, for any (α, β, n) , and any $K \leq n$, there are values of r satisfying $\gamma(1, K) > r > P(G|0, 1)$, so for which condition (5) holds. For these parameter values, a single sellout leads the consumer to ignore her private signal and generates a positive cascade.

The profit of a seller that sets a capacity constraint $K \leq n$ satisfying (5) are

$$\pi_C(K) = \sum_{i=1}^{2n} iP(i, 2n) + \sum_{i=K+1}^{2n} (K - i)P(i, 2n) + \frac{\delta}{1 - \delta} \sum_{i=K}^{2n} KP(i, 2n). \quad (7)$$

The trade-off to restricting capacity is as follows. The seller loses $i - K$ in period-1 sales, compared to being unconstrained, if there are $i \geq K + 1$ good signals in period 1. If moreover there are $i \geq n + 1$ good signals in period 1, then he also loses $2n - K$ in each subsequent period, since an unconstrained seller would also experience a positive cascade. But if there are $K \leq i \leq n - 1$ good signals in period 1, then the seller gains K in each subsequent period, since the constraint is what generates a positive (rather than a negative) cascade. The seller also gains if there are precisely $i = n$ good signals in period 1, since consumers in the second cohort would then have followed their own private signals if the seller was unconstrained.

The following Theorem shows that restricting capacity, to any level, can at times increase seller profits.

Theorem 1. *For any $n > 1$, $K \leq n$ and $\delta > \sqrt{\frac{2n-K}{2n}}$, there are $(\alpha, \beta) \in (0, 1)^2$ and $r > 0$ for which the seller can increase its profit above the unconstrained level by restricting capacity to K .*

Remarkably, for any capacity constraint $K \leq n$, parameters can be found such that (i) selling out a single time at this capacity triggers a cascade, and (ii) setting this constraint

generates higher profits than remaining unconstrained. As such, Theorem 1 identifies situations where it is optimal to restrict capacity. It does not state that for arbitrary K , there exists parameters such that K is the optimal constraint, or that selling out once at the optimal constraint must necessarily trigger a cascade.

The proof of Theorem 1 relies on constructing a sequence (α_m, β_m) such that demand of exactly K from consumers in the first cohort become arbitrary more likely than any strictly higher demand. This sequence has the property that $\alpha_m \rightarrow 1$ and $\beta_m \rightarrow 0$ as m becomes large, which implies that the environment becomes increasingly pessimistic: quality is likely bad and consumers are likely to learn about it. As the environment becomes more unpromising, the probability of any positive number of good signals falls. But in particular, the probability of $i > K$ good signals falls very quickly relative to the probability of exactly $i = K$ good signals, so that restricting capacity to K becomes unlikely to cost the seller any sales.

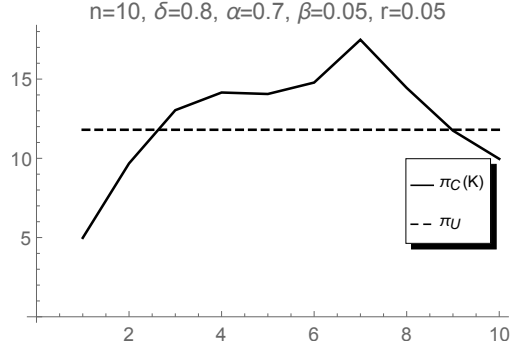
Naturally, in a very optimistic environment, early sales are likely to be high and generate a positive cascade. It is then optimal to remain unconstrained. This result is summarized in the following Proposition.

Proposition 1. *Suppose that $\sum_{i=n+1}^{2n} P(i, 2n) \geq 1/2$. Then the seller should not set a capacity constraint: $\pi_U > \pi_C(K)$ for all $K \leq n$.*

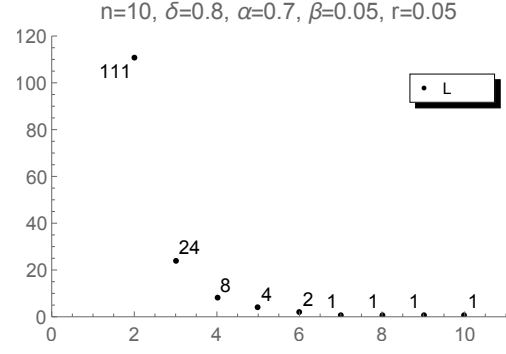
Proof. See Appendix. □

Although it only makes sense to restrict capacity when the environment is quite unpromising, the situation need not be extreme. Our proof considers the simultaneous case of low prior beliefs and high signal precision, where constrained profits exceed unconstrained profits but are nonetheless very low. Our numerical analysis, presented below, shows that restricting capacity can also be beneficial for parameters that are far from this limit case, and for which the optimal constraint generates reasonably high profits.

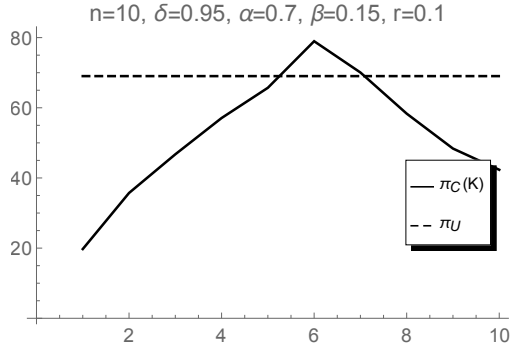
Figure 1: Seller profits and L .



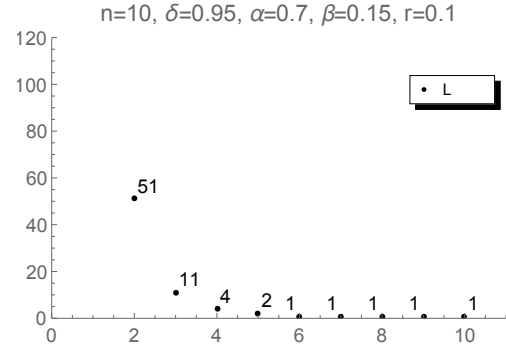
(a)



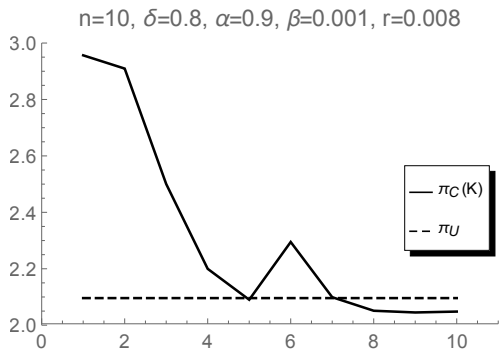
(b)



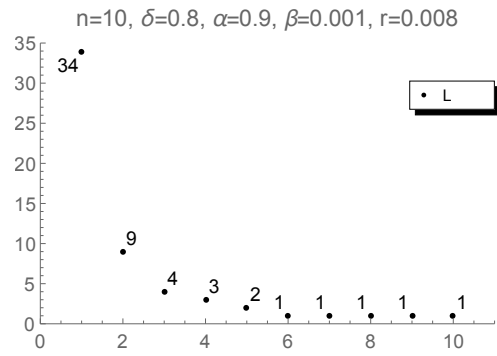
(c)



(d)



(e)



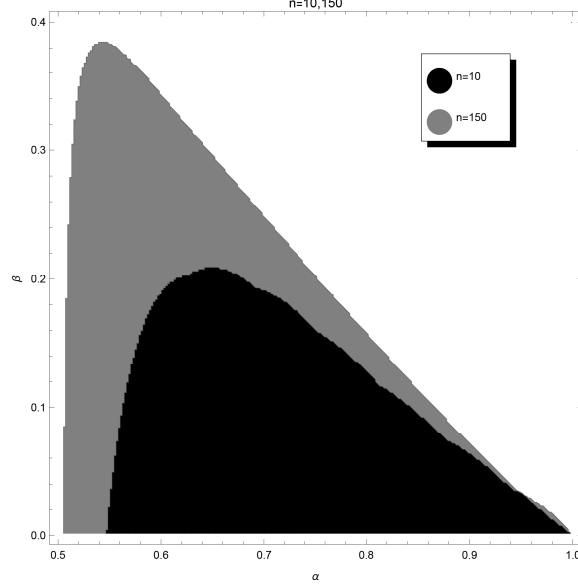
(f)

This analysis is presented in Figure 1. The panels on the left-hand-side compare unconstrained profits to profits with a capacity constraint, as a function of K . Each panel does so for different values of α , β , and δ , where the number of consumers in each cohort is fixed at $2n = 20$. The panels on the right-hand-side describe the minimum number of sellouts needed to generate a positive cascade, L , as a function of K , for these same parameter values.

The numerical analysis provides us with a few interesting observations. First, profits can still be quite substantial in situations where restricting capacity is optimal, where β can be significantly larger than zero and α (along with δ) can be much lower than one. Second, for different parameter values, the optimal capacity constraint can take on a variety of values. Some are closer to $K = n = 10$, in panels 1(a) and 1(c), whereas the optimal constraint in panel 1(e) is $K = 1$. Third, for low levels of K , the seller needs multiple sellouts to trigger the cascade, i.e. $L > 1$. The smaller the constraint, the greater the number of sellouts that are needed, since each sellout shifts consumer beliefs by a smaller amount. Fourth, the optimal constraint does not always yield $L = 1$. Although this is the case for panels 1(a)-1(d), it is not for panels 1(e)-1(f), where the seller must sell out thirty four(!) times at an optimal capacity of $K = 1$ to start a cascade. This latter situation is typical for very pessimistic scenarios. Fifth, profits might not be quasi-concave in K and sudden jumps can occur. There are three main forces created by changes in K : larger K gives a lower chance of a sellout in any given period, higher sales conditional on selling out or on being in a positive cascade, and may also mean that fewer sellouts are necessary for the cascade to begin. When multiple sellouts are required to trigger a cascade, the balance between these forces might swing very sharply. In particular, panels 1(e)-1(f) show that increasing K from 5 to 6 reduces the number of sellouts necessary to start a cascade from $L = 2$ to $L = 1$, which pushes up profits.

To provide a better idea about which values of (α, β) can make it optimal to restrict capacity, we impose $r = \gamma(0, 1)$, and compute the region of (α, β) -space for which there

Figure 2: Region where restricting capacity can be optimal



is some constraint K that increase profits over the unconstrained level. We do so for two cases with different numbers of consumers. For each case, Figure 2 displays the region where restricting capacity is optimal, in dark grey for $n = 10$ and light grey for $n = 150$. Different points in each region correspond to different values of r and to different optimal constraints.¹³

Figure 2 shows that the parameter region for which restricting capacity can be optimal is larger in the case with a larger number of consumers. It also illustrates an interesting interaction between the precision of the signal and the prior. When the signal is very imprecise ($\alpha \approx 1/2$), it is relatively likely that at least half of consumers will buy, regardless of the state. The seller then finds it optimal to remain unconstrained and bet on generating a positive cascade. As signal precision increases, an unconstrained seller becomes increasingly likely to experience a negative cascade in situations where the state is bad. This makes it

¹³Figure 2 can be seen as the counterpart to Theorem 1. In the figure, a particular pair (α, β) is included in the shaded region if and only if there exist values of $K \leq n$, r , and δ such that (i) selling out once triggers a positive cascade (which is always achieved by an appropriate choice of r), and (ii) constrained profits exceed unconstrained profits.

optimal to restrict capacity when the bad state is sufficiently likely, so for small values of β .

A further increase in signal precision has two effects: it makes restricting capacity increasingly helpful when the state is bad, but increasingly harmful when the state is good, since in the latter case a positive cascade is very likely without a constraint. For intermediate values of α , the first effect dominates. Restricting capacity can then be optimal for values of β in a reasonably large range. The second effect dominates when α is sufficiently large, because a positive cascade is then only likely in the bad state if there is a very low constraint, and such a constraint dramatically reduces profits if the state turns out to be good. For $\alpha \approx 1$, restricting capacity in this way can only make sense if the good state is extremely unlikely, so if $\beta \approx 0$.

Figure 3 explicitly shows the size of the optimal capacity constraint when $n = 10$. It depicts this capacity on the vertical axis, for each point (α, β) in the horizontal plane. The most striking feature in the region where restricting capacity can be profitable is that an increase in signal precision always decreases the size of the optimal constraint. The reason for this negative relationship is that restricting capacity generally only pays off in the bad state. When the signal is very precise, sales are likely to be very low when the state is bad, so that the firm can only hide information from consumers by setting a low constraint.

Figure 3: Size of the optimal constraint

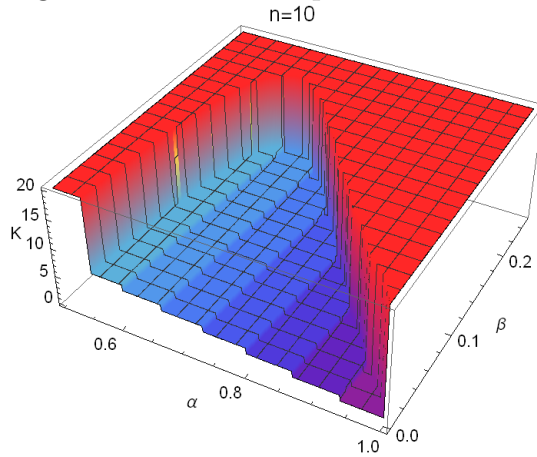
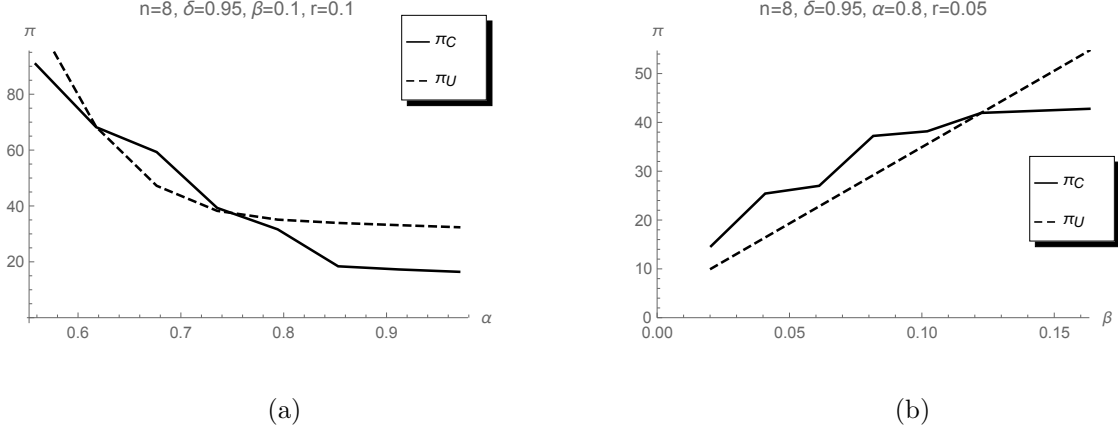


Figure 4: Comparative statics in α and β



The comparative statics in Figure 4 further illustrate how signal precision and the ex ante probability of the state being good affect seller profits. Each panel depicts profits both when the seller restricts capacity to the optimal $K \leq n$, and when it remains unconstrained. Holding fixed the probability of the good state, constrained profits exceed unconstrained profits for intermediate levels of signal precision, but not when precision is quite high or low. Moreover, for a fixed signal precision, the seller finds it optimal to restrict capacity when the bad state is sufficiently likely, and otherwise prefers to remain unconstrained. Naturally, the higher the precision of the signal (provided that $\beta < 1/2$), or the smaller the probability of the good state, the lower the profits.

Although the result of Theorem 1 holds for a range of sufficiently high δ , intuitively the seller should have more incentive to restrict capacity when δ is particularly large. This is because a capacity constraint always reduces expected profits in period 1 and can only generate potential gains in the periods that follow. The following Proposition formulates this monotonicity result: for small values of δ the seller tends to be unconstrained, while for large δ it pays off to limit capacity.

Proposition 2. *Suppose that for parameters $(\alpha, \beta, r, n, K, \delta_0)$, we have $\pi_C(K) = \pi_U$, and that condition (5) holds, so that selling out at capacity K in period 1 triggers a positive*

cascade. Then for $(\alpha, \beta, r, n, K, \delta')$ such that $\delta' > \delta_0$ the seller strictly prefers to limit capacity to K , rather than be unconstrained, i.e. $\pi_C(K) > \pi_U$, and for $(\alpha, \beta, r, n, K, \delta')$ such that $\delta' < \delta_0$ the seller prefers to be unconstrained, rather than limit capacity to K , i.e. $\pi_C(K) < \pi_U$.

Proof. See Appendix. □

Two remarks are in order. The first is that Proposition 2 gives an additional insight into the situations captured earlier by Theorem 1. Theorem 1 shows that for any given (n, K) with $1 \leq K \leq n$, we can find parameter values (α, β, r) and δ such that condition (5) holds, and it is more profitable to restrict capacity to K than it is to remain unconstrained. It can never be profitable to restrict capacity when δ is sufficiently close to 0, since a constraint always limits expected period-1-sales. Notice moreover that profits are continuous in δ . Thus, for any parameter values that give $\pi_C(K) > \pi_U$ in Theorem 1, we can say the following, holding all parameters other than δ constant: profits with capacity constraint K will exceed unconstrained profits as long as δ exceeds a threshold value, where the threshold corresponds to $\pi_C(K) = \pi_U$ in Proposition 2.

The second remark is that Proposition 2 compares the unconstrained case to a seller with *some* constraint K , not necessarily the *optimal* K . In particular, the result does not state that for $\delta' < \delta$ the seller should be unconstrained, it just states that the seller can improve by removing the constraint. The seller may potentially do even better still by moving to some other, possibly larger, constraint K' .

In our analysis, we assumed that the seller cannot adjust capacity over time. This assumption is not crucial. If a seller could adjust its capacity, it would set it at the optimal level in the first period and then, in the case of a positive cascade, remove the constraint. This insight only reinforces our result that restricting capacity can be optimal, as future gains following a cascade would no longer be bounded by the size of the initial constraint. However, the fact that the seller could adjust capacity over time would affect the optimal value of K . Since the losses from setting a low constraint would be limited by a single

period, the seller would be more aggressive in period 1 in order to increase the probability of a sellout (provided that one sellout triggers a cascade), and then capture the resulting gains in future periods.¹⁴ The following Proposition formally describes this result.

Proposition 3. *Suppose that K_1 is the optimal period-1 capacity constraint when the seller can adjust capacity over time, and that K_0 is the optimal constraint when the seller cannot do so. Suppose furthermore that selling out once at either K_0 or K_1 triggers a positive cascade, i.e. condition (5) holds for both constraints. Then $K_0 \geq K_1$.*

Proof. See Appendix. □

4 Discussion and Conclusions

In this paper, we show that a firm can benefit from limiting capacity, as sellouts make consumers more optimistic about product quality. Capacity constraints are a useful tool in many industries. For example, Ferrari will decide on the production number of a particular model well before putting it on the production line. All cars may even be sold out before production actually begins. If we treat each new Ferrari model as a new period, their sales correspond closely to the picture in our model: there is a persistent positive cascade, with repeated sellouts in subsequent periods. Moreover, as most luxury car producers die out fairly soon, one can argue that this industry corresponds broadly to the conditions for optimality of capacity restriction in our analysis: ex ante beliefs about quality should be relatively low, relative to consumers' outside options (and in practice also relative to the price). In the restaurant industry, between 59 and 95 percent of new restaurants shut down

¹⁴By this same logic, if the seller could adjust its capacity over time, and only cared about influencing consumer inference so as to maximize expected profits as of period 2, then it would be even more aggressive in period 1. The optimal period-1-capacity would then never exceed the lowest $K \leq n$ that satisfies condition (5), if such a K exists, since this constraint would maximize the probability of immediately triggering a positive cascade.

in their first three years¹⁵, which again suggests that ex ante beliefs about quality should be quite low. That being said, when restaurants do succeed (so when realized quality turns out to be high), demand can far exceed supply, resulting in long waiting lists.

To put our mechanism driving capacity restrictions into broader context, notice that the firm here effectively tries to hide information from consumers. That is, if the firm sets a capacity constraint and demand then exceeds capacity, consumers are left with coarser information than they would have received if the firm had remained unconstrained. This is in stark contrast to Debo et al. (2012), as described in the Introduction, where precisely the opposite occurs: the firm may reduce service speed in order to reveal information to consumers, not to hide it. Slow service tends to generate a relatively long queue, which suggests to uninformed consumers that demand is likely high. Our notion of restricting service to hide information from consumers also differs from Stock and Balachander (2005), who consider (as in Debo et al. (2012)) a firm that knows its own quality. They show that the firm may choose to ration uninformed consumers with positive probability, in order to reveal to these consumers that earlier demand was likely high.

Callander and Hörner (2009) consider how a capacity constraint can limit consumer information about demand, but in a setting where some consumers are better informed than others. They are interested in whether it can then be optimal for consumers to follow the decision taken by the minority, rather than the majority, of earlier arrivals. The assumption of limited capacity mainly serves to simplify their analysis in a setting with continuous time, random consumers arrivals, and turnover. They do not consider whether restricting capacity can increase profits, which is the main focus of our analysis.

Papanastasiou et al. (2014) show that a firm may hide information from consumers in a setting with social learning, but in a very different context from ours, where rationing is driven by bounded rationality. They assume that consumers who arrive later to the market

¹⁵See <http://possector.com/management/restaurant-failure>.

observe the product reviews from those who purchased earlier. However, these consumers do not correctly use Bayes' rule in inferring product quality from the reviews. This may lead the firm to impose early shortages, if restricting sales can lead to better reviews on average. Unlike this work, our analysis does not rely on assumptions of bounded rationality, and hence gives insights into strategic situations with sophisticated consumers.

Johnson and Myatt (2006) also consider a firm's incentive to hide or provide information, but they focus on information that drives up the willingness to pay of some consumers and drives down that of others. Providing this information rotates the demand curve by increasing the dispersion of consumer valuations. In contrast, in our setting, revealing information has the same effect on all consumers, leading to a shift in demand in any particular cohort, rather than a rotation. Dispersion also plays a role in our analysis but in a different way. We show that hiding information from consumers can be attractive in situations that appear *a priori* to be quite grim, where most consumers likely have the same low valuation, because most of them likely receive the same negative signal.

In a slightly modified version of our setting, the fact that a capacity constraint can hide information from consumers would likely do more than increase the probability of a positive cascade; it may well also increase cascade stability. We show that if the firm restricts capacity, and there is a positive cascade, then consumers will observe a sell out in each period but not the exact level of demand. This hidden information about excess demand does not matter in our setting because signals are boundedly informative and cascades last forever. But this would likely change if we assumed there was always a small probability that each consumer followed her own signal, either because her private signal was too strong to ignore, or because she was a 'resolute' or 'naive' behavioral type. A consumer who observed that others refused to buy would then become less optimistic about product quality, which could potentially reverse a positive cascade. This ties in with the notion from Bikhchandani et al. (1992) and Bikhchandani et al. (1998) that cascades may sometimes be fragile. But

for a capacity-constrained firm with excess demand, a consumer will only notice that others refuse to buy in the rare event that at least $2n - K + 1$ do so simultaneously, since only then would the firm fail to sell out.

It may appear surprising that a firm which does not initially know its own quality, and faces consumers who are fully rational, can still successfully engage in persuasion. Here we mean persuasion in the sense of manipulating the information available to consumers so as to increase profits. The reason persuasion is possible in our setting is that overall profits are highly non-linear in period-2 beliefs, and hence in period-1 sales. In order to generate a positive purchase cascade, consumers in the second cohort must observe sales that push their beliefs about quality over a threshold value. This threshold would leave a consumer indifferent about buying if she herself received a bad signal. By setting the appropriate capacity constraint, the firm can increase the probability that period-2 beliefs exceed this threshold. This can help the firm despite the fact that consumers are not fooled on average.

Now we briefly discuss the role of various assumptions in our paper. The fact that consumers arrive in cohorts is not very important, provided that the first $2n$ consumers do not observe one another's actions. Later stages of the model can be adapted for a wide class of arrival processes, assuming that a capacity constraint limits the number of consumers who can be served within a certain time frame. The assumption of bounded private signals is important in our case as it insures that cascades are persistent. If this assumption were relaxed (and all previous sales were observed) then social learning would be asymptotically efficient. Our assumption that the firm sets capacity before knowing product quality also plays an important role, as otherwise the firm and consumers would engage in a sender-receiver game with potentially quite different outcomes.

We assume throughout that consumers in each cohort only observe sales in the cohort that immediately precedes them. As argued in Section 3, our results would remain unchanged if consumers could observe sales from each and every previous cohort. Our results would

also remain unchanged if consumers could only observe total sales from all cohorts that preceded them. In any particular cohort, all signals realisations except for (at most) one will immediately trigger a cascade, and cascades last forever, which allows consumers who observe total sales to infer the precise earlier sales path.

It is certainly true that assumptions about observability might matter in other social learning models.¹⁶ Behavior would also change in our setting if we assumed, for example, that consumers were uncertain about the identity of their own cohort, and only observed sales from a random sample of earlier cohorts. But while this assumption would affect consumer beliefs, there is little to suggest that it would affect the mechanism driving our results on capacity constraints. Consumers who observed a sellout from any earlier cohort would still infer that demand might have exceeded capacity, making them more optimistic about product quality. When deciding whether to restrict capacity, the seller would consider this positive effect of sellouts on consumer beliefs, along with the negative effect of constraining sales in each cohort, just as in our analysis.

Appendix: Proofs

Proof of Lemma 1. Denote

$$A \equiv \sum_{j=K}^{2n} \binom{2n}{j} \alpha^j (1-\alpha)^{2n-j}$$

$$B \equiv \sum_{j=K}^{2n} \binom{2n}{j} \alpha^{2n-j} (1-\alpha)^j$$

Then expression $\gamma(l, K) < \gamma(l+1, K)$ is equivalent to

$$\frac{(1-\alpha)\beta A^l}{(1-\alpha)\beta A^l + \alpha(1-\beta)B^l} < \frac{(1-\alpha)\beta A^{l+1}}{(1-\alpha)\beta A^{l+1} + \alpha(1-\beta)B^{l+1}}$$

¹⁶See the argument in Acemoglu et al. (2011). Indeed, in the particular settings examined by Çelen and Kariv (2004) and Callander and Hörner (2009), assuming full observability of earlier actions would affect their results.

which holds if and only if $B < A$. Now as

$$A = \sum_{j=K}^{2n} \binom{2n}{j} \alpha^j (1-\alpha)^{2n-j} = 1 - \sum_{j=0}^{K-1} \binom{2n}{j} \alpha^j (1-\alpha)^{2n-j}$$

$$B = \sum_{j=K}^{2n} \binom{2n}{j} \alpha^{2n-j} (1-\alpha)^j = 1 - \sum_{j=0}^{K-1} \binom{2n}{j} \alpha^{2n-j} (1-\alpha)^j$$

condition $B < A$ is equivalent to

$$\sum_{j=0}^{K-1} \binom{2n}{j} \alpha^{2n-j} (1-\alpha)^j > \sum_{j=0}^{K-1} \binom{2n}{j} \alpha^j (1-\alpha)^{2n-j}$$

which always holds as each term $\alpha^{2n-j} (1-\alpha)^j > \alpha^j (1-\alpha)^{2n-j}$ as long as $j \leq K \leq n$ and $\alpha > 1/2$. Thus, $\gamma(l, K)$ increases in l . Moreover, we can rewrite

$$\gamma(l, K) = \frac{(1-\alpha)\beta}{(1-\alpha)\beta + \alpha(1-\beta) \left(\frac{B}{A}\right)^l}$$

which, as $B < A$, approaches 1 with l approaching infinity.

□

Proof of Theorem 1. From the text preceding the Theorem we know that it is possible to choose r in such a way, that cascade occurs after one sellout for any (α, β, K, n) . Now, rearrange the terms in (3) to obtain

$$[1 - \delta P(n, 2n)] \pi_U = S_1 + \frac{\delta}{1-\delta} 2n S_2, \quad (8)$$

where $S_1 = \sum_{i=1}^{2n} i P(i, 2n)$ and $S_2 = \sum_{i=n+1}^{2n} P(i, 2n)$. Note that as long as $\alpha < 1$, the term $[1 - \delta P(n, 2n)]$ is bounded away from 0. In a similar way,

$$\pi_C(K) = S_1 - \sum_{i=K}^{2n} (i-K) P(i, 2n) + \frac{\delta}{1-\delta} K \left[\sum_{i=K}^{n-1} P(i, 2n) + P(n, 2n) + S_2 \right] >$$

$$S_1 - (2n-K) \left[\sum_{i=K}^{n-1} P(i, 2n) + P(n, 2n) + S_2 \right] + \frac{\delta}{1-\delta} K \left[\sum_{i=K}^{n-1} P(i, 2n) + P(n, 2n) + S_2 \right],$$

where the inequality follows from replacing all terms $(i - K)$ with the larger term $2n - K$. Moreover, for $\delta > \frac{2n-K}{2n}$ the penultimate term is smaller than the discounted sum of future profits, and thus

$$\pi_C(K) > S_1 + \frac{\delta K - (2n - K)(1 - \delta)}{1 - \delta} [P(n, 2n) + S_2]$$

This implies that $\pi_C(K = n) > \pi_U$ if

$$\left\{ S_1 + \frac{\delta K - (2n - K)(1 - \delta)}{1 - \delta} [P(n, 2n) + S_2] \right\} [1 - \delta P(n, 2n)] \geq S_1 + \frac{\delta}{1 - \delta} 2n S_2,$$

which can be rewritten as

$$\begin{aligned} -\delta P(n, 2n) S_1 + \frac{\delta K - (2n - K)(1 - \delta)}{1 - \delta} P(n, 2n) [1 - \delta P(n, 2n)] \geq \\ \frac{S_2}{1 - \delta} \{2n\delta - [\delta K - (2n - K)(1 - \delta)][1 - \delta P(n, 2n)]\}. \end{aligned}$$

S_1 is just the first period expected sales for an unconstrained seller, which implies $S_1 < 2n$.

Thus, the above inequality holds as long as

$$\begin{aligned} \frac{P(n, 2n)}{1 - \delta} \{[\delta K - (2n - K)(1 - \delta)][1 - \delta P(n, 2n)] - 2n\delta(1 - \delta)\} \geq \\ \frac{S_2}{1 - \delta} \{2n\delta - [\delta K - (2n - K)(1 - \delta)][1 - \delta P(n, 2n)]\} \end{aligned}$$

Now note that for all $\delta \in (0, 1)$, the expression

$$f_R \equiv 2n\delta - [\delta K - (2n - K)(1 - \delta)][1 - \delta P(n, 2n)] = 2\delta^2 n P(n, 2n) + (2n - K)[1 - \delta P(n, 2n)]$$

is always positive. Moreover, the expression

$$f_L \equiv [\delta K - (2n - K)(1 - \delta)][1 - \delta P(n, 2n)] - 2n\delta(1 - \delta)$$

is also strictly positive if $\delta > \frac{\sqrt{[(2n-K)P(n,2n)]^2 + 8n(2n-K)[1-P(n,2n)]} - (2n-K)P(n,2n)}{4n-4nP(n,2n)} \equiv \delta^*(P(n, 2n); n, K)$,

where the critical value $\delta^*(P(n, 2n); n, K)$ is increasing in $P(n, 2n)$, equals $\sqrt{\frac{2n-K}{2n}}$ at $P(n, 2n) = 0$, and approaches 1 as $P(n, 2n) \rightarrow 1$. Note, that $\sqrt{\frac{2n-K}{2n}} > \frac{2n-K}{2n}$, and therefore condition

we used for the approximation of the constrained profit is automatically satisfied as long as $\delta > \sqrt{\frac{2n-K}{2n}}$. Thus, for any $P(n, 2n) < 1$, there exists $\delta \in \left(\sqrt{(2n-K)/2n}, 1\right)$ such that the right-hand-side of

$$\frac{P(n, 2n)}{S_2} = \frac{P(n, 2n)}{\sum_{i=n+1}^{2n} P(i, 2n)} \geq \frac{f_R}{f_L} \quad (9)$$

is positive and finite. Moreover, such δ can be chosen arbitrarily close to $\sqrt{\frac{2n-K}{2n}}$, as long as $P(n, 2n)$ is sufficiently close to zero. Now, we focus on the ratio on the left-hand-side of (9). This left-hand-side becomes arbitrarily large as α tends to 1 and β tends to zero. In order to see this rewrite the left hand side of (9) as

$$\frac{P(n, 2n)}{\sum_{i=n+1}^{2n} P(i, 2n)} = \frac{\binom{2n}{n} \alpha^n (1-\alpha)^n}{\beta \sum_{i=n+1}^{2n} \binom{2n}{i} \alpha^i (1-\alpha)^{2n-i} + (1-\beta) \sum_{i=n+1}^{2n} \binom{2n}{i} \alpha^{2n-i} (1-\alpha)^i}.$$

Now, our aim is to show that for any arbitrary M , we can find (α, β) such that

$$P(n, 2n) / \left(\sum_{i=n+1}^{2n} P(i, 2n) \right) > M. \quad (10)$$

Denote

$$A_\alpha = \binom{2n}{n} \alpha^n (1-\alpha)^n, \quad B_\alpha = \sum_{i=n+1}^{2n} \binom{2n}{i} \alpha^i (1-\alpha)^{2n-i}, \quad C_\alpha = \sum_{i=n+1}^{2n} \binom{2n}{i} \alpha^{2n-i} (1-\alpha)^i$$

Then, inequality (10) holds if

$$\frac{A_\alpha - C_\alpha M}{M(B_\alpha - C_\alpha)} > \beta$$

Now take a sequence $\alpha_m \rightarrow 1$. Note that $B_{\alpha_m} > C_{\alpha_m}$ for all $\alpha > 1/2$.

Recall that for any $\delta > \sqrt{(2n-K)/2n}$, there exists a (sufficiently small) value P , such that $f_L > 0$ holds for all $P(n, 2n) < P$. Note as well that $\lim_{\alpha_m \rightarrow 1} P(n, 2n) = 0$. Therefore, there exists an index m_0 such that, for all $m > m_0$, f_L is positive, so the ratio f_R/f_L is bounded from above.

From the definition of A_{α_m} and C_{α_m} , we have

$$\lim_{\alpha_m \rightarrow 1} \frac{A_{\alpha_m}}{C_{\alpha_m}} = \lim_{\alpha_m \rightarrow 1} \frac{\binom{2n}{n}}{\sum_{i=n+1}^{2n} \binom{2n}{i} \alpha_m^{n-i} (1-\alpha_m)^{i-n}} = \infty.$$

Thus, for any M , there exists m_1 such that, for all $m > m_1$, the expression $A_{\alpha_m} > C_{\alpha_m}M$ holds. Now define a sequence $\beta_m = \frac{A_{\alpha_m} - C_{\alpha_m}M}{2M(B_{\alpha_m} - C_{\alpha_m})}$ ¹⁷. It follows that for any $m > \max\{m_0, m_1\}$, the ratio of probabilities in (10) is arbitrary large, while f_R/f_L is bounded, and hence the constrained profit is larger than the unconstrained.

□

Proof of Proposition 1. First, notice that

$$\pi_C(K) < \sum_{i=1}^{2n} iP(i, 2n) + \frac{\delta}{1-\delta} \sum_{i=K}^{2n} KP(i, 2n)$$

and

$$\pi_U > \sum_{i=1}^{2n} iP(i, 2n) + \frac{\delta}{1-\delta} \sum_{i=n+1}^{2n} 2nP(i, 2n)$$

Hence,

$$\pi_U - \pi_C(K) > \frac{\delta}{1-\delta} \sum_{i=n+1}^{2n} 2nP(i, 2n) - \frac{\delta}{1-\delta} \sum_{i=K}^{2n} KP(i, 2n)$$

or, equivalently,

$$\pi_U - \pi_C(K) > \frac{\delta}{1-\delta} \left(\sum_{i=n+1}^{2n} (2n - K)P(i, 2n) - \sum_{i=K}^n KP(i, 2n) \right)$$

The result then follows from $2n - K \geq K$ and $\sum_{i=n+1}^{2n} P(i, 2n) \geq 1/2 \geq \sum_{i=K}^n P(i, 2n)$. □

Proof of Proposition 2. Write out the profits of constrained and unconstrained firms at some δ assuming (5) holds.

$$\begin{aligned} \pi_C(K) &= \sum_{i=1}^{2n} iP(i, 2n) + \sum_{i=K+1}^{2n} (K - i)P(i, 2n) + \frac{\delta}{1-\delta} \sum_{i=K}^{2n} KP(i, 2n) \\ \pi_U &= \sum_{i=1}^{2n} iP(i, 2n) + \delta P(n, 2n)\pi_U + \frac{\delta}{1-\delta} \sum_{i=n+1}^{2n} 2nP(i, 2n) \end{aligned}$$

¹⁷Since A_{α_m} and C_{α_m} both approach zero as $\alpha_m \rightarrow 1$, while B_{α_m} approaches one, we have $\beta_m \in (0, 1)$ for sufficiently large m , with $\lim_{m \rightarrow \infty} \beta_m = 0$

Comparing and simplifying shows that

$$\begin{aligned} \pi_C(K) - \pi_U = & \underbrace{\frac{\delta}{1 - \delta p_1} \left((K - \sum_{i=0}^{2n} iP(i, 2n))p_1 - (2n - K)p_2 + \frac{\delta}{1 - \delta} [p_1(1 - p_1 - p_2)K - (2n - K)p_2] \right)}_{\equiv X(\delta)} \\ & - \sum_{i=K+1}^{2n} (i - K)P(i, 2n) \end{aligned}$$

where $p_1 = P(n, 2n)$ and $p_2 = \sum_{i=n+1}^{2n} P(i, 2n)$. Denote the first term of this expression as $X(\delta)$. As $\pi_C(K) = \pi_U$ given discount factor δ_0 we get

$$X(\delta_0) = \sum_{i=K+1}^{2n} (i - K)P(i, 2n) > 0$$

.

We consider four possible cases. First case: $(K - \sum_{i=0}^{2n} iP(i, 2n))p_1 - (2n - K)p_2 < 0$ and $p_1(1 - p_1 - p_2)K - (2n - K)p_2 < 0$. Then $X(\delta') < 0$ for all $\delta' \in (0, 1]$, which is a contradiction.

Second case: $(K - \sum_{i=0}^{2n} iP(i, 2n))p_1 - (2n - K)p_2 > 0$ and $p_1(1 - p_1 - p_2)K - (2n - K)p_2 < 0$. This in turn implies

$$(K - \sum_{i=0}^{2n} iP(i, 2n))p_1 > p_1(1 - p_1 - p_2)K,$$

or equivalently

$$\sum_{i=n}^{2n} KP(i, 2n) > \sum_{i=0}^{2n} iP(i, 2n)$$

which cannot hold by $K \leq n$ (we have $\pi_C(K) < \pi_U$ whenever $n + 1 \leq K \leq 2n - 1$).

Third case: $(K - \sum_{i=0}^{2n} iP(i, 2n))p_1 - (2n - K)p_2 > 0$ and $p_1(1 - p_1 - p_2)K - (2n - K)p_2 > 0$. Then $X(\delta)$ is strictly increasing in δ , and we have our result.

Fourth case: $(K - \sum_{i=0}^{2n} iP(i, 2n))p_1 - (2n - K)p_2 < 0$ and $p_1(1 - p_1 - p_2)K - (2n - K)p_2 > 0$. Then the expression in large brackets in $X(\delta)$ is strictly increasing in the discount factor.

Moreover, this expression is negative for δ sufficiently small, and positive for δ sufficiently large. Thus, there exists δ'' such that $X(\delta)$ is negative for all $\delta < \delta''$, and equals zero when $\delta = \delta''$. Notice as well that $X(\delta)$ is strictly increasing in δ for all $\delta \geq \delta''$. Putting things together, we have our result. □

Proof of Proposition 3. The optimality of K_0 and K_1 implies:

$$\begin{aligned} \sum_{K_0+1}^{2n} (K_0 - i)P(i, 2n) + \frac{\delta}{1-\delta} K_0 \sum_{i=K_0}^{2n} P(i, 2n) \\ \geq \sum_{K_1+1}^{2n} (K_1 - i)P(i, 2n) + \frac{\delta}{1-\delta} K_1 \sum_{i=K_1}^{2n} P(i, 2n) \end{aligned}$$

and

$$\begin{aligned} \sum_{K_1+1}^{2n} (K_1 - i)P(i, 2n) + \frac{\delta}{1-\delta} 2n \sum_{i=K_1}^{2n} P(i, 2n) \\ \geq \sum_{K_0+1}^{2n} (K_0 - i)P(i, 2n) + \frac{\delta}{1-\delta} 2n \sum_{i=K_0}^{2n} P(i, 2n) \end{aligned}$$

Summing it up yields

$$\begin{aligned} \frac{\delta}{1-\delta} K_0 \sum_{i=K_0}^{2n} P(i, 2n) + \frac{\delta}{1-\delta} 2n \sum_{i=K_1}^{2n} P(i, 2n) \\ \geq \frac{\delta}{1-\delta} K_1 \sum_{i=K_1}^{2n} P(i, 2n) + \frac{\delta}{1-\delta} 2n \sum_{i=K_0}^{2n} P(i, 2n) \end{aligned}$$

or

$$\sum_{i=K_1}^{2n} (2n - K_1)P(i, 2n) \geq \sum_{i=K_0}^{2n} (2n - K_0)P(i, 2n)$$

which implies that $K_1 \leq K_0$ as $S(K) = \sum_{i=K}^{2n} (2n - K)P(i, 2n)$ is a decreasing function of K .

□

References

- Daron Acemoglu, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.
- Itai Arieli. Payoff externalities and social learning. Mimeo, 2016.
- Abhijit Banerjee and Drew Fudenberg. Word-of-mouth learning. *Games and Economic Behavior*, 46(1):1–22, 2004.
- Abhijit V Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, pages 797–817, 1992.
- Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, pages 992–1026, 1992.
- Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 12(3):151–170, 1998.
- Subir Bose, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund. Dynamic monopoly pricing and herding. *The RAND Journal of Economics*, 37(4):910–928, 2006.
- Subir Bose, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund. Monopoly pricing in the binary herding model. *Economic Theory*, 37(2):203–241, 2008.

- Steven Callander and Johannes Hörner. The wisdom of the minority. *Journal of Economic theory*, 144(4):1421–1439, 2009.
- Boğaçhan Çelen and Shachar Kariv. Observational learning under imperfect information. *Games and Economic Behavior*, 47(1):72–86, 2004.
- Laurens G Debo, Christine Parlour, and Uday Rajan. Signaling quality via queues. *Management Science*, 58(5):876–891, 2012.
- Erik Eyster, Andrea Galeotti, Navin Kartik, and Matthew Rabin. Congested observational learning. *Games and Economic Behavior*, 87:519–538, 2014.
- Justin P Johnson and David P Myatt. On the simple economics of advertising, marketing, and product design. *The American Economic Review*, 96(3):756–784, 2006.
- Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.
- Ignacio Monzón. Aggregate uncertainty can lead to incorrect herds. *American Economic Journal: Microeconomics*, forthcoming.
- Ignacio Monzón and Michael Rapp. Observational learning with position uncertainty. *Journal of Economic Theory*, 154:375–402, 2014.
- Yiangos Papanastasiou, Nitin Bakshi, and Nicos Savva. Scarcity strategies under quasi-bayesian social learning. *History*, 2014.
- Luis Rayo and Ilya Segal. Optimal information disclosure. *Journal of Political Economy*, 118(5):949–987, 2010.
- Daniel Sgroi. Optimizing information in the herd: Guinea pigs, profits, and welfare. *Games and Economic Behavior*, 39(1):137–166, 2002.

Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

Axel Stock and Subramanian Balachander. The making of a hot product: A signaling explanation of marketers scarcity strategy. *Management Science*, 51(8):1181–1192, 2005.