Learning in Crowded Markets *

Péter Kondor[†] Adam Zawadowski[‡]

London School of Economics Central European University

May 30, 2016

Abstract

We study how competition among investors affects the efficiency of capital allocation, the speed of capital, and welfare. In our model, investors learn about other entrants in a fully flexible way. We find that competition increases the speed of capital, but does not necessarily improve the efficiency of capital allocation: there is persistent over- or underinvestment. As speed is a by-product of costly over-learning, increasing competition decreases welfare. We describe how the speed of capital and the level of over- or underinvestment depend on market and investor characteristics. With investors of heterogeneous skills, more sophisticated investors might harm welfare.

*For comments and suggestions, we are grateful to Patrick Bolton, Thomas Chemmanur, Johannes Hörner, Victoria Vanasco, Marianne Andries, Ming Yang, Martin Oehmke and seminar participants at Boston University, Central European University, ESSET 2014 (Gerzensee), Brigham Young University, GLMM 2015 (Boston), Paul Woolley Conference 2015 (LSE), University of Vienna, 6th Annual Financial Market Liquidity Conference (Budapest), American Finance Association 2016 Annual Meeting (San Fransisco), Adam Smith Workshop 2016 (Oxford), University of Maryland, University of Naples, University of Copenhagen, 4nation cup workshop (Copenhagen), Federal Reserve Bank of New York. We thank Pellumb Reshidi for his excellent research assistance. The support of the European Research Council Starting Grant #336585 is acknowledged.

[†]e-mail: p.kondor@lse.ac.uk, http://personal.lse.ac.uk/kondor

[‡]e-mail: zawa@ceu.edu, http://www.personal.ceu.edu/staff/Adam_Zawadowski/

1 Introduction

Global markets are increasingly dominated by "smart money", such as hedge funds. These investors devote vast resources to finding new investment opportunities before their competititors. Does this increasing competition make markets more efficient? Does it increase welfare?

We study a capital reallocation problem where some investors find a new investment opportunity. For example, this might be an emerging market or a new technology sector. Each investor is uncertain about how many of her competitors have already invested, but can gather information optimally and decide accordingly.[1] Our main observation is that increasing competition does not effect the efficiency of capital allocation. However, it makes capital reallocation faster. Increasing speed is a by-product of costly over-learning, therefore, it decreases welfare. Instead of the level of competition, the amount of capital reallocated is determined by the characteristics of the market and the investors. A practical implication of our results is that neither the speed of capital, nor allocational efficiency can be used as a proxy to determine the welfare effects of increased competition.

Our model is an entry-game where each investor decides whether to invest into a new market with insufficient capital. The new market is characterized by a decreasing returns to scale technology. The investor's pay-offs depends on her initially unknown relative type. The earlier the type of an investor, the earlier she recognizes the new investment opportunity. Each entering investor's pay-off depends both on the mass of investors entering before and after her. Due to the decreasing marginal return to capital, the model exhibits two main properties. First, trades can become "crowded" since the more investors enter, the lower the payoff for all. Second, early entrants can buy capital at a low cost where it is abundant and transfer it to another location where it is very valuable. This implies earlier type investors benefit the most. To the contrary, late types might even lose if they invest. Thus investor's payoff depends crucially on her type, implying a potential "rat race". In fact, the rat race property

[1] As Stein (2009) points out, several popular trading strategies are "unanchored", that is, the price gives little information on how much capital has been already invested in that given market. It is a major concern for investors when these strategies become "crowded" as managers inflict negative externalities on each other. (e.g. Khandani and Lo (2011), Sun, Wang, and Zheng (2012))

captures the zero-sum part of the trading process: early entrants gain at the expense of late entrants. Allowing for idiosyncratic or aggregate liquidity shocks changes the relative strengths of these two properties.

To avoid entering crowded trades, investors can learn about their type. In equilibrium, this is equivalent to learning about how many investors have already invested. Following the rational inattention approach of Sims (1998), Woodford (2008) and Yang (2015a), we allow investors to acquire optimally chosen flexible signals about their type. This choice is subject to a constant marginal cost of reducing uncertainty around their type as measured by entropy. By information theory, this marginal cost is the cost of an additional line in a computer code optimally mapping a large amount of information into a single decision whether to invest. Formally, each investor chooses a function mapping its ex ante unknown type into a probability of entry. For example, entering with a constant probability regardless of its type is free of learning cost, as this function does not require any reduction in uncertainty around its type. However, entering only when very few other investors have entered before is costly, because this requires a large reduction in uncertainty. This approach has several advantages. First, it parsimoniously captures the joint choice of entry and learning. Second, it allows for full flexibility in learning. Third, it also has an axiomatic foundation based on information theory.

Our main focus is to analyze how the allocation and speed of capital and welfare change as the competition among investors increases. We model increased competition by increasing the mass of investors who may learn about the investment opportunity and invest. Unless the mass is so small that the entry decision is trivial, increasing competition does not improve the efficiency of capital allocation. Instead, the total capital reallocated stabilizes at an inefficiently low or high level. To understand this result, note that investors adjust their entry decisions along two main dimensions as competition increases. First, the relative benefit of entering early compared to late is increasing because there are more early investors. Thus investors choose entry strategies that are more contingent on their types. We refer to this as the "rat race effect". Second, with more investors present, crowding becomes a bigger concern. This means each investor is less likely to enter on average, we refer to this as the “crowding effect”. With flexible learning these two effects cancel out in the aggregate. Thus aggregate entry remains constant and thus allocative efficiency does not improve due to increased competition.

Nevertheless, as competition increases, welfare decreases. The key insight is that the “rat race effect” increases with competition, thus investors choose to learn more. Thus early investors are more and more likely to enter compared to late investors. Increased competition leads to faster reallocation of capital in the aggregate, that is, the speed of capital increases. However, this learning is socially wasteful since the planner is only interested in the efficiency of capital allocation which is determined by aggregate entry as opposed to the order of entry. This is why welfare decreases in competition. In equilibrium, there is too much learning and capital moves too fast. In an extension, we show that speed is excessive even if there is social value to speed.

Whether there is over- or under-entry compared to the planners’ solution is independent of the mass of investors. However, it crucially depends the characteristics of the new investment opportunity and the potential shocks. First, a market with more investors provides easier exit opportunities for those hit by an idiosyncratic liquidity shock. This is not internalized by market participants leading to under-entry of investors in situations such as twin-stocks. Second, over-entry is more likely for investments with more likely and more severe aggregate shocks, since investors do not internalize the fire sale externalities. This means we should expect over-investment in trades like carry trades that are mostly subject to aggregate risk. Third, markets with higher cost of learning are more prone to over-investment. Examples for such markets include emerging technologies and markets where learning is hard due to the novel nature of the investment.

Allowing for flexible information acquisition is crucial for the specific result that aggregate entry is independent of competition. We show this in an extension where investors can only acquire Gaussian signals about their type subject to the same entropy cost as before. If investors only have to pay for the part of the Gaussian signal they use in their decision, the results are qualitatively very similar to our baseline, though entry is not exactly constant. If investors have to pay for all information they acquire, including unused information, then information acquisition becomes too costly and aggregate entry is increasing as crowding gets worse. However, our main insight still holds: allocational efficiency, speed and welfare behave differently as competition increases. Since professional investors in the real world are quite inventive in gathering and processing information, we believe the results for flexible information structure are important in understanding market behavior.

We also extend our model to the case in which there is heterogeneity across investors: some are more sophisticated than others. Keeping the mass of all investors fixed but increasing the share of sophisticated investors might also decrease welfare. Having some sophisticated investors increases welfare since it raises the average sophistication of investors and this can alleviate over-entry. As sophisticated investors start dominating, less sophisticated investors are afraid of being ripped off and exit the market. Once less sophisticated investors exit, sophisticated investors engage in a vicious "rat race" of learning which leads to decreasing welfare.

Our main contribution is to embed learning in a model of capital allocation. Our paper is connected to various branches of literature. First, there is a recent and growing literature on slow moving capital, see Pedersen, Mitchell, and Pulvino (2007), Duffie (2010), Duffie and Strulovici (2012), Oehmke (2009),and Greenwood, Hanson, and Liao (2015). They also assume that capital does not immediately move to markets where it is scarce but focus on the asset pricing implications. We focus on the endogenous choice of the amount of capital transferred and show that even though increasing the amount of investors might make capital transfer faster, markets do not converge to efficiency and welfare deteriorates.

Second, there is a literature analyzing entry/exit in the presence of externalities from other investors. Stein (2009) introduces a simple model of crowded markets but leaves the effect of learning in such models for future research. Abreu and Brunnermeier (2003) and Moinas and Pouget (2013) show that the inability to learn about one's relative position versus that of other investors' is a key ingredient in sustaining excessive investment in bubbles. This highlights our contribution in adding learning to a model of crowded markets with potential over-entry.

Third, a growing literature analyzes the consequences of limited information processing capacity based on the rational inattention approach pioneered by Sims (1998) and Sims (2003). Maćkowiak and Wiederholt (2009), Hellwig and Veldkamp (2009) and Kacperczyk, Nieuwerburgh, and Veldkamp (2016) study the allocation of limited attention across signals but restrict the signals to be Gaussian. Fully flexible information acquisition in rational inattention models is employed by Matějka and McKay (2015), Woodford (2008), Yang (2015a) and Yang (2015b). Typically, these papers focus on learning about common value uncertainty, while we focus on learning on private values. Also, none of the above papers directly analyze capital allocation.

Fourth, there are numerous papers showing excessive investment in learning or effort. In models of high frequency trading, Budish, Cramton, and Shim (2015) and Biais, Foucault, and Moinas (2015) show that there is excessive investment in speed if trading is continuous in time. Our framework is conceptually different: investors cannot change their individual type (speed), but more learning results higher speed in the aggregate. Also, our insights work on longer time horizons. There is also a distinct literature on the social value of private learning where prices reveal private information which can change ex ante incentives for insurance and learning or ex-post trading opportunities, e.g. Hirshleifer (1971), Grossman and Stiglitz (1980), Glode, Green, and Lowery (2012). More generally, socially inefficient effort choice has also been emphasized in very different settings: e.g. Tullock (1967), Krueger (1974), and Loury (1979). Our focus is different compared to the above papers: we are interested in how learning affects capital allocation.

The rest of the paper is structured as follows. In Section 2 we present our reduced form model and also give a structural microfoundation. In Section 3 we present the optimal choice of entry and learning and analyze its implications for aggregate entry, market efficiency, speed and welfare. In Section 4 we analyze different variations of the payoff function, cost function and also allow for heterogeneity in investor sophistication. Section 5 concludes. All proofs are relegated to Appendix A. In Online Appendix B, we show the applicability of the reduced-form model in variety of different settings.

2 A model of learning and investing in crowded markets

In this part we describe our set up. We first present the reduced form payoff function, then describe a micro-foundation. We then introduce the flexible learning technology and define the real outcomes.

2.1 Payoffs

The heart of our model is an entry game with a continuum (mass M) investors, each with a type $\theta \in [0,1]$. Each investor can decide to take an action: whether to enter in a market or not. θ is interpreted as the time when investor θ can make this decision. The utility gain (or loss, if negative) from entry is given by

$$\Delta u(\theta) = 1 - \beta \cdot b(\theta) + \alpha \cdot a(\theta) \tag{1}$$

where α and β are constant parameters. $a(\theta)$ denotes the equilibrium mass of entrants whose type is higher than θ, i.e. the investors who enter *a*fter investor θ. $b(\theta)$ denotes the equilibrium mass of entrants with a type lower than θ, i.e. investors who enter *b*efore investor θ. We show in the microfoundation that the following two assumptions are natural. First, $\beta + \alpha > 0$, such that entering earlier is better than later: we call this property rat-race. Second, $\beta - \alpha > 0$, such that the average entrant imposes a negative externality on others: we call this property crowding. The two assumptions together imply that $\beta > 0$ while α could be positive or negative. As we specify below, players do not know their type, but can gather information about it through a costly learning process.

While throughout the paper we work with the reduced form payoff (1), to clarify the economic interpretation of the parameters α and β it is useful to develop a model microfounding the reduced form (1). In the next part we present the core of such a model in the context of capital arbitrage: this is our leading microfoundation. A richer model is presented in Section 4.1.1. However, note that there are many other potential microfoundations of this reduced-form model. The critical feature is that each player's pay-off is lower if better types also enter, while worse entrants can both help or hurt. In Online Appendix B, we describe potential examples such as production decisions with spill-overs, consumption decisions when early adoption of a trend gives private benefit, and academic publication tournament.

2.2 Microfoundation: Capital arbitrage

There are two islands A and B indexed by $i \in \{A, B\}$. There are two types of agents: a worker on each island and a continuum of investors with mass M uniformly distributed over types denoted by $\theta \in [0, 1]$. There are three types of goods: capital, a specialized consumption good produced by capital, and a numeraire good. The numeraire good is used as a method of exchange and all investors are endowed with sufficient numeraire goods to make transactions possible. Time is continuous and denoted by $t \in [0, 1]$. At $t = 0$ capital is inefficiently distributed, the worker on island A is endowed with $k_{A,0}$ capital, while the worker on island B has none ($k_{B,0} = 0$). We think of island B as an emerging idea/industry/country representing a profitable investment opportunity found sequentially by investors.

At time t, investor of type $\theta = t$ has the opportunity to buy one unit of capital on island A, transport it to island B and sell it there. For simplicity, we assume workers do not have any bargaining power and buy and sell capital at a price equal to its marginal product at the given capital level at the time.[2] Production only happens at $t = 1$: on island i, k_i capital produces c_i consumption good according to the production function $c_i = \gamma \cdot k_i - \delta \cdot \frac{k_i^2}{2}$, thus on island i the marginal product or value of capital, given the capital level at time t, becomes:

$$\frac{dc_i}{dk_{i,t}} = \gamma - \delta \cdot k_{i,t}. \tag{2}$$

[2]This is a shortcut to capture that workers know even less than investors and thus cannot capture any of the surplus generated by trade.

All consumption happens at $t = 1$ after production has been completed. Investors only consume the n_j numeraire goods with utility $U_{investor} = n_j$. Worker i consumes n_i numeraire goods it possesses and c_i consumption goods it produces with utility $U_i = n_i + c_i$.

Lemma 1. ***Microfoundation of reduced form parameters.*** *Choosing $k_{A,0} = \frac{1}{\delta}$, the expected payoff of investor θ from transporting capital (given that investor θ can enter at time t) simplifies to (1) with $\alpha = 0$ and $\beta = 2\delta$. This results in strictly positive crowding and rat-race parameters $\beta - \alpha = \alpha + \beta = 2\delta > 0$.*

Note that both our rat-race and our crowding properties are driven solely by δ which measures the extent of decreasing return to scale. That is, entering early is beneficial because the investment opportunity is more profitable when not many investors have entered yet. This shows that in financial market it is natural to assume that crowding comes hand-in-hand with a rat race even in the absence of any externalities. While in this simple setup $\alpha = 0$, in Section 4.1.1, we show how introducing shocks in this same model leads to $\alpha \neq 0$.

2.3 Learning cost based on entropy

Before entry, investors can engage in costly learning about their type. Observe that if $H(\cdot)$ is any intuitive measure of uncertainty then $H(\theta) - H(\theta|s)$, the reduction of uncertainty after observing signal s, is a measure of learning induced by signal s. Following Sims (1998), we measure uncertainty by specifying $H(\cdot)$ as the Shannon-entropy of a random variable.[3] Therefore, we specify the cost of learning a signal s as being proportional to the induced reduction in entropy of $\theta : H(\theta) - H(\theta|s)$. This quantity is often called as the mutual information in θ and s. Sims (1998) argues that the advantage of such a specification is that it both allows for flexible information acquisition and can be derived based on information theory. Note that the payoff (1) for a given θ in our model is linear in entry.

[3]The entropy of a discrete variable is defined as $\sum_x P(x) \log \frac{1}{P(x)}$, where the random variable takes on the value x with probability $P(x)$, see MacKay (2003).

Woodford (2008) derives the optimal signal structure and entry decision rule for such problems which we restate in the lemma below.

Lemma 2. *The optimal signal structure is binary: investors choose to receive signal* $s = 1$ *with probability* $m(\theta)$ *and* $s = 0$ *with probability* $1 - m(\theta)$*, given their type* θ*. The optimal entry decision conditional on the signal is: enter if* $s = 1$*, stay out if* $s = 0$*.*

Thus, similar to Yang (2015a), $m(\theta)$ is the only choice variable (learning and entry strategy combined) which in turn is the conditional probability of entry. The intuition for the binary signal structure is that the only reason investors want to learn about θ is to be able to make a binary decision of whether or not to enter. Given the linearity of the problem, the "cheapest" signal to implement the optimal entry strategy is also binary, it simply tells the investor whether or not to enter.

We now write the cost of learning, defined by the reduction in entropy, in case of a binary information structure. Denote the amount of learning L using the mutual information in signal s defined in Lemma 2 and in θ,

$$\begin{aligned} L(m) \equiv\ & H(\theta) - H(\theta|s) = H(s) - H(s|\theta) = \\ & \left(-p \log\left[\frac{1}{p}\right] - (1-p)\log\left[\frac{1}{1-p}\right]\right) - \int_0^1 \left(-m(\theta)\log\left[\frac{1}{m(\theta)}\right] - (1-m(\theta))\log\left[\frac{1}{1-m(\theta)}\right]\right) \mathrm{d}\theta \end{aligned} \tag{3}$$

where p denotes the unconditional probability of entry and is defined by:

$$p = \int_0^1 m(\tilde{\theta})\mathrm{d}\tilde{\theta} \tag{4}$$

and the first equation is a property of Shannon-entropy. The expression for learning (3) can be understood in the following way. There is no learning if the signal is uninformative of the state, that is, if it prompts the investor to enter with probability p unconditional on its type θ. Indeed, it is easy to check that when $m(\theta)$ is constant at p then $L(m) = 0$. Thus, learning depends on how much information the signal contains of the state. Intuitively, the steeper $m(\theta)$ becomes in θ (keeping average entry p constant), the more the investor is differentiating its entry decision according to its type and the higher the entropy reduction, thus the higher the learning cost. The highest cost is achieved when $m(\theta)$ is a step function. Note that L is bounded from above but might generate infinite marginal cost of learning.

Our measure of cost of learning induced by a signal defined in Lemma 2 is $\mu \cdot L(m)$ where μ is an exogenous marginal cost parameter. We assume that investors have to decide about the amount of information acquisition ex ante without any knowledge about the action of others. We interpret this as the cost of building an information gathering and evaluation "machine" which includes the costs of gathering and optimally evaluating the right data. While this machine might collect and evaluate information dynamically, none of the trader can interfere with its "code" once it is in operation. Each investor waits until the machine gives them a signal to invest or not and proceed accordingly.

Conveniently, standard results in information theory implies that the entropy of a random variable is proportional to the average number of bits needed to optimally convey its realization. Hence, the parameter μ can be interpreted as the cost of building a marginally larger information gathering and evaluating machine or writing a longer "code".[4] We believe that given today's financial markets where professional traders invest vast resources in systems and practices of mapping large amount of data into investment strategies, this might be the appropriate modelling approach.

Note also that μ might vary with the nature of the trading strategy. With low μ it is easy for the investor to determine how many investors have entered before, e.g. because the price's relation to the fundamentals reveals this. Examples for a trade like this would be that of twin stocks or on-the-run-off-the-run bonds: it is clear from the price difference whether an investor is early (large price gap) or late (small price gap). Another example is merger arbitrage, where the price offered by the bidder is known. On the other hand, with high μ, it is very hard for the investor to determine

[4] An alternative would be to think of capacity as limited and μ being the Lagrange multiplier of the capacity constraint. We choose to use a fixed μ instead of a fixed capacity because we think in this context learning capacity can be expanded freely at a fixed marginal cost: e.g. it is always possible to hire new staff or allocate more attention to this specific trade at the expense of other trades.

whether to enter, e.g. because there is no clear price signal whether the trade is still profitable. In the language of Stein (2009), high μ represents unanchored strategies. Examples for such trades include: emerging markets, carry trade, January effect. While the assumption that learning from prices is also a part of a more general learning process makes the problem very tractable, it is not without loss of generality. It might be that certain θ's are easier to learn from prices than others: we leave this theoretical question for future research.

We only consider learning about the private information θ in this model, investors do not learn about any common fundamentals of the investment. However, one can interpret the private signal as learning about the fundamental from the perspective of the specific investor. They learn how good the fundamentals are at the moment when they learn about the investment possibility.

2.4 Definition of allocative efficiency, speed of capital and welfare

In this part, we define our main economic objects of interest.

Note first, that under symmetric strategies, the mass of lower types entering ("before" investor θ) becomes:

$$b(\theta) = M \cdot \int_0^{\theta} m(\tilde{\theta}) \mathrm{d}\tilde{\theta} \tag{5}$$

mass of higher types entering ("after" investor θ):

$$a(\theta) = M \cdot \int_{\theta}^{1} m(\tilde{\theta}) \mathrm{d}\tilde{\theta} \tag{6}$$

thus $M \cdot p = b(\theta) + a(\theta)$ is the aggregate entry of investors.

The expected revenue of an investor, before taking into account the cost of information acquisition, is:

$$R \equiv \int_0^1 m(\theta) \cdot \Delta u(\theta) \cdot \mathrm{d}\theta \tag{7}$$

Recall from section 2.2 that in our leading application, investors gain in the aggregate if and only if they reduce the difference in marginal returns of capital across locations, i.e. equate the price of capital in the two regions. Therefore, in this economy, aggregate revenue $M \cdot R$ can be also interpreted as a measure of efficiency of capital allocation or allocative efficiency.

The total expected payoff (value) per unit of investor is the revenue from entering, net of the ex ante cost of learning:

$$V \equiv R - \mu \cdot L \tag{8}$$

which is what investors maximize. In equilibrium all surplus ends up with the investor: workers capture none, since they are myopic and sell/buy capital at a price equal to its marginal product. Thus the overall welfare in the whole economy can be computed as $W \equiv M \cdot V$.

The speed of capital can be measured many ways. We choose to measure it by the inverse of the half-time $\frac{1}{\tau}$, where τ is the $\theta = \tau$ (time) by which half of the capital has been transported. The lower τ, the faster the capital is. τ is formally defined by:

$$\int_0^{\tau} M \cdot m(\theta) \mathrm{d}\theta = \frac{M \cdot p}{2} \tag{9}$$

3 Model Solution

In this section we present our main results. First, we formulate the investor's problem. Second, we derive the optimal strategies of investors for general levels of learning cost. Third, we analyze how aggregate entry, speed and welfare changes as the mass of investors increases.

3.1 Optimal strategies

The private problem of any investor is to choose its conditional entry $m(\theta)$ to maximize its value function V, which can be written as the following:

$$\max_{m(\theta)} \int_0^1 (m(\theta) \cdot \Delta u(\theta) - \mu \cdot L(m))\, \mathrm{d}\theta. \tag{10}$$

We contrast the private solution with that of a social planner who can choose the amount of learning and entry for all investors. This gives us a benchmark against which we can evaluate learning and entry decisions in the competitive equilibrium. The main difference between the competitive solution and the social planer's one is that the social planner takes into account the externalities that investors exert on each other: it takes into account that Δu depends not only on θ but on the choice function of all other investors m. The social planner chooses the symmetric function $m_s(\theta)$ to maximize

$$\max_{m_s(\theta)} \int_0^1 (m_s(\theta) \cdot \Delta u(\theta, m_s)\mathrm{d}\theta - \mu \cdot L(m_s))\, \mathrm{d}\theta \tag{11}$$

We derive the first order condition (FOC) of these problems using the variation method, i.e. we look for the function $m(\theta)$ such that if we take a very small variation around the function, the value function of the investors does not change.

Lemma 3. ***First order conditions.*** *Denote the strategy function of all other players as $\tilde{m}(\theta)$. The first-order condition of the competitive problem is:*

$$M \cdot \alpha \cdot \int_\theta^1 \tilde{m}(\tilde{\theta})d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \tilde{m}(\tilde{\theta})d\tilde{\theta} + 1 = \mu \cdot \left[\log\left(\frac{m(\theta)}{1-m(\theta)}\right) - \log\left(\frac{p}{1-p}\right)\right]. \tag{12}$$

The first-order condition of the social problem (assuming the same entry function m_s for all investors) is:

$$M \cdot (\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[\log\left(\frac{m_s(\theta)}{1-m_s(\theta)}\right) - \log\left(\frac{p}{1-p}\right)\right] \tag{13}$$

Using the FOC one can derive an ordinary differential equation for $m(\theta)$ where the original FOC at $\theta = 0$ (an integral equation) is the boundary condition. The solution of this ordinary differential equation can be expressed up to a constant (boundary value) $m(0)$.

Proposition 1. ***Competitive entry strategies.*** *If M is below a threshold $M \leq \bar{M}$, all investors enter without learning $m(\theta) = 1$. If M is above the threshold $M > \bar{M}$, the optimal entry function is given by:*

$$m(\theta) = \frac{1}{1 + W_0\left(e^{M \cdot \frac{\alpha+\beta}{\mu} \cdot \theta + \frac{1-m(0)}{m(0)} + \log\left(\frac{1-m(0)}{m(0)}\right)}\right)}, \tag{14}$$

where W_0 denotes the upper branch of the Lambert function[5] *and $m(0)$ is pinned down by the boundary condition ((12) evaluated at $\theta = 0$):*

$$M \cdot \alpha \cdot p + 1 = \mu \cdot \left[\log\left(\frac{m(0)}{1 - m(0)}\right) - \log\left(\frac{p}{1-p}\right)\right]. \tag{15}$$

The threshold $\bar{M}$ is pinned down by the following implicit equation:

$$\frac{\bar{M} \cdot (\alpha + \beta)}{\mu} = e^{-\frac{1-\beta \cdot \bar{M}}{\mu}} - e^{-\frac{1+\alpha \cdot \bar{M}}{\mu}}. \tag{16}$$

Proposition 2. ***Socially optimal entry strategies.*** *If M is below a threshold $M \leq \bar{M}_s = \frac{1}{\beta - \alpha}$ then all investors enter $m_s(\theta) = 1$. If M is above the threshold $M > \bar{M}_s$ then the socially optimal entry function is flat in θ:*

$$m_s(\theta) = \frac{1}{M \cdot (\beta - \alpha)}. \tag{17}$$

Note that investors want to differentiate between states, but the planner does not. The planner chooses a flat entry function. The reason for this over-learning is the rat race ($\alpha + \beta > 0$) that accompanies the crowding: every investor wants to know whether he is ahead of the other investors even if this is wasteful from the social planner's point of view.

[5]The definition of the upper branch of Lambert function is $z = W_0(z) \cdot e^{W_0(z)}$ if $z > 0$.

To better understand the optimal strategies, we first look at the extreme cases of $\mu \to 0$ (full information) and $\mu \to \infty$ (no information).

Lemma 4. ***Entry under full and no information.*** *For full information ($\mu = 0$), the competitive functions $m(\theta)$ is a step function, resulting in the first $M \cdot p$ investors entering. In the social planner's optimum, $M \cdot p_s$ mass of investors enter but many (symmetric) strategies are permissible, e.g. all investors entering with unconditional probability p_s. The aggregate amount of entrants in competitive and social planner's optimum are, respectively:*

$$M \cdot p|_{\mu=0} = \min\left(\frac{1}{\beta}, 1\right) \tag{18}$$

$$M \cdot p_s|_{\mu=0} = \min\left(\frac{1}{\beta - \alpha}, 1\right) \tag{19}$$

For no information ($\mu \to \infty$), both the competitive and social planner's entry functions $m(\theta)$ are flat. All investors enter with the same unconditional probability. The aggregate amount of entrants in competitive and social planner's optimum are, respectively:

$$M \cdot p|_{\mu\to\infty} = \min\left(\frac{2}{\beta - \alpha}, 1\right) \tag{20}$$

$$M \cdot p_s|_{\mu\to\infty} = \min\left(\frac{1}{\beta - \alpha}, 1\right). \tag{21}$$

Under full information, whether there is under- or over-entry, compared to the social planner's choice, depends on the sign of α. There is competitive under-entry (under-entry) if $\alpha > 0$ ($\alpha < 0$), since investors with higher θ do not take into account the positive (negative) effect of their entry that accrues to entrants with lower θ.

Under no information, there is competitive over-entry under any parameter values: investors enter twice as often than they should. The intuition is analogous to the "tragedy of commons". While each investor internalizes that if others enter more often, that reduces its own revenue, it does not internalize that when she enters that reduces the benefit of entry for everyone else. Note that there is over-entry even in our benchmark microfoundation without externalities where $\alpha = 0$ In this particular case, the "externality" comes from the decreasing returns to scale technology and the lack of information, like a well observable price.

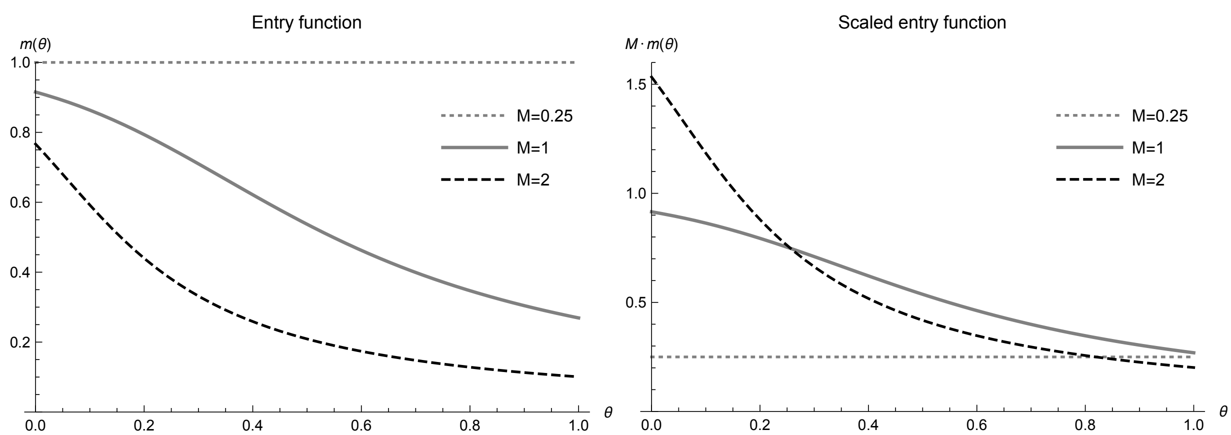
Lemma 4 makes it clear that simply increasing the mass of investors in a market does not mean that entry converges to the socially optimal level. Using the above analysis, on can draw implications about specific markets. Trades where it is easy to learn (low μ) with high α do not have enough investors entering. One example for such a market is twin stocks: the price difference between the two stocks reveals whether it is still profitable to enter but early entrants need later entrants to be able to exit the trade with a profit. Thus the model can give a potential explanation of why there is insufficient entry into trades like twin stocks and why mispricing persists. On the other hand, it also shows why trades where it is hard to learn (high μ), such as carry trade or momentum, might see too much entry: in the extreme case of $\mu \to \infty$ even driving the revenue of investors to zero, like in a tragedy of commons game.

To understand the effect of M on incentives in the general solution, see Figure 1 which shows the competitive and social planner's optimal entry function for different levels of M. The left panel shows the unscaled functions, while the right panel is scaled by M, thus showing the aggregate entry by type in equilibrium. For small $M = 0.25$, investors enter for sure and there is no need for learning since revenues in the market are high. To gain intuition on how $m(\theta)$ changes as M increases from 1 to 2, consider the effect of larger M on the benefit of entry $\Delta u(\theta)$ for a given investor, keeping the others' strategy constant. First, note that we can measure the relative incentive for entering earlier by differentiating $\Delta u(\theta)$ in θ, giving $M \cdot (\alpha + \beta) \cdot \tilde{m}(\theta)$. Therefore, keeping other investors' strategies fixed, the incentive to learn more and follow a more differentiated strategy is increasing in M. Loosely speaking, this results in a steeper $m(\theta)$ as it is apparent on the right panel of Figure 1. Given that this effect is scaled by the rat race parameter, $(\alpha + \beta)$, we refer to this as the rat race effect. Second, note that the benefit of entry for the average investor is $\Delta u\left(\frac{1}{2}\right) = M \cdot \alpha \cdot \int_{\frac{1}{2}}^{1} \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_{0}^{\frac{1}{2}} \tilde{m}(\tilde{\theta}) \mathrm{d}\tilde{\theta}$, hence, keeping others' strategy constant

$$\frac{\partial \Delta u\left(\frac{1}{2}\right)}{\partial M}=\alpha\cdot\int_{\frac{1}{2}}^{1}\tilde{m}(\tilde{\theta})\mathrm{d}\tilde{\theta}-\beta\cdot\int_{0}^{\frac{1}{2}}\tilde{m}(\tilde{\theta})\mathrm{d}\tilde{\theta}<(\alpha-\beta)\frac{p}{2}<0$$

where the first inequality comes from the fact that $\tilde{m}(\theta)$ is decreasing in equilibrium. This suggests that for the average investor with $\theta=\frac{1}{2}$, increasing M is deceasing the incentive to enter as it is apparent in the left panel of Figure 1 Given that this effect is scaled by the crowding parameter $(\alpha-\beta)$, we refer to this as the crowding effect. While in equilibrium the strategy of other investors, $\tilde{m}(\tilde{\theta})$, also changes, implying further adjustments, as Figure 1 demonstrate, the total effect is still driven by this intuition.

Figure 1: **Competitive entry functions for different levels of competition**



Entry functions for the competitive entry (m, left panel) and scaled competitive entry function ($M\cdot m$, right panel) for $M=0.25$ (dotted line), $M=1$ (solid line), and $M=2$ (dashed line). Parameters: $\beta=4$, $\alpha=2$, $\mu=0.5$.

3.2 Allocative efficiency: over- and under-entry

Since the allocative efficiency of capital in the markets depends on the overall entry of all investors, in this subsection we analyze how aggregate entry $M\cdot p$ changes as the mass of investors M grows.

Proposition 3. ***Entry in the competitive and socially optimal solution.*** *The competitive aggregate entry is* $M \cdot p = \min\left(M, \bar{M}\right)$*. The aggregate entry in the social planner's solution is* $M \cdot p_s = \min\left(M, \frac{1}{\beta-\alpha}\right)$.

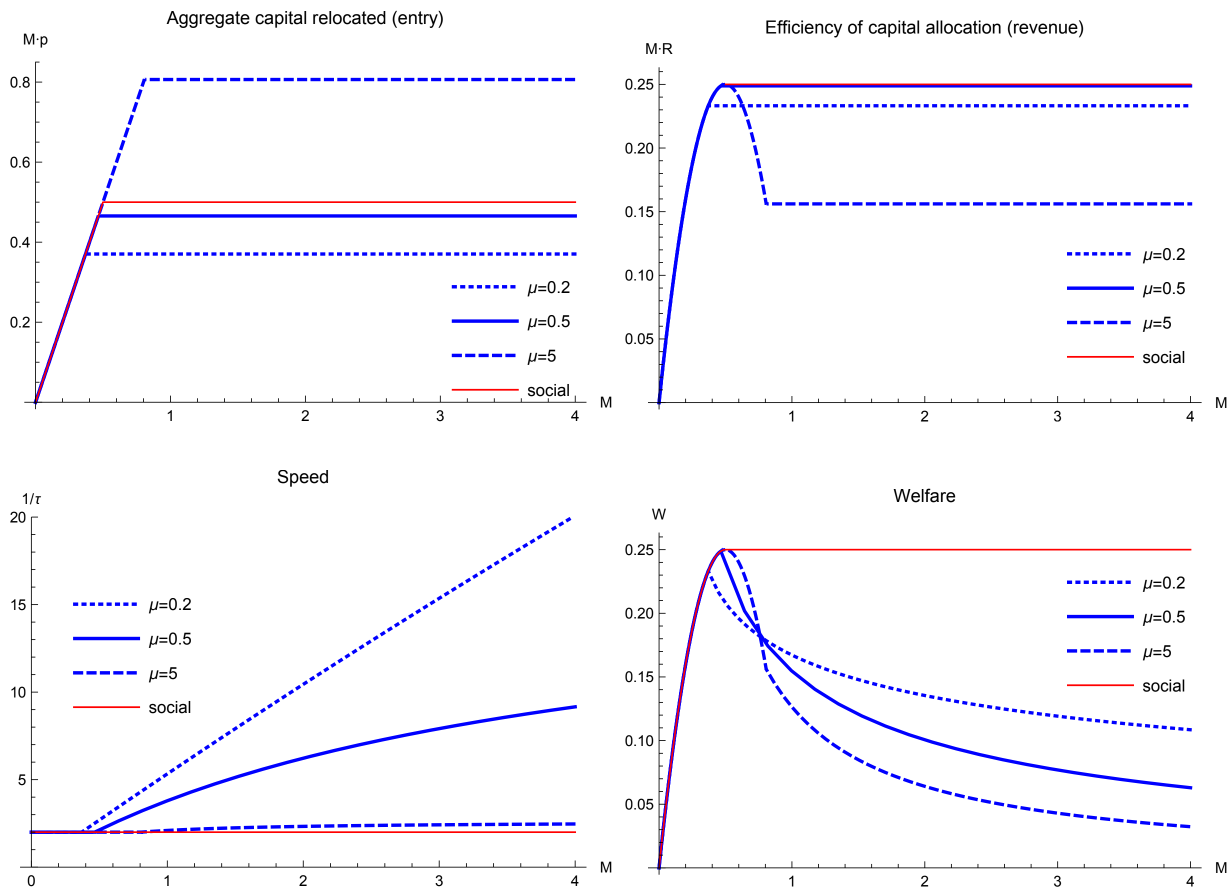
Just as in the planner's solution, investors in the competitive equilibrium also enter with probability one for small M and aggregate entry $M \cdot p$ is constant when M is large. However, that constant level, $\bar{M}$, is in general different from the social optimum $\bar{M}_s = \frac{1}{\beta-\alpha}$. That is, whenever $M > \bar{M}$, increasing the number of investors neither improves the efficiency of capital allocation, nor does it lead to additional crowding. Figure 2 illustrates this part by showing the amount of total entry $M \cdot p$ as a function of the mass of investors M.

For the intuition, recall that changing M changes the optimal strategy $m(\theta)$ for every investor through the rat race effect and the crowding effect. As the rat race effect primarily affects the slope of the entry function, as opposed to its level, it has little influence on $M \cdot p$. In contrast, due to the crowding effect the average entry, p, decreases. In equilibrium, the decrease in p is exactly proportional to the increase in M, keeping $M \cdot p$ constant. This can be observed on Figure 1: even though the entry functions look very different in case of $M = 1$ and $M = 2$, the areas under the scaled entry functions, i.e. aggregate entry, are the same.

Allowing investors to flexibly choose their information structure is crucial in generating the result of constant entry as the mass of investors increases. With flexible learning the investors can optimally devise their information to exactly counter the increase in the mass of investors and thus enter at a constant aggregate rate. When learning is constrained, this is not necessarily the case: we demonstrate this in Section 4.2 in which investors can only buy Gaussian signals about their type subject to the same entropy cost as before.

We formally analyze the effects of model parameters on allocative efficiency in Section 4.1.1. For now, just note that the cost of learning μ does not only influence the amount of entry but also the welfare for a given mass M of investors. Figure 2 gives us some insights: First, for high levels of M, easier learning (lower μ) means higher welfare W. This holds irrespective of the fact that a very low

Figure 2: **Real outcomes as a function of investor competition**



Aggregate entry, revenue, speed and welfare as a function of the mass M of investors allowed to invest. The thin solid line is the social optimum, the thick lines are the competitive outcomes for three different values of μ: $\mu = 0.2$ (dotted line), $\mu = 0.5$ (solid line), $\mu = 5$ (dashed line). Parameters: $\beta = 4$, $\alpha = 2$.

μ might lead to less entry than the social optimum. The reason is that with many investors, they all have to spend an increasing fraction of their revenues on learning in order to stabilize entry and this is more costly if the marginal cost of learning μ is high. This also highlights that the possibility to learn is beneficial from a welfare point of view, especially if the cost is not that high. Second, for lower mass M of investors, welfare might be higher when learning is more expensive. The intuition here is that for $\alpha > 0$, higher learning cost μ deters learning and thus helps avert under-entry.

3.3 Decoupling of welfare, speed and allocative efficiency

Now we turn to the question of welfare as more and more sophisticated investors enter. We show that the presence of some investors ($M < \bar{M}$) unambiguously increases welfare in the competitive equilibrium. Note that in the small M case, welfare does not depend on μ since no resources are spent on learning. The total mass of sophisticated investors is small in this range, hence they do not try to beat each other by learning about their relative type. Instead, all decide to enter without putting resources in learning. As their mass is marginally increasing, in terms of our microfoundation, they are able to allocate more capital to the new market, which increases the efficiency of capital allocation. Thus allocative efficiency and welfare go hand-in-hand. The above insight that larger M means (at least weakly) higher welfare and a more efficient capital allocation remains true in case of the social planner's optimum since no learning is chosen in that case.

In the competitive equilibrium, raising M above $\bar{M}$ leads to decoupling of welfare and allocative (or market) efficiency: while allocative efficiency stays constant (even though at a suboptimal level), welfare decreases, see Figure 2. The reason is that as the amount of investors in the market grows, they start worrying about crowding and thus their relative type θ, inducing them to learn about it. A rat race ensues with increasing amounts invested in learning and reduced welfare. Thus an increasing mass of sophisticated investors leads to a drop in welfare not because of crowding (the total amount of investors entering is constant) but because of increased spending on learning.

Proposition 4. ***Welfare.*** *If $M > \bar{M}$, the efficiency of capital allocation (aggregate revenue of investors) stays constant as we increase M. However, welfare becomes decoupled from allocative efficiency, welfare converges to zero from above as $M \to \infty$:*

$$W(\bar{M}) > \lim_{M\to\infty} W(M) = 0 \tag{22}$$

The welfare in the social planner's optimum for $M > \bar{M}_s$ is constant:

$$W_s(M) = \frac{1}{2 \cdot (\beta - \alpha)}. \tag{23}$$

Note that learning is useful in limiting crowding, albeit at a cost. One can see this by comparing the above positive welfare for any $M > 0$ with the case of $M > \frac{2}{\beta - \alpha}$ when learning is prohibitively expensive as $\mu \to \infty$. In that case, investors do not learn but enter until their payoff is zero, leading to zero welfare. In the context of our structural model, the above result also implies that increasing the mass of investors M from below $\bar{M}$ to above $\bar{M}$ might make markets more efficient from an allocative point of view and decrease welfare at the same time. This is due to the fact that increased allocative efficiency is achieved at the cost of spending on learning.

Interestingly, the costly over-learning in our model manifests itself in increasing speed of capital. The right panel of Figure 1 gives the intuition why this increasing speed is indeed a side-effect of more learning. As more learning implies a steeper scaled entry function $M \cdot m(\theta)$, early types enter with higher probability, while late types enter with lower probability. Therefore, in the aggregate, each unit of capital is reallocated sooner. We state the formal result in the next proposition.

Proposition 5. ***Speed of capital.*** *If the mass of investors is small $M \leq \bar{M}$, the speed of capital reallocation in the competitive solution is $\frac{1}{\tau} = 2$. If the mass of investors is large $M > \bar{M}$, the speed of capital reallocation in the competitive solution is higher $\frac{1}{\tau} > 2$ and as the number of investors increases, in the limit it converges to:*

$$\lim_{M \to \infty} \frac{1}{\tau} = 1 + e^{\frac{\bar{M} \cdot (\alpha + \beta)}{2\mu}} \tag{24}$$

The speed of capital reallocation in the social planner's optimum is always $\frac{1}{\tau_s} = 2$.

Thus increasing the number of investors increases the speed at which capital is reallocated. Note, however, that in our baseline model, there is no social benefit of this increased speed. In fact speed just destroys welfare since more information is necessary for higher speed and learning is costly. In Section 4.1.2 we introduce a variation in the model in which early entry, thus speed, is valuable and show that this does not change any of the major insights.

Contrary to what one might expect, the equilibrium speed is not infinite even with a large amount of investors, i.e. τ does not converge to zero. The intuition is that the amount of aggregate learning is bounded from above by the total revenue. If only the very first were to enter, that would necessitate large amounts of aggregate learning. We further discuss how the equilibrium speed depends on deep parameters of the model in Section 4.1.1.

If there are other ways to limit entry without learning, that might be welfare improving. In fact if we limit the mass of investors indiscriminately, before they learn their type, one can improve welfare. The intuition is that limiting entry decreases the effective M and thus limits the incentive to learn. Note that in a more general model where workers (or consumers) benefit from entry of investors, the welfare analysis changes, see Online Appendix B.4.

4 Discussion and extensions

4.1 The payoff function

We now analyze the payoff function in more detail. First, we provide a more detailed microfoundation with direct externalities that can generate reduced form parameters $\alpha \neq 0$. We use this to analyze how deep parameters of the model effect the amount of entry in the model. Second, we add a term to the reduced form payoff function such that early entry, thus speed is more valuable.

4.1.1 Introducing shocks in the microfoundation

Remember from the baseline microfoundation in Section 2.2, that production only happens at $t = 1$: on island i, k_i capital produces c_i consumption good according to the production function $c_i = \gamma \cdot k_i - \delta_i \cdot \frac{k_i^2}{2}$, thus on island i the marginal product of capital, given the capital level at time t, becomes:

$$\frac{dc_i}{dk_{i,t}} = \gamma - \delta_i \cdot k_{i,t}. \tag{25}$$

In the present extension, we allow δ_i to be state-contingent.

To explore the effect of various direct externalities in our problem, we consider the possibility that the capital reallocation is subject to shocks. In particular, suppose that the transport is successful only with probability $1-\nu$, with probability $\nu \geq 0$, the investor is hit by an idiosyncratic ("liquidity") shock and has to go back to island A and sell the capital there at $t = 1$. Furthermore, with ex ante probability $\eta \geq 0$, even if the capital transfer is successful, upon arriving on island B, there is an aggregate shock ("crisis") and all investors have to sell their capital in a fire sale. While on island A, we still always have $\delta_i = \delta$, on island B, $\delta_i = \delta$ only if there is no crisis, but $\delta_i = \delta_c + \delta$ if there is a crisis (aggregate shock) where $\delta_c > 0$.

Note that unless there is an idiosyncratic shock, investors buy capital at its marginal return on island A and sell capital at its marginal return on island B. Whenever this activity is profitable, it also decreases the difference between the marginal return on capital across the two islands, that is, it increases market efficiency. Idiosyncratic shock complicates this picture only to the extent that it introduces some redistribution among investors; an element which washes out by aggregation. Therefore, we can still interpret the aggregate revenue of investors as a measure of market efficiency.

We interpret our deep parameters as follows. δ captures the extent of decreasing return to scale in each market. Conceptually, this is a technological parameter of the sectors or firms which are subject to the capital reallocation. δ_c characterizes the depth of the financial market where claims on these firms and sectors are traded. When δ_c is large, a sudden selling pressure of the participating investors drives the price down significantly. In contrast, η and ν characterizes the investors as opposed to the markets. A large η is interpreted as a large probability for a common liquidity shock for all investors, for example, because of large common exposure to risk factors outside of the model. A large ν is a large probability of an idiosyncratic liquidity shock, for example, because investors are professional investors with a volatile investor base. The below lemma shows that this model is equivalent to the reduced form payoff of Section 2.2.

Lemma 5. ***Reduced form parameters with shocks.*** *The microfounded model with shocks is equivalent to the reduced form model with parameters:*

$$\alpha = \left(\frac{1}{2} + (1-\nu)^2\left(\frac{(1-\eta)}{2} - 1\right)\right)\cdot\delta - \frac{1}{2}(1-\nu)^2\cdot\eta\cdot\delta_c \tag{26}$$

$$\beta = \left(\frac{1}{2} + (1-\nu)^2\left(\frac{(1-\eta)}{2} + 1\right)\right)\cdot\delta + \frac{1}{2}(1-\nu)^2\cdot\eta\cdot\delta_c \tag{27}$$

The above parameters satisfy the rat race $\beta - \alpha > 0$ and crowding $\beta + \alpha > 0$ properties.

To interpret α and β, it is useful to first consider the case without idiosyncratic shock $(\nu = 0)$. In this case, $\alpha = -\frac{1}{2}\eta\,(\delta + \delta_c)$ and $\beta = 2\delta + \frac{1}{2}\eta(\delta_c - \delta)$, implying $\beta - \alpha = 2\delta + \eta\cdot\delta_c$ and $\alpha + \beta = (2-\eta)\cdot\delta$. Note that the crowding parameter, $\beta - \alpha$, is increasing in the probability of the aggregate liquidity shock, η, and the illiquidity of the market, δ_c. That is, entrants impose a negative externality on each other, because it is more costly to exit when more investors want to exit at the same time. Finally, note that α is negative without idiosyncratic shock, because of the same logic: late entrants harm early entrants because they aggravate crowding. Without idiosyncratic shock, the effect of more late entrants in a liquidity crisis is the same as the effect of more entrants, late or early.

While the introduction of idiosyncratic shock affects all our reduced form parameters, its main qualitative effect is that it changes the sign of α. Indeed, α is monotonically increasing in ν, reaching $\frac{1}{2}\cdot\delta > 0$ when $v = 1$. The intuition is that for large v early entrants benefit from late entrants since if they have to liquidate their position, they can do so at a higher price. This means that α is likely to be positive in markets where entrants need enough subsequent liquidity to exit at a reasonable price.

Comparing the cases with and without shocks, it is clear that in both cases (1) (with positive crowding and rat-race parameters) is a suitable reduced form representation of payoffs. The main qualitative difference is that without shocks $\alpha = 0$, while with shocks α can be both negative and positive. We show in Section 3.2 that in the case with shocks, changing the parameters affecting α and β help understand how pay-off externalities affect the learning and entry decisions of our investors.

In Proposition 3 we showed that whether there is under- or over-entry for $M > \bar{M}$ is independent of the mass of investors considering to enter. Instead, as we state in Proposition 6, it is determined by all the other characteristics of the capital reallocation problem.

Proposition 6. ***Comparative statics of crowding.*** *Under certain assumptions*[6] *(see proof), one can show that if there is a sufficient mass $M > \max[\bar{M}, \bar{M}_s]$ of investors, the relative amount of competitive aggregate entry to social aggregate entry $\frac{\bar{M}}{\bar{M}_s}$ is*

1. *increasing in μ, the marginal cost of information*
2. *decreasing in δ, the rate of decreasing returns to scale of the technology in the presence of aggregate shocks, i.e. if $\eta > 0$*
3. *increasing in δ_c, i.e. decreasing in market depth in crisis in the presence of aggregate shocks, i.e. if $\eta > 0$*
4. *increasing in η, the probability of aggregate shocks*
5. *decreasing in ν, the probability of idiosyncratic shocks*

More frequent aggregate liquidity shocks (larger η) and less market depth (higher δ_c) make markets more crowded since they increase fire sales externalities. More costly information leads to more crowding, because the game is closer to a tragedy of commons problem as explained in Section 3.1. A slower decrease in marginal product of capital (higher δ) in the technology also makes the market more crowded in the presence of aggregate shocks. On the other hand, more frequent idiosyncratic liquidity shocks (larger ν) makes the market less crowded since it leads to under-entry due to late entrants not internalizing the positive effect they have on earlier entrants.

[6] Note that both Propositions 6 and 7 are stated as holding under certain conditions. The conditions are stated explicitly in the proof and are employed only to facilitate the proof. In fact we found no admissible parameters for which the assumed conditions do not hold, however, we could not prove this explicitly.

In Proposition 5 we showed that the speed of capital allocation converges to a constant as we increase the mass of investors. In the following Proposition 7, we show how the equilibrium speed in the limit is determined by the deep parameters of the market and the investors.

Proposition 7. ***Comparative statics of speed.*** *Under certain assumptions (see proof), one can show that in the limit of a large mass of investors ($M \to \infty$), the equilibrium speed of entry $\frac{1}{\tau}$ is*

1. *decreasing in μ, the marginal cost of information*
2. *increasing in δ, the rate of decreasing returns to scale of the technology in the presence of aggregate shocks, i.e. if $\eta > 0$*
3. *decreasing in δ_c, i.e. increasing in market depth in crisis*
4. *decreasing in η, the probability of aggregate shocks*
5. *increasing in ν, the probability of idiosyncratic shocks*

In the limit, with $M \to \infty$, welfare goes to zero, see Proposition 4. Thus in the limit all revenues from improving the capital allocation are used for learning. Thus it seems obvious that the easier it is to learn (lower μ), the higher the equilibrium speed of trading, since holding the amount of expenditure fixed, more can be learned at lower cost, increasing the speed. The intuition for the other results can also be understood from a similar perspective: the higher the rate of decreasing returns to scale δ, the higher revenues from capital reallocation and thus more can be spent for learning. Markets with more severe (higher δ_c) or more likely (higher η) aggregate shocks offer less revenues in expectation, again decreasing the amount spent on learning and thus equilibrium speed.

4.1.2 Introducing value of speed

Consider a modification of Equation 1 in which there is a value of capital arriving early, i.e. of low θ types transferring the capital. This means there is now social value of leaning. We measure this value by $\kappa > 0$ and write the payoff to entry as:

$$\Delta u = 1 - \kappa \cdot \theta - \beta \cdot b(\theta) + \alpha \cdot a(\theta) \tag{28}$$

Note that the baseline specification (1) is a special case of this more general formalization with $\kappa = 0$.

Proposition 8. *If speed has value ($\kappa > 0$), the competitive optimal strategy $m(\theta)$ in the symmetric equilibrium solves the differential equation*

$$(\alpha + \beta) \cdot m(\theta) + \kappa = -\mu \cdot \frac{m'(\theta)}{m(\theta) \cdot (1 - m(\theta))} \tag{29}$$

with the boundary condition

$$\alpha \cdot p + 1 = \mu \cdot \left[\log \left(\frac{m(0)}{1 - m(0)} \right) - \log \left(\frac{p}{1 - p} \right) \right]. \tag{30}$$

If speed has value ($\kappa > 0$), the socially optimal strategy $m_s(\theta)$ in the symmetric equilibrium solves the differential equation

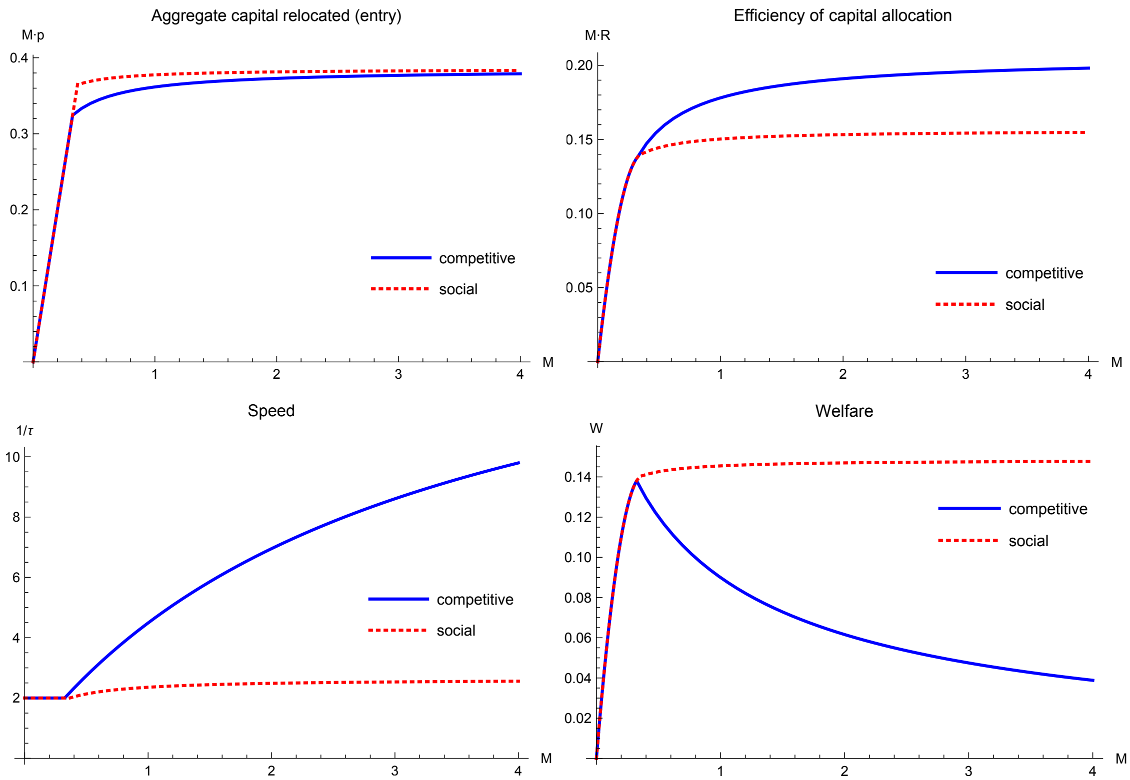
$$\kappa = -\mu \cdot \frac{m_s'(\theta)}{m_s(\theta) \cdot (1 - m_s(\theta))} \tag{31}$$

subject to the boundary condition

$$(\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[\log \left(\frac{m_s(0)}{1 - m_s(0)} \right) - \log \left(\frac{p_s}{1 - p_s} \right) \right]. \tag{32}$$

From the differential equation (29) it is obvious, that the social planner also wants to differentiate between states: m_s is no longer flat, it is also downward sloping. However, the incentive for private learning is even higher in (31) since private incentives include the rat race $\alpha + \beta > 0$: every investor wants to know whether it is ahead of the others. No closed form can be attained in general for the private solution,[7] thus we resort to numerical simulations.

Figure 3: **Competitive and social optimum if speed has value**



The red dotted lines denote the social optimum and the blue solid lines the competitive solution. Other parameters are: $\beta = 4$, $\alpha = 2$, $\mu = 0.5$, $\kappa = 0.5$.

Figure 3 shows the competitive and social outcome for different κ parameters. More learning leads to faster capital entry (lower half-time of entry) even in the social solution as speed has social benefit. Note that more investors imply higher entry in both the competitive and the social optimum because early entry has increasing benefit as the mass of investors increases and there are more potential early entrants. Competitive investors learn too much about their type, so competitive total entry increases too much with the mass of investors M. Total revenue is increasing faster in the mass M of potential entrants than entry because most of the additional revenue comes from the better timing of entry, not

[7] While (31) can be solved in closed form up to constant, to our best knowledge, (29) can only be solved for the special case $\alpha + \beta = \kappa$.

simply more entry. Investors are also motivated by the rat race so capital is sill too fast, though the welfare loss is partially offset by the welfare gain from earlier entry (lower θ entry). Nevertheless, the revenue gains from better timing of entry, which is a side-effect of over-learning, cannot offset the loss from excessive learning, thus welfare still converges to zero.

4.2 The cost function: Suboptimal Gaussian learning

In this section, we investigate how our assumptions in the baseline analysis influence our main results. In particular, we contrast our framework with fully flexible learning with a, perhaps more standard, Gaussian formalization (see e.g. Hellwig and Veldkamp (2009)): suppose that each investor observes a Gaussian signal about its type θ of a chosen precision and enter if and only if this signal is larger than a chosen threshold. We show that as long as we specify the cost of learning analogously to our baseline model, this alternative structure amounts to a restriction on the functional form of $m(\theta)$. We refer to this as the Gaussian problem and show how this restriction affects the results. In the following, $\Phi(.;\sigma)$ and $\phi(.;\sigma)$ denote, respectively, the cdf and the pdf of a normally distributed variable with zero mean and σ standard deviation. $\Phi^{-1}(.;\sigma)$ denotes the inverse of $\Phi(.;\sigma)$.

First, we introduce the transformed type variable $\eta_i = \Phi^{-1}(\theta_i;\sigma_\eta)$. Clearly, as θ_i is uniform on the unit interval, $\eta_i \sim N\left(0,\sigma_\eta^2\right)$. Investor i with type η_i can, at a cost $C\left(\sigma_{\varepsilon_i}\right)$, choose the standard deviation σ_{ε_i} of a signal $s_i = \eta_i + \varepsilon_i$ about its type where $\varepsilon_i \sim N\left(0,\sigma_{\varepsilon_i}^2\right)$ and ε_i is independent of η_i.[8] After having received the signal s_i, investor i decides whether to enter.

As for the cost of learning function, we consider two cases, both of which use the reduction in entropy. In the first specification, which we denote partial cost $C_p\left(\sigma_{\varepsilon_i}\right)$, we assume the cost is identical to our baseline specification (3) with the only exception that the entry function $m(\theta)$ resulting from the entry decision based on the received signal is restricted. Intuitively, under this specification investors pay only for the information they use for their binary actions, instead of all the information they learn.

[8] This is equivalent to drawing the type and signal from a bivariate normal: $\begin{pmatrix} \eta_i \\ s_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sqrt{\sigma_\eta^2+\sigma_{\varepsilon_i}^2} \end{pmatrix}\right)$.

Since the investor does not have to pay for unused information, we call this partial cost learning. In the second specification, which we denote full cost, we specify the cost of learning $C_f(\sigma_{\varepsilon_i})$ as the reduction in entropy in knowledge after the observation of the signal s_i, which, by the property of the normal distribution is:

$$C_f(\sigma_{\varepsilon_i}) = \frac{1}{2} \cdot \log\left(1 + \frac{\sigma_{\varepsilon_i}^2}{\sigma_\eta^2}\right). \tag{33}$$

On the one hand, this second specification means an additional departure from the baseline model since not only is the exact form of the entry function $m(\theta)$ constrained but investors have to pay for unnecessary information. On the other hand, this specification is closer to that employed in the literature.

Proposition 9. *In a symmetric equilibrium with Gaussian learning, the optimal strategy of the investor can be fully described by a choice of the signal noise σ_ε and the entry cutoff $\bar{s}$. The investor enters if and only if it receives a signal $s_i < \bar{s}$. The entry function has the form of*

$$m_G(\theta) = \Phi\left(\bar{s} - \Phi^{-1}(\theta; \sigma_\eta); \sigma_\varepsilon\right), \tag{34}$$

and the probability of unconditional entry is:

$$p = \Phi\left(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}\right). \tag{35}$$

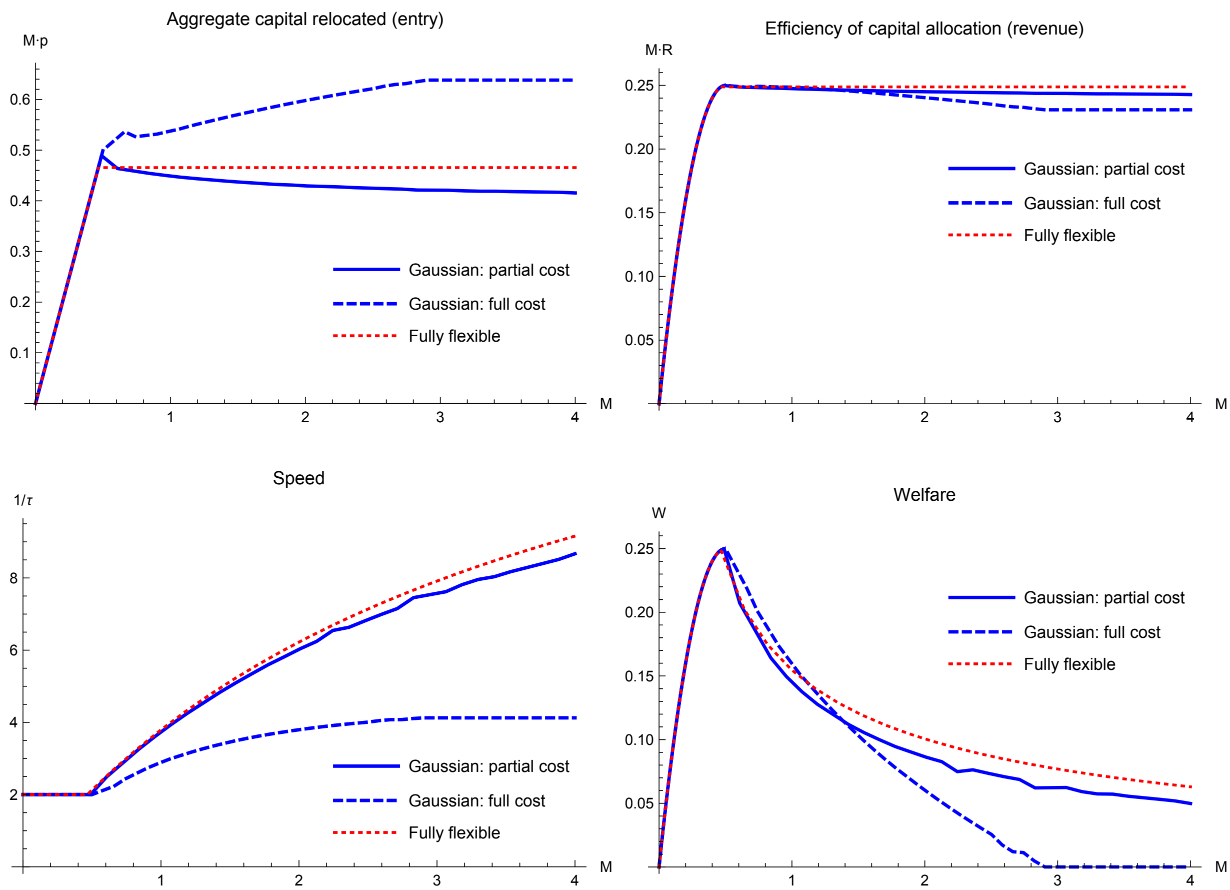
The equilibrium σ_ε and $\bar{s}$ are pinned down by their respective first-order conditions for both cost sepcification (see Appendix).

With full cost learning, if M is large enough no symmetric equilibrium exists, but there is an equilibrium with some investors learning and entering, other investors not entering at all. In this mixed-strategy equilibrium, all investors achieve zero profits.

Note that $m_G(.)$ can be fully described using two parameters: the standard deviation of signal noise σ_ε and the entry cutoff $\bar{s}$. $m_G(.)$ is thus constrained compared to the completely unrestricted choice of

$m(.)$ in the baseline model (which yields a Lambert function). The entry function is constrained due to the fact that the signal structure is constrained and that the investor has to decide on entering based on this restricted signal. Thus constraining learning automatically means constraining the entry strategies. In fact, for partial cost learning this restriction is the only deviation from the baseline model.

Figure 4: **Increasing the mass of investors using different learning costs**



Outcomes as a function of the mass M of investors allowed to invest with different cost specifications: Gaussian learning where the investors only has to pay for the information used (solid line), Gaussian learning where the investors have to pay for all reduction in entropy (dashed line), baseline model with entropy (dotted line). Parameters: $\beta = 4$, $\alpha = 2$, $\mu = 0.5$.

Since we could not attain a closed form solution, we perform a numerical analysis. Figure 4 illustrates the outcome of the Gaussian problem with both cost specifications. For comparability, we included the competitive outcome from the baseline model using the same parameters (originally in

Figure 2). For partial cost learning, there is only a small difference between the Gaussian specification and the baseline which all comes from the restriction on the form of the entry function (34) implied by the Gaussian specification. Aggregate entry is not completely flat in the mass of investors but relatively close to the benchmark.

For full cost learning, the difference is larger compared to the baseline. Total entry is monotonically increasing up to a point as an increasing mass of investors have the possibility to enter to the new market. The intuition is that learning is so expensive, due to having to pay for unnecessary information, that investors cannot learn enough and the equilibrium looks more and more like the tragedy of commons. Another interesting observation is that if M is large enough, learning s so expensive that in the symmetric equilibrium all investors entering would get negative payoffs. Thus some investors decide to stay out ex ante without even learning. In equilibrium enough stay out, such that the payoff to all learning and entering is zero, same as for those choosing the stay out. This is similar to the mixed strategy in Grossman and Stiglitz (1980), even though here it happens even though learning is a continuous choice.

To sum up, our exercise in this subsection emphasizes the importance of the flexible specification for entry and learning. It also highlights that flexible learning is more tractable than the Gaussian framework in our context. While changing the cost to be Gaussian changes the exact behavior of the observable outcomes, the main insight remains to be true: the behavior of market efficiency, speed and welfare do not coincide as we raise the amount of competition.

4.3 Heterogenous investors

In this section we consider an extension with heterogenous investors to analyze how changing the composition of investors, instead of the total mass, influences allocative efficiency and welfare. This is interesting, as the level of sophistication among investors is very heterogenous: e.g. pension funds might be less sophisticated than hedge funds. We consider two groups: $\omega \cdot M$ mass of investors is sophisticated and faces a lower learning cost of μ_L, while $(1-\omega)\cdot M$ mass of investors is unsophisticated and faces a higher learning cost of $\mu_H > \mu_L$. Both groups of investors have types θ that are evenly distributed over $[0,1]$. We consider the symmetric equilibrium in which sophisticated investors choose the same entry strategy of $m_L(\theta)$, while unsophisticated investors choose the same $m_H(\theta)$. To simplify the problem, we assume that the unsophisticated cannot learn at all, i.e. $\mu_H \to \infty$, resulting in a constant m_H in θ. Otherwise the solution would be a set of two joint differential equations which cannot be easily solved.[9]

Proposition 10. *If $\mu_H \to \infty$, the optimal interior solution for $m_L(\theta)$ and m_H is given by the following system of equations. The optimal strategy m_L of the sophisticated is given by*

$$\omega \cdot \log\left(\frac{(1-\omega)\cdot m_H + \omega \cdot m_L(\theta)}{m_L(\theta)}\right) + (1-\omega)\cdot m_H \cdot \log\left(\frac{1-m_L(\theta)}{m_L(\theta)}\right) - \frac{M\cdot(\alpha+\beta)\cdot(1-\omega)\cdot m_H \cdot ((1-\omega)\cdot m_H + \omega)}{\mu_L}\cdot\theta =$$

$$\omega \cdot \log\left(\frac{(1-\omega)\cdot m_H + \omega \cdot m_L(0)}{m_L(0)}\right) + (1-\omega)\cdot m_H \cdot \log\left(\frac{1-m_L(0)}{m_L(0)}\right) \tag{36}$$

and m_H is pinned down by the indifference condition of the unsophisticated

$$1 - M\cdot(\beta-\alpha)\cdot(1-\omega)\cdot m_H + M\cdot\omega\cdot\int_0^1\left[\alpha\cdot\int_\theta^1 m_L(\tilde\theta)\mathrm{d}\tilde\theta - \beta\cdot\int_0^\theta m_L(\tilde\theta)\mathrm{d}\tilde\theta\right]\mathrm{d}\theta = 0 \tag{37}$$

where $m_L(0)$ is pinned down by the boundary condition (the FOC of the sophisticated)

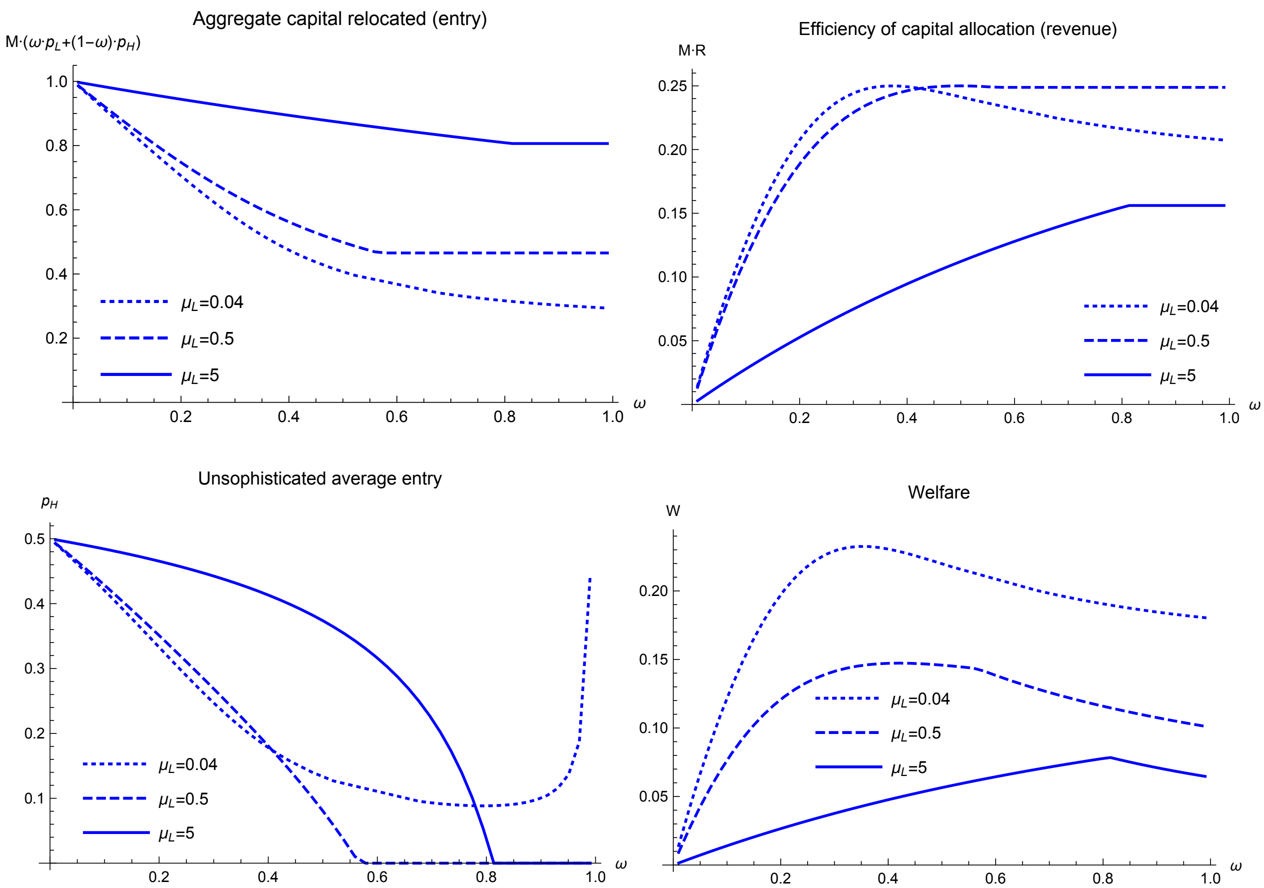
$$M\cdot\alpha\cdot[\omega\cdot p_L + (1-\omega)\cdot m_H] + 1 = \mu_L\cdot\left[\log\left(\frac{m_L(0)}{1-m_L(0)}\right) - \log\left(\frac{p_L}{1-p_L}\right)\right] \tag{38}$$

and $p_L = \int_0^1 m_L(\tilde\theta)\mathrm{d}\tilde\theta$ is the average entry of the sophisticated abritrageur.

We solve the above set of equations numerically since it is analytically intractable. In Figure 5 we vary the portion ω of sophisticated investors who can learn with cost μ_L. Thus the overall mass of investors M is kept constant but a growing fraction of investors is sophisticated. At $\omega = 0$ only unsophisticated are present and thus they enter until revenue is zero (given that M is large enough), yielding zero welfare. As ω initially increases, welfare increases since the average investor is

[9] See the proof of Proposition 10 for the full problem.

Figure 5: **Real outcomes with varying composition of investors**



Here we change the portion ω of sophisticated investors who can learn with low cost μ_L, while $1-\omega$ cannot learn. The mass of investors M is kept constant. Parameters: $\beta = 4$, $\alpha = 2$, while μ takes three different values: $\mu_L = 0.04$ (dotted line), $\mu_L = 0.5$ (dashed line), $\mu_L = 5$ (solid line). In all cases the social planner would allow each investor to enter with probability $\frac{1}{M\cdot(\beta-\alpha)} = \frac{1}{4}$, yielding total entry of $\frac{1}{\beta-\alpha} = \frac{1}{2}$.

more sophisticated. This is very similar to the result in the case of homogenous investors that welfare increases as the average sophistication of investors increases (i.e. as μ decreases). There are two effects leading to decreasing welfare as ω increases further. First, if the sophisticated are very sophisticated (low μ_L) then having lots of sophisticated leads to under-entry for $\alpha > 0$, thus decreasing welfare. This is like the case of homogenous investors where lowering μ is welfare reducing at low levels of μ since it aggravates under-entry. Second, and more interestingly, welfare can be decreasing in the share of sophisticated investors ω even for high μ_L in the absence of under-entry. The reason is that as ω increases, the unsophisticated are less likely to enter (m_H decreases) and above a threshold they are completely driven out of the market. Once the unsophisticated are not present, welfare is decreasing in ω. The intuition is similar to the baseline result in case of homogenous entrepreneurs that increasing the number of investors M, welfare eventually decreases as investors spend their revenue on learning.

Figure 5 also highlights the intricate interplay between the entry and learning strategies of the sophisticated and the unsophisticated. First, the remaining unsophisticated are less likely to enter as the fraction ω of sophisticated increases because there is more and more aggregate entry at low θ, cream-skimming the market and leaving less revenues for unsophisticated investors who enter indiscriminately. Second, unsophisticated investors are completely driven out of the market for high ω when sophisticated investors are also not perfectly sophisticated (if $\mu_L > 0$), thus they are competitors of the unsophisticated, cannibalizing their revenues and eventually driving them out. The intuition is similar to that in high frequency trading where some investors may stay out of the market because they are afraid of very fast investors front-running them.

In fact, if the sophisticated investors are sophisticated enough (μ_L close to zero), unsophisticated investors will never be completely driven out of the market, see Figure 5. The reason is that perfectly sophisticated investors follow cutoff strategies with the last entrant at the cutoff getting zero payoff and being indifferent (see $\mu_L = 0$ of the baseline model). If only sophisticated investors are present in the market, then an unsophisticated investor with a uniform prior about its θ knows it can get positive payoff if its θ is smaller than the cutoff of the perfectly sophisticated investors and gets zero payoff (equal to that of the last perfectly sophisticated to enter) with θ higher than the cutoff since there are no other entrants with higher θ in equilibrium. Figure 5 shows that in this case with ω close to one, all unsophisticated enter.

The above analysis also highlights how a not very well informed (unsophisticated) investor should behave if it learns about an arbitrage opportunity. It should enter with relatively high probability if it thinks investors in the market are predominantly sophisticated but only if it believes that the sophisticated investors are very sophisticated. On the other hand, it should not enter at all, if it thinks the other sophisticated investors are not extremely sophisticated. It may also choose to enter if it thinks that investors are predominantly unsophisticated.

5 Conclusions

We develop a model of capital reallocation to analyze whether the presence of more investors improve the efficiency and speed of capital allocation and welfare. Trades can become crowded due to imperfect information and externalities but investors can devote resources to learn about the number of earlier entrants. In general, more investors having the choice to enter into a trade neither improves the efficiency of capital allocation nor does it aggravate crowding. In fact, whether there is eventually too little or too much capital allocated to the new sector is determined solely by the technology in that sector, the cost of learning, the depth of the market, and the severity of the potential shocks, but not the mass of investors present. However, the presence of more investors decreases welfare, as they use more aggregate resources learning about each others' position. This excessive learning leads to inefficiently fast moving capital. Overall, our analysis cautions against using market efficiency or speed of capital allocation as goal in order to improve welfare.

References

Abreu, Dilip, and Markus K. Brunnermeier, 2003, Bubbles and crashes, *Econometrica* 71, 173–204.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2015, Equilibrium fast trading, *Journal of Financial Economics* 116, 292 – 313.

Budish, Eric, Peter Cramton, and John Shim, 2015, The high-frequency trading arms race: Frequent batch auctions as a market design response, *Quarterly Journal of Economics* 130, 1547–1621.

Duffie, Darrell, 2010, Asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1238–1268.

——— , and Bruno Strulovici, 2012, Capital mobility and asset pricing, *Econometrica* 80, 2469– 2509.

Glode, Vincent, Richard C. Green, and Richard Lowery, 2012, Financial expertise as an arms race, *Journal of Finance* 67, 1723–1759.

Greenwood, Robin, Samuel G. Hanson, and Gordon Y. Liao, 2015, Asset price dynamics in partially segmented markets, Harvard University Working Paper.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Hellwig, Christian, and Laura Veldkamp, 2009, Knowing what others know: Coordination motives in information acquisition, *Review of Economic Studies* 76, 223–251.

Hirshleifer, Jack, 1971, The private and social value of information and the reward to inventive activity, *American Economic Review* 61, 561–574.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Khandani, Amir E, and Andrew W Lo, 2011, What happened to the quants in august 2007? evidence from factors and transactions data, *Journal of Financial Markets* 14, 1–46.

Krueger, Anne O., 1974, The political economy of the rent-seeking society, *American Economic Review* 64, pp. 291–303.

Loury, Glenn C., 1979, Market structure and innovation, *Quarterly Journal of Economics* 93, 395–41.

MacKay, David J.C., 2003, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press).

Maćkowiak, Bartosz, and Mirko Wiederholt, 2009, Optimal sticky prices under rational inattention, *American Economic Review* 99, 769–803.

Matějka, Filip, and Alisdair McKay, 2015, Foundation for the multinomial logit model, *American Economic Review* 105, 272–98.

Moinas, Sophie, and Sebastien Pouget, 2013, The bubble game: An experimental study of speculation, *Econometrica* 81, 1507–1539 University of Toulouse working paper.

Oehmke, Martin, 2009, Gradual arbitrage, Columbia University, Working Paper.

Pedersen, Lasse, Mark Mitchell, and Todd Pulvino, 2007, Slow moving capital, *American Economic Review* 97, 215–220.

Sims, Christopher A., 1998, Stickiness, *Carnegie-Rochester Conference Series on Public Policy* 49, 317–356.

——— , 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–90.

Stein, Jeremy C., 2009, Presidential address: Sophisticated investors and market efficiency, *Journal of Finance* 64, 1517–1548.

Sun, Zheng, Ashley Wang, and Lu Zheng, 2012, The road less traveled: Strategy distinctiveness and hedge fund performance, *Review of Financial Studies* 25, 96–143.

Tullock, Gordon, 1967, The welfare costs of tariffs, monopolies, and theft, *Economic Inquiry* 5, 224–232.

Woodford, Michael, 2008, Inattention as a source of randomized discrete adjustment, New York University working paper.

Yang, Ming, 2015a, Coordination with rational inattention, *Journal of Economic Theory* 158, 721–738.

——— , 2015b, Optimality of debt under flexible information acquisition, Duke University working paper.

A Proofs

Proof of Lemma 1

Proof. Denote by $b(t)$ the mass of investors who chose to enter (i.e. engage in capital transport) before time t and $a(t)$ the mass of investors who enter after time t. Thus $k_{A,t} = k_{A,0} - b(t)$ and $k_{B,t} = b(t)$: there is more capital on island B if already $b(t)$ investors have decided to transport capital there from island A. Then, the revenue of an investor that chooses to transport capital at time t is

$$\underbrace{\Big[\gamma - \delta \cdot b(t)\Big]}_{\text{sell price}} - \underbrace{\Big[\gamma - \delta \cdot [k_{A,0} - b(t)]\Big]}_{\text{buy price}} = \delta \cdot k_{A,0} - 2\delta \cdot b(t)$$

Choosing $k_{A,0} = \frac{1}{\delta}$ yields α and β in the Lemma.

□

Proof of Lemma 2

Proof. The proof follows that of Lemma 1 in Woodford (2008). □

Proof of Lemma 3

Proof. For the private FOC we use a perturbation method similar to the proof in Yang (2015a). In the first order perturbation we set $m(\theta) + \nu \cdot \epsilon(\theta)$ as $m(\theta)$, while we keep the entry decision of the others $\tilde{m}$ fixed:

$$\int_0^1 ((m(\theta) + \nu \cdot \epsilon(\theta)) \cdot \Delta u(\tilde{m}, \theta) - \mu \cdot L(m(\theta) + \nu \cdot \epsilon(\theta)))\,\mathrm{d}\theta. \tag{39}$$

We then take derivative wrt ν and then set $\nu = 0$ yielding the FOC:

$$\int_0^1 \epsilon(\theta) \cdot \left(\Delta u(\tilde{m}, \theta) - \mu \cdot \left[\log\left(\frac{m(\theta)}{1 - m(\theta)}\right) - \log\left(\frac{\int_0^1 m(\tilde{\theta})d\tilde{\theta}}{1 - \int_0^1 m(\tilde{\theta})d\tilde{\theta}}\right)\right]\right) \mathrm{d}\theta = 0. \tag{40}$$

Since the original equation is an optimum, the above equality has to hold for any $\epsilon(\theta)$: thus the part multiplying $\epsilon(\theta)$ has to be zero for all θ. Setting $\tilde{m} = m$ we arrive at the symmetric solution we get (12).

For the social FOC we also use a perturbation method similar to the proof in Yang (2015a). In the first order perturbation we set $m_s(\theta) + \nu \cdot \epsilon(\theta)$ as $m_s(\theta)$, take derivative w.r.t. ν and then set $\nu = 0$ in order to arrive at the following equation that has to hold for any function $\epsilon(\theta)$:

$$\int_0^1 \epsilon(\theta) \cdot \left(M \cdot \alpha \cdot \int_\theta^1 m_s(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta m_s(\tilde{\theta}) d\tilde{\theta} - \mu \cdot \left[\log \left(\frac{m_s(\theta)}{1 - m_s(\theta)} \right) - \log \left(\frac{\int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}}{1 - \int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}} \right) \right] \right) \mathrm{d}\theta + \tag{41}$$

$$+ \int_0^1 m_s(\theta) \cdot \left(M \cdot \alpha \cdot \int_\theta^1 \epsilon(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \epsilon(\tilde{\theta}) d\tilde{\theta} \right) \mathrm{d}\theta = 0 \tag{42}$$

We choose $\epsilon(\theta) = \delta_{\hat{\theta}}(\theta)$ where $\delta_{\hat{\theta}}$ is the Dirac-Delta function. Thus $\int_\theta^1 \epsilon(\tilde{\theta}) d\tilde{\theta} = \mathbf{1}_{\theta < \hat{\theta}}$ where $\mathbf{1}$ is the heaviside function. Substituting $\hat{\theta} = \theta$, the equation becomes:

$$M \cdot \alpha \cdot \int_\theta^1 m_s(\tilde{\theta}) \mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta m_s(\tilde{\theta}) \mathrm{d}\tilde{\theta} + 1 - \mu \cdot \left[\log \left(\frac{m_s(\theta)}{1 - m_s(\theta)} \right) - \log \left(\frac{\int_0^1 m_s(\tilde{\theta}) \mathrm{d}\tilde{\theta}}{1 - \int_0^1 m_s(\tilde{\theta}) \mathrm{d}\tilde{\theta}} \right) \right] + \tag{43}$$

$$+ M \cdot \alpha \cdot \int_0^\theta m_s(\tilde{\theta}) \mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_\theta^1 m_s(\tilde{\theta}) \mathrm{d}\tilde{\theta} = 0 \tag{44}$$

which simplifies to (13). □

Proof of Proposition 1

Proof. Differentiating the FOC (12) we arrive at the following differential equation:

$$(M \cdot \alpha + M \cdot \beta) \cdot \tilde{m}(\theta) = - \frac{\mu \cdot m'(\theta)}{m(\theta) \cdot (1 - m(\theta))}. \tag{45}$$

thus the competitive equilibrium strategy $m(\theta)$ in the symmetric equilibrium ($m = \tilde{m}$) has to solve the above differential equation with the original FOC (e.g. evaluated at $\theta = 0$) as a boundary condition which is (15). If there is an interior solution (s.t. $m(\theta) \neq 1$), it can be written in the form

$$\frac{\frac{1}{m(\theta)} + \log \left(\frac{1 - m(\theta)}{m(\theta)} \right)}{M(\alpha + \beta)} = C + \frac{\theta}{\mu} \tag{46}$$

for an appropriate constant C. Setting $\theta = 0$ above and subtracting from the above we can eliminate C and thus arrive at (47).

$$\frac{1}{m(\theta)} + \log \left(\frac{1 - m(\theta)}{m(\theta)} \right) - \frac{M(\alpha + \beta)}{\mu} \cdot \theta = \frac{1}{m(0)} + \log \left(\frac{1 - m(0)}{m(0)} \right), \tag{47}$$

Taking logs and using the definition of the Lambert function (upper branch if $z > 0$) yields (14).

To calculate the level of $\bar{M}$ we use the observation (independently proven in Proposition 3) that $M \cdot p$ is constant, including in the limit as $M \to \infty$. At $\bar{M}$ still all investors enter with probability 1, thus $p = 1$ and $\bar{M}$ can be expressed as:

$$\bar{M} = \lim_{\mu \to \infty} (M \cdot p) \tag{48}$$

Thus we focus on expressing $M \cdot p$ in the limit for large M. As a first step note that as $M \to \infty$, given that $M \cdot p$ is constant, $m(\theta) \to 0$ for every θ. Thus the implicit equation (47) for $m(\theta)$ can be approximated by

$$\frac{1}{m} - M(\alpha + \beta)\left(C + \frac{\theta}{\mu}\right) = 0 \tag{49}$$

since for $m \approx 0$: $\frac{1}{m} \gg \log\left(\frac{1}{m}\right)$. A closed form solution can be obtained in this limit case:

$$m(\theta) = \frac{\mu}{M(\alpha + \beta)(C\mu + \theta)} \tag{50}$$

for a specific C. By the definition of the average entry p this implies

$$M \cdot p = M \cdot \int_0^1 m(\theta) \mathrm{d}\theta = \frac{\mu}{\alpha + \beta} \cdot \log\left(\frac{1}{C\mu} + 1\right). \tag{51}$$

Substituting this into the boundary condition (15) yields:

$$\alpha \frac{\mu}{\alpha + \beta} \log\left(\frac{1}{C\mu} + 1\right) + 1 = \mu \left[\log\left(\frac{1}{CM(\alpha + \beta) - 1}\right) - \log\left(\frac{M(\alpha + \beta) - \mu \log\left(\frac{1}{C\mu} + 1\right)}{\mu \log\left(\frac{1}{C\mu} + 1\right)}\right)\right] \tag{52}$$

Since $M \cdot p$ is a constant for any $M > \bar{M}$, C also has to converge to a finite constant as $M \to \infty$. Using this insight, one can take the limit of the above equation as $M \to \infty$:

$$\mu(-\alpha - \beta) \log\left(\frac{1}{C}\right) + (\alpha + \beta)\left(\mu \log\left(\mu \log\left(\frac{1}{C\mu} + 1\right)\right) + 1\right) + \alpha\mu \log\left(\frac{1}{C\mu} + 1\right) = 0 \tag{53}$$

Using the relation between C and $M \cdot p$ in (51), one can eliminate C:

$$(\alpha + \beta)\left(\mu \log(M \cdot p \cdot (\alpha + \beta)) - \mu \log\left(\mu\left(e^{\frac{M \cdot p \cdot (\alpha + \beta)}{\mu}} - 1\right)\right) + \alpha M \cdot p + 1\right) = 0 \tag{54}$$

using (48) and rearranging yields equation (16) in the proposition. □

Proof of Proposition 2

Proof. The derivative of FOC (13) w.r.t. θ delivers the differential equation

$$0 = -\frac{\mu \cdot m_s'(\theta)}{m_s(\theta) \cdot (1 - m_s(\theta))} \tag{55}$$

subject to the boundary condition (setting $\theta = 0$ in (13))

$$M \cdot (\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[\log\left(\frac{m_s(0)}{1 - m_s(0)} \right) - \log\left(\frac{p_s}{1 - p_s} \right) \right]. \tag{56}$$

This trivially yields

$$m_s(\theta) = C \tag{57}$$

for some constant C, implying $p_s = C$. The boundary condition (56) simplifies to

$$M \cdot (\alpha - \beta) p_s + 1 = 0 \tag{58}$$

implying (17). If the implied entry probability is > 1, then we have the corner solution that all enter with $m(\theta) = 1$. □

Proof of Lemma 4

Proof. Under complete information, in the competitive optimum the last one to enter $\bar{\theta}$ is indifferent between entering and not:

$$- M \cdot \beta \cdot \bar{\theta} + 1 = 0 \tag{59}$$

yielding Eq. 18. Because learning is free and only the aggregate amount of entrants matters from the social planner, we could choose many symmetric entry functions. For simplicity, let us choose the strategy in which all investors with $\theta < \bar{\theta}$ enter, the others stay out. $\bar{\theta}$ is given by maximizing:

$$\int_0^{\bar{\theta}} \big((M \cdot \alpha) \cdot (\bar{\theta} - \theta) - M \cdot \beta \cdot \theta + 1 \big) \, \mathrm{d}\theta = \frac{M \cdot \alpha - M \cdot \beta}{2} \cdot \bar{\theta}^2 + 1 \cdot \bar{\theta} \tag{60}$$

yielding the interior optimum in Eq. 19 if $M \cdot (\beta - \alpha) > 1$. If on the other hand, $M \cdot (\beta - \alpha) < 1$, everyone enters: $m(\theta) = 1$ is optimal.

Under no information, in the competitive equilibrium every investor enters with probability p and they are all indifferent given they do not know their θ and use a uniform prior. Expected payoff to entering:

$$\int_0^1 (M \cdot \alpha \cdot (1-\theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1)\, \mathrm{d}\theta = 0 \tag{61}$$

yielding the unconditional entry probability in Eq. 20. If M is low and the implied entry is > 1, then the revenue is not driven to zero and everyone enters for sure implying $p = 1$. In the social planner's optimum every investor enters with probability p and they maximize social planner's welfare

$$\int_0^1 p \cdot (M \cdot \alpha \cdot (1-\theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1)\, \mathrm{d}\theta \tag{62}$$

taking derivative w.r.t. p and setting to zero, this implies the entry probability in Eq. 21. As before, of the implied entry probability is > 1, then everyone enters for sure $m(\theta) = 1$ implying $p_s = 1$. Note that there are infinite other solutions since the social planner does not care about who exactly enters. □

Proof of Proposition 3

Proof. To show that $M \cdot p$ is constant in M once the solution m is interior, first write the system of 3 equations determining p. First, the difference of FOC (12) at $\theta = 0$ and $\theta = 1$.

$$p = \frac{\mu \left(\log\left(\frac{m(0)}{1-m(0)}\right) - \log\left(\frac{m(1)}{1-m(1)}\right)\right)}{M(\alpha+\beta)} \tag{63}$$

Second, the boundary condition (12) at $\theta = 0$

$$\alpha M p + 1 = \mu \left(\log\left(\frac{m(0)}{1-m(0)}\right) - \log\left(\frac{p}{1-p}\right)\right). \tag{64}$$

Third, the implicit equation for $m(\theta)$ evaluated at $\theta = 1$.

$$\log\left(\frac{m(0)}{1-m(0)}\right) - \log\left(\frac{m(1)}{1-m(1)}\right) = \frac{M(\alpha+\beta)}{\mu} + \frac{1}{m(0)} - \frac{1}{m(1)} \tag{65}$$

Substituting

$$x_0 = \log\left(\frac{m(0)}{1-m(0)}\right) \tag{66}$$

and

$$x_1 = \log\left(\frac{m(1)}{1-m(1)}\right) \tag{67}$$

the system of three equations can be written as:

$$p = \frac{\mu(x_0 - x_1)}{M(\alpha+\beta)} \tag{68}$$

$$\alpha M p + 1 = \mu\left(x_0 - \log\left(\frac{p}{1-p}\right)\right) \tag{69}$$

$$x_0 - x_1 = \frac{M(\alpha+\beta)}{\mu} + e^{-x_0} - e^{-x_1} \tag{70}$$

Substituting out p from (68), (69), (70) we arrive at a system of two equations:

$$F = \mu\left(x_0 - \log\left(\frac{\mu(x_0-x_1)}{M(\alpha+\beta)+\mu(x_1-x_0)}\right)\right) - \left(\frac{\alpha\mu(x_0-x_1)}{\alpha+\beta} + 1\right) = 0 \tag{71}$$

$$G = \frac{M(\alpha+\beta)}{\mu} - (x_0 - x_1) + e^{-x_0} - e^{-x_1} = 0 \tag{72}$$

To prove $M \cdot p$ is constant, it is sufficient to prove $\frac{\partial(M \cdot p)}{\partial M} = 0$ which from (68) is equivalent to

$$\frac{\partial x_0}{\partial M} = \frac{\partial x_1}{\partial M} \tag{73}$$

We apply Cramer's rule both for x_0 and x_1 to the system of equations (71) and (72):

$$\frac{\partial x_0}{\partial M} = \frac{\begin{vmatrix} \frac{\partial F}{\partial x_0} & -\frac{\partial F}{\partial M} \\ \frac{\partial G}{\partial x_0} & -\frac{\partial G}{\partial M} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} \\ \frac{\partial G}{\partial x_0} & \frac{\partial G}{\partial x_1} \end{vmatrix}} \tag{74}$$

$$\frac{\partial x_1}{\partial M} = \frac{\begin{vmatrix} -\frac{\partial F}{\partial M} & \frac{\partial F}{\partial x_1} \\ -\frac{\partial G}{\partial M} & \frac{\partial G}{\partial x_1} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} \\ \frac{\partial G}{\partial x_0} & \frac{\partial G}{\partial x_1} \end{vmatrix}} \tag{75}$$

We check numerically that the denominator (which is the same for both derivatives) is non-zero thus the two equations are indeed independent. It is thus sufficient to show that the numerators of the Cramer rule for the two derivatives are equal, yielding the sufficient condition

$$\frac{(\alpha+\beta)e^{-x_0-x_1}\left(e^{x_0+x_1}(M(\alpha+\beta)+\mu(x_1-x_0)) - \mu e^{x_0} + \mu e^{x_1}\right)}{M(\alpha+\beta)+\mu(x_1-x_0)} = 0. \tag{76}$$

It follows from (70) that the denominator is non-zero if $x_0 \neq x_1$. Thus it is sufficient to prove that

$$\frac{M(\alpha+\beta)}{\mu} + (x_1 - x_0) + \frac{1}{e^{x_0}} - \frac{1}{e^{x_1}} = 0, \tag{77}$$

which is exactly the function $G = 0$ defined in (72). Thus the identity holds and we have proved that $M \cdot p$ is constant in M for interior solutions.

□

Proof of Proposition 4

Proof. By Proposition 1 when $M < \bar{M}$ all investors enter with probability 1. Hence, all equilibrium objects are the same for the planner and in the decentralized solution. In particular, average entry of an investor is $p = 1$ thus expected aggregate entry is M. Total revenue and welfare are

$$M \cdot R = W = M \cdot R_s = W_s = M \cdot \int_0^1 (M \cdot \alpha \cdot (1-\theta) - M \cdot \beta \cdot \theta + 1)\, \mathrm{d}\theta = M - \frac{M^2(\beta-\alpha)}{2}. \tag{78}$$

To arrive at a formula for $W(M)$ one can rearrange the aggregate learning from (3) to get:

$$M \cdot L = M \int_0^1 m(\theta) \cdot \mu \cdot \left(\log\left(\frac{m(\theta)}{1-m(\theta)}\right) - \log\left(\frac{p}{1-p}\right)\right) \cdot \mathrm{d}\theta - M \int_0^1 \mu \log\left(\frac{1-p}{1-m(\theta)}\right) \cdot \mathrm{d}\theta \tag{79}$$

where the interior part of the first integral multiplying $m(\theta)$ can be replaced using the FOC (12) to yield:

$$M \cdot L = M \int_0^1 m(\theta) \cdot \left[M \cdot \alpha \cdot \int_\theta^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \tilde{m}(\tilde{\theta}) d\tilde{\theta} + 1 \right] \mathrm{d}\theta - M \int_0^1 \mu \log \left(\frac{1-p}{1-m(\theta)} \right) \cdot \mathrm{d}\theta \quad (80)$$

thus the first integral is exactly the definition of aggregate revenue. Since $M \cdot p$ is constant if $M \geq \bar{M}$ (Propostion 3), so is aggregate revenue $M \cdot R$. Rearranging yields the below expression for W: In general, one can write:

$$W(M) = M \cdot \int_0^1 \log \left(\frac{1-p}{1-m(\theta)} \right) \mathrm{d}\theta \quad (81)$$

We now show that welfare converges to zero for large M. For large M, $m \approx 0$ and $p \approx 0$ thus in the $M \to \infty$ limit (81) converges to zero. This convergence happens from above, since the payoff per investor $\frac{W}{M}$ cannot be negative, otherwise investors would choose not to enter.

In the social planner's interior optimum every investor enters with probability $p = \frac{1}{M \cdot (\beta - \alpha)}$ and thus welfare becomes

$$W_s = M \cdot \int_0^1 p \cdot (M \cdot \alpha \cdot (1-\theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1) \, \mathrm{d}\theta = \frac{1}{2 \cdot (\beta - \alpha)}. \quad (82)$$

□

Proof of Proposition 5

Proof. For flat entry function $m(\theta)$, the median entrant is exactly at $\theta = \frac{1}{2}$, implying $\tau = \frac{1}{2}$. For $M > \bar{M}$, the privately optimal entry function is downward sloping. This can be seen by observing that (45) implies that $m'(\theta)$ is always strictly negative. A decreasing $m(\theta)$ implies that the median entrant is smaller than $\frac{1}{2}$, thus $\tau < \frac{1}{2}$.

Finally, for $M \to \infty$ we use the result from the proof of Proposition 1 that in the limit $m(\theta)$ converges to (50) where C converges to a finite constant. Solving for τ using the approximation in the limit:

$$\int_0^\tau M \cdot m(\theta) d\theta = \int_0^\tau \frac{\mu}{(\alpha + \beta)(C\mu + \theta)} d\theta = \frac{\bar{M}}{2} \quad (83)$$

Evaluating the integral and using the relationship between C and $\bar{M}$ in (51) to substitute out C, one gets the expression for τ stated in the Lemma.

□

Proof of Lemma 5

Proof. With probability ν the investor is reverted to island A and sells at the average marginal product of capital: $\nu \cdot (a(t) + b(t))$ capital is sold by investors at $t = 1$ in random order. Note that on average half of the entrants sell before the investor in a fire sale. Thus the expected price is higher than if it was sold at a market clearing price in e.g. an auction. This assumption simplifies the analysis by not allowing workers to capture any of the surplus.

In case of a crisis, $(1-\nu)\cdot(a(t)+b(t))$ capital is sold on island B in a fire sale, in random order. Overall, the revenue of an investor that chooses to transport capital at time t is:

$$(1-\eta)\cdot(1-\nu)\cdot\underbrace{\left[\gamma-\delta\cdot(1-\nu)\cdot b(t)\right]}_{\text{sell price (no shock)}}+\eta\cdot(1-\nu)\cdot\underbrace{\left[\gamma-(\delta_c+\delta)\cdot(1-\nu)\cdot\frac{a(t)+b(t)}{2}\right]}_{\text{sell price (crisis)}}+$$

$$\nu\cdot\underbrace{\left[\gamma-\delta\cdot\left(k_{A,0}-[a(t)+b(t)]+\nu\cdot\frac{a(t)+b(t)}{2}\right)\right]}_{\text{sell price (idiosyncratic shock)}}-\underbrace{\left[\gamma-\delta\cdot[k_{A,0}-b(t)]\right]}_{\text{buy price}} \tag{84}$$

Choosing $k_{A,0} = \frac{1}{\delta\cdot(1-\nu)}$, the expected payoff of investor θ from transporting capital (given that investor θ can enter at time t) simplifies to (1) if α and β are given by the equations in the lemma.

Resulting in positive crowding and rat-race parameters of:

$$\beta-\alpha=(1-\nu)^2\cdot(\eta\cdot\delta_c+2\delta)>0 \tag{85}$$

$$\alpha+\beta=\left(1+(1-\nu)^2(1-\eta)\right)\cdot\delta>0 \tag{86}$$

for all parameter values. □

Proof of Proposition 6

Proof. Denote

$$A=e^{\frac{\bar{M}(\alpha+\beta)}{\mu}}-2e^{\frac{\alpha\bar{M}+1}{\mu}}+1 \tag{87}$$

and

$$B=\beta\left(e^{\frac{\alpha\bar{M}+1}{\mu}}-e^{\frac{\bar{M}(\alpha+\beta)}{\mu}}\right)+\alpha\left(e^{\frac{\alpha\bar{M}+1}{\mu}}-1\right). \tag{88}$$

To facilitate the proof we assume that $A > 0$, $B < 0$ and $\frac{B\cdot\bar{M}+e^{\frac{\bar{M}(\alpha+\beta)}{\mu}}-1}{B} > 0$. These are the conditions under which we state the Proposition. In fact we did not find any counterexamples to these restrictions given our assumptions about α and β.

Remember that $\bar{M}$ is defined by the implicit equation 16:

$$F = \frac{\bar{M}\cdot(\alpha+\beta)}{\mu} - e^{-\frac{1-\beta\cdot\bar{M}}{\mu}} + e^{-\frac{1+\alpha\cdot\bar{M}}{\mu}} = 0. \tag{89}$$

Using the result from Eq. 17 that $\bar{M}_s = \frac{1}{\beta-\alpha}$ and the implicit function theorem, the derivative of interest $\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial .}$ for any parameter "." becomes:

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial .} = \frac{\partial\bar{M}}{\partial .}\cdot(\beta-\alpha) + \frac{\partial(\beta-\alpha)}{\partial .}\cdot\bar{M} = -\frac{\frac{\partial F}{\partial .}}{\frac{\partial F}{\partial\bar{M}}}\cdot(\beta-\alpha) + \frac{\partial(\beta-\alpha)}{\partial .}\cdot\bar{M} \tag{90}$$

Basic algebra and the conditions stated at the beginning of the proof yield:

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\beta} = -\frac{A}{B}\cdot\alpha\cdot\bar{M} \tag{91}$$

the sign of which is the same as the sign of α, and

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\alpha} = \frac{A}{B}\cdot\beta\cdot\bar{M} < 0 \tag{92}$$

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\mu} = \frac{B\cdot\bar{M}+e^{\frac{\bar{M}(\alpha+\beta)}{\mu}}-1}{B}\cdot\frac{\beta-\alpha}{\mu} > 0 \tag{93}$$

The last expression proves the first part of the Proposition. For the other parts, we use the total derivative to get the effect of the parameters of our full model:

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\delta} = \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\beta}\cdot\frac{\partial\beta}{\partial\delta} + \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\alpha}\cdot\frac{\partial\alpha}{\partial\delta} = -\left(-\frac{A}{B}\cdot\bar{M}\right)\cdot\frac{1}{2}\delta_c\cdot\eta\cdot(1-\nu)^2\cdot\big((1-\eta)(1-\nu)^2+1\big) \leq 0 \tag{94}$$

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\delta_c} = \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\beta}\cdot\frac{\partial\beta}{\partial\delta_c} + \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\alpha}\cdot\frac{\partial\alpha}{\partial\delta_c} = \left(-\frac{A}{B}\cdot\bar{M}\right)\cdot\frac{1}{2}\delta\cdot\eta\cdot(1-\nu)^2\cdot\big((1-\eta)(1-\nu)^2+1\big) \geq 0 \tag{95}$$

with equality if and only if $\eta = 0$. Furthermore,

$$\frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \eta} = \frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \eta} + \frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \eta} = \left(-\frac{A}{B} \cdot \bar{M}\right) \cdot \frac{1}{2}\delta \cdot (1-\nu)^2 \cdot \big((1-\nu)^2(2\delta + \delta_c) + \delta_c\big) > 0 \tag{96}$$

$$\frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \nu} = \frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \nu} + \frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \nu} = -\left(-\frac{A}{B} \cdot \bar{M}\right) \cdot \delta \cdot (1-\nu) \cdot (2\delta + \delta_c \eta) < 0 \tag{97}$$

where we used $\delta > 0$, $A > 0$, $B < 0$ and $\bar{M} > 0$.

□

Proof of Proposition 7

Proof. Again, the condition under which we prove the Proposition is $B < 0$, where B is defined by (88).

Remember that $\bar{M}$ is defined by the implicit equation (89) in an implicit form as $F(\bar{M}) = 0$. Denote $\lim_{M\to\infty} \tau = \bar{\tau}$. Using the result that $\bar{\tau} = \frac{1}{e^{\frac{\bar{M}\cdot(\alpha+\beta)}{2\mu}}+1}$ and the implicit function theorem, the derivative of interest $\frac{d\bar{\tau}}{d.}$ for any parameter "." becomes:

$$\frac{d\bar{\tau}}{d.} = \frac{\partial \bar{\tau}}{\partial \bar{M}} \cdot \frac{\partial \bar{M}}{\partial .} + \frac{\partial \bar{\tau}}{\partial .} = \frac{\partial \bar{\tau}}{\partial \bar{M}} \cdot \left(-\frac{\frac{\partial F}{\partial .}}{\frac{\partial F}{\partial \bar{M}}}\right) + \frac{\partial \bar{\tau}}{\partial .} \tag{98}$$

Basic algebra yields:

$$\frac{\partial \bar{\tau}}{\partial \beta} = \frac{\alpha \bar{M} C}{2(-B)\mu} \tag{99}$$

the sign of which is the same as the sign of α.

$$\frac{\partial \bar{\tau}}{\partial \alpha} = -\frac{\beta \bar{M} C}{2(-B)\mu} < 0 \tag{100}$$

$$\frac{\partial \bar{\tau}}{\partial \mu} = \frac{(\alpha+\beta)C}{2(-B)\mu^2} > 0 \tag{101}$$

where C is defined by:

$$C = \frac{e^{\frac{\bar{M}(\alpha+\beta)}{2\mu}}}{\left(e^{\frac{\bar{M}(\alpha+\beta)}{2\mu}}+1\right)^2} \cdot \left(e^{\frac{\bar{M}(\alpha+\beta)}{\mu}} - 1\right). \tag{102}$$

Where we have assumed that $B < 0$. Also, $\frac{\bar{M}(\alpha+\beta)}{\mu} > 0$ implies $e^{\frac{\bar{M}(\alpha+\beta)}{\mu}} > 1$ and thus $C > 0$. This yields the above signs. The last expression proves the first part of the Proposition. For the other parts, we use the total derivative to get the effect of the parameters of our full model:

$$\frac{\partial \bar{\tau}}{\partial \delta} = \frac{\partial \bar{\tau}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \delta} + \frac{\partial \bar{\tau}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \delta} = \frac{C\delta_c \eta (1-\nu)^2 \bar{M}}{4B\mu} \left((1-\eta)(1-\nu)^2 + 1 \right) < 0 \tag{103}$$

$$\frac{\partial \bar{\tau}}{\partial \delta_c} = \frac{\partial \bar{\tau}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \delta_c} + \frac{\partial \bar{\tau}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \delta_c} = -\frac{C\eta (1-\nu)^2 \bar{M}}{4B\mu} (\alpha + \beta) > 0 \tag{104}$$

Furthermore,

$$\frac{\partial \bar{\tau}}{\partial \eta} = \frac{\partial \bar{\tau}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \eta} + \frac{\partial \bar{\tau}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \eta} = -\frac{C(1-\nu)^2 \bar{M}}{4B\mu} ((\beta - \alpha) \cdot \delta + (\beta + \alpha) \cdot \delta_c) > 0 \tag{105}$$

$$\frac{\partial \bar{\tau}}{\partial \nu} = \frac{\partial \bar{\tau}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \nu} + \frac{\partial \bar{\tau}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \nu} = \frac{C\delta (1-\nu) \bar{M} (2\delta + \delta_c \eta)}{2B\mu} < 0 \tag{106}$$

where we used $B < 0$, $C > 0$, $\bar{M} > 0$ and the parametric assumptions.

□

Proof of Proposition 8

Proof. Denote the strategy function of all other players as $\tilde{m}(\theta)$. Following the same steps as in the proof of Lemma 3 we arrive at the FOC:

$$1 - \kappa \cdot \theta + \alpha \cdot \int_{\theta}^{1} \tilde{m}(\tilde{\theta}) d\tilde{\theta} - \beta \cdot \int_{0}^{\theta} \tilde{m}(\tilde{\theta}) d\tilde{\theta} = \mu \cdot \left[\log \left(\frac{m(\theta)}{1 - m(\theta)} \right) - \log \left(\frac{p}{1-p} \right) \right]. \tag{107}$$

Differentiating this we arrive at the differential equation:

$$(\alpha + \beta) \cdot \tilde{m}(\theta) + \kappa = -\mu \cdot \frac{m'(\theta)}{m(\theta) \cdot (1 - m(\theta))}. \tag{108}$$

Imposing symmetry $\tilde{m}(\theta) = m(\theta)$ results in Equation 29. The boundary condition is given by the original integral-differential Equation 107 evaluated at any θ: in Equation 30 we set $\theta = 0$.

For the social planner's problem again we follow the steps of Lemma 3 which we reiterate here. The social planner chooses the symmetric choice function $m_s(\theta)$ to maximize

$$\int_{0}^{1} m_s(\theta) \cdot \Delta u(\theta, m_s) d\theta - \mu \cdot I(m_s) \tag{109}$$

where it takes into account that Δu depends not only on θ but on the information choice function of all other investors m.

We use a perturbation method similar to the proof in Yang (2015a). In the first order perturbation we set $m_s(\theta)+\nu\cdot\epsilon(\theta)$ as $m_s(\theta)$, take derivative wrt ν and then set $\nu=0$ in order to arrive at the following equation that has to hold for any function $\epsilon(\theta)$:

$$\int_0^1 \epsilon(\theta)\cdot\left(\alpha\cdot\int_\theta^1 m_s(\tilde{\theta})d\tilde{\theta}-\beta\cdot\int_0^\theta m_s(\tilde{\theta})d\tilde{\theta}-\kappa-\mu\cdot\left[\log\left(\frac{m_s(\theta)}{1-m_s(\theta)}\right)-\log\left(\frac{\int_0^1 m_s(\tilde{\theta})d\tilde{\theta}}{1-\int_0^1 m_s(\tilde{\theta})d\tilde{\theta}}\right)\right]\right)d\theta+$$

$$+\int_0^1 m_s(\theta)\cdot\left(\alpha\cdot\int_\theta^1 \epsilon(\tilde{\theta})d\tilde{\theta}-\beta\cdot\int_0^\theta \epsilon(\tilde{\theta})d\tilde{\theta}\right)d\theta=0 \tag{110}$$

We choose $\epsilon(\theta)=\delta_{\hat{\theta}}(\theta)$ where $\delta_{\hat{\theta}}$ is the Dirac-Delta function. Thus $\int_\theta^1 \epsilon(\tilde{\theta})d\tilde{\theta}=\mathbf{1}_{\theta<\hat{\theta}}$ where $\mathbf{1}$ is the heaviside function. Substituting $\hat{\theta}=\theta$, the equation becomes:

$$\alpha\cdot\int_\theta^1 m_s(\tilde{\theta})d\tilde{\theta}-\beta\cdot\int_0^\theta m_s(\tilde{\theta})d\tilde{\theta}+1-\kappa\cdot\theta-\mu\cdot\left[\log\left(\frac{m_s(\theta)}{1-m_s(\theta)}\right)-\log\left(\frac{\int_0^1 m_s(\tilde{\theta})d\tilde{\theta}}{1-\int_0^1 m_s(\tilde{\theta})d\tilde{\theta}}\right)\right]+$$

$$+\alpha\cdot\int_0^\theta m_s(\tilde{\theta})d\tilde{\theta}-\beta\cdot\int_\theta^1 m_s(\tilde{\theta})d\tilde{\theta}=0 \tag{111}$$

which simplifies to:

$$(\alpha-\beta)\cdot p_s+1-\kappa\cdot\theta-\mu\cdot\left[\log\left(\frac{m(\theta)}{1-m(\theta)}\right)-\log\left(\frac{p_s}{1-p_s}\right)\right]=0 \tag{112}$$

The derivative of Equation 112 w.r.t. θ delivers Equation 31, while setting $\theta=0$ in Equation 112 gives the boundary condition Equation 32. □

Proof of Proposition 9

Proof. First, note that the expected payoff is monotonously decreasing in either η_i or θ_i. Also, if $s_i>s_j$ then $g(\eta_i|s_i;\sigma_\eta,\sigma_{\varepsilon_i})$ first order stochastically dominates $g(\eta_j|s_j;\sigma_\eta,\sigma_{\varepsilon_i})$ where $g(\cdot|\cdot)$ is the distribution of η_i conditional on the signal s_i. Note that a lower signal means a higher expected payoff. Thus the unique optimal strategy of investor i is to enter based on a cutoff $\bar{s}_i$, entering whenever $s_i<\bar{s}_i$. Thus, conjecturing that the choice of σ_{ε_i} and $\bar{s}_i$ are symmetric and dropping the subscripts, each agent solves

$$\max_{\sigma_\varepsilon, \bar{s}} \int_{-\infty}^{\infty} \left[1 + \alpha M \int_{\eta}^{\infty} f\left(s < \tilde{\bar{s}} | \eta'; \tilde{\sigma}_\varepsilon, \sigma_\eta\right) \phi\left(\eta'; \sigma_\eta\right) d\eta' - \beta M \int_{-\infty}^{\eta} f\left(s < \tilde{\bar{s}} | \eta'; \tilde{\sigma}_\varepsilon, \sigma_\eta\right) \phi\left(\eta'; \sigma_\eta\right) d\eta'\right] \cdot f\left(s < \bar{s} | \eta; \sigma_\varepsilon, \sigma_\eta\right) \phi\left(\eta; \sigma_\eta\right) d\eta - C\left(\sigma_\varepsilon\right) \quad (113)$$

where the tilde denotes the choice of others, and $f\left(s < \bar{s} | \eta; \sigma_\varepsilon, \sigma_\eta\right)$ is the conditional probability of $s < \bar{s}$ given η with choice σ_ε.

Note that we can rewrite the expected revenue in (113) as follows

$$\int_{-\infty}^{\infty} \left(1 + \alpha M \int_{\eta}^{\infty} f\left(s < \tilde{\bar{s}} | \eta'; \tilde{\sigma}_\varepsilon, \sigma_\eta\right) \phi\left(\eta'; \sigma_\eta\right) d\eta' - \beta M \int_{-\infty}^{\eta} f\left(s < \tilde{\bar{s}} | \eta'; \tilde{\sigma}_\varepsilon, \sigma_\eta\right) \phi\left(\eta'; \sigma_\eta\right) d\eta'\right) \cdot f\left(s < \bar{s} | \eta; \sigma_\varepsilon, \sigma_\eta\right) \phi\left(\eta; \sigma_\eta\right) d\eta =$$

$$= \int_{-\infty}^{\infty} \left[1 + \alpha M \int_{\eta}^{\infty} \Phi\left(\tilde{\bar{s}} - \eta'; \tilde{\sigma}_\varepsilon\right) \phi\left(\eta'; \sigma_\eta\right) d\eta' - \beta M \int_{-\infty}^{\eta} \Phi\left(\tilde{\bar{s}} - \eta'; \tilde{\sigma}_\varepsilon\right) \phi\left(\eta'; \sigma_\eta\right) d\eta'\right] \cdot \Phi\left(\bar{s} - \eta; \sigma_\varepsilon\right) \phi\left(\eta; \sigma_\eta\right) d\eta =$$

$$= \int_{0}^{1} \left[1 + \alpha M \underbrace{\int_{\theta}^{1} \Phi\left(\tilde{\bar{s}} - \Phi^{-1}\left(\theta'; \sigma_\eta\right); \tilde{\sigma}_\varepsilon\right) d\theta'}_{a(\theta)} - \beta M \underbrace{\int_{0}^{\theta} \Phi\left(\tilde{\bar{s}} - \Phi^{-1}\left(\theta'; \sigma_\eta\right); \tilde{\sigma}_\varepsilon\right) d\theta'}_{b(\theta)}\right] \cdot \Phi\left(\bar{s} - \Phi^{-1}\left(\theta; \sigma_\eta\right); \sigma_\varepsilon\right) d\theta. \quad (114)$$

In the first equation we used that $f\left(s < \bar{s} | \eta; \sigma_\varepsilon, \sigma_\eta\right) = \Phi\left(\bar{s} - \eta; \sigma_\varepsilon\right)$ based on our assumptions, while in the second equation we used the rule of integration by substitution to replace η with θ.

Note that the last equation has the same form as the expected revenue in our baseline model with the restriction that in the Gaussian problem the entry function $m\left(\theta\right)$ is restricted to have the form of

$$m_G\left(\theta\right) = \Phi\left(\bar{s} - \Phi^{-1}\left(\theta; \sigma_\eta\right); \sigma_\varepsilon\right), \quad (115)$$

The probability of unconditional entry is $p = \Phi\left(\bar{s}; \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}\right)$ because the standard deviation of the signal is $\sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}$ due to the independence of η_i and ε_i.

The solution can be obtained by setting marginal cost and benefit of both parameters σ_ε and $\bar{s}$ equal while keeping the others' choice constant and then imposing symmetry, such that $\tilde{\sigma}_\varepsilon = \sigma_\varepsilon$ and $\tilde{\bar{s}} = \bar{s}$. For both cost functions, we numerically search for the solution of the two first order conditions.

First, consider the partial cost of learning, where the function $C_p\left(\sigma_{\varepsilon_i}\right)$ is defined by:

$$C_p(\sigma_{\varepsilon_i}) = \mu \cdot \left[\left(p \log\left[\frac{1}{p}\right] + (1-p)\log\left[\frac{1}{1-p}\right]\right) - \int_0^1 \left(m_G(\theta)\log\left[\frac{1}{m_G(\theta)}\right] + (1-m_G(\theta))\log\left[\frac{1}{1-m_G(\theta))}\right]\right) \mathrm{d}\theta\right] \tag{116}$$

This implies the two first order conditions:

$$\frac{\partial C_p}{\partial \bar{s}} = \mu \cdot \left[\log\left(\frac{1-p}{p}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2+\sigma_\varepsilon^2)}}}{\sqrt{2\pi}\sqrt{\sigma_\eta^2+\sigma_\varepsilon^2}} - \int_0^1 \log\left(\frac{1-m_G(\theta)}{m_G(\theta)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta\right] = \int_0^1 \left(1+\alpha\cdot M\cdot a\left(\theta\right) - \beta\cdot M\cdot b\left(\theta\right)\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta = \frac{\partial R}{\partial \bar{s}} \tag{117}$$

$$\frac{\partial C_p}{\partial \sigma_\varepsilon} = \mu \cdot \left[-\log\left(\frac{1-p}{p}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2+\sigma_\varepsilon^2)}}\cdot\sigma_\varepsilon}{\sqrt{2\pi}(\sigma_\eta^2+\sigma_\varepsilon^2)^{\frac{3}{2}}} + \int_0^1 \log\left(\frac{1-m_G(\theta)}{m_G(\theta)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}\cdot\left(\bar{s}-\Phi^{-1}\left(\theta\right)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2}d\theta\right] = -\int_0^1 \left(1+\alpha\cdot M\cdot a\left(\theta\right) - \beta\cdot M\cdot b\left(\theta\right)\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}\cdot\left(\bar{s}-\Phi^{-1}\left(\theta\right)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2}d\theta = \frac{\partial R}{\partial \sigma_\varepsilon} \tag{118}$$

That is one has to find σ_ε and $\bar{s}$ that jointly solves:

$$\mu \cdot \left[\log\left(\frac{1-\Phi\left(\bar{s};\sqrt{\sigma_\eta^2+\sigma_\varepsilon^2}\right)}{\Phi\left(\bar{s};\sqrt{\sigma_\eta^2+\sigma_\varepsilon^2}\right)}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2+\sigma_\varepsilon^2)}}}{\sqrt{2\pi}\sqrt{\sigma_\eta^2+\sigma_\varepsilon^2}} - \int_0^1 \log\left(\frac{1-\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\eta\right);\sigma_\varepsilon\right)}{\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\eta\right);\sigma_\varepsilon\right)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta\right] = \int_0^1 \left(1+\alpha\cdot M\cdot\int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\eta\right);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot\int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\eta\right);\sigma_\varepsilon\right)d\theta'\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta \tag{119}$$

$$\mu \cdot \left[-\log\left(\frac{1-\Phi\left(\bar{s};\sqrt{\sigma_\eta^2+\sigma_\varepsilon^2}\right)}{\Phi\left(\bar{s};\sqrt{\sigma_\eta^2+\sigma_\varepsilon^2}\right)}\right) \frac{e^{-\frac{\bar{s}^2}{2(\sigma_\eta^2+\sigma_\varepsilon^2)}} \cdot \sigma_\varepsilon}{\sqrt{2\pi}(\sigma_\eta^2+\sigma_\varepsilon^2)^{\frac{3}{2}}} + \int_0^1 \log\left(\frac{1-\Phi\left(\bar{s}-\Phi^{-1}(\theta;\sigma_\eta);\sigma_\varepsilon\right)}{\Phi\left(\bar{s}-\Phi^{-1}(\theta;\sigma_\eta);\sigma_\varepsilon\right)}\right) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}} \cdot \left(\bar{s}-\Phi^{-1}(\theta)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2} d\theta \right] =$$

$$-\int_0^1 \left(1+\alpha\cdot M\cdot \int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}(\theta';\sigma_\eta);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot\int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}(\theta';\sigma_\eta);\sigma_\varepsilon\right)d\theta'\right) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}} \cdot \left(\bar{s}-\Phi^{-1}(\theta)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2} d\theta \tag{120}$$

Second, consider the full cost of learning, where the function $C_f(\sigma_{\varepsilon_i})$ is defined by (33). This implies the two first order conditions:

$$\frac{\partial C_f}{\partial \bar{s}} = 0 = \int_0^1 (1+\alpha\cdot M\cdot a(\theta) - \beta\cdot M\cdot b(\theta)) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon} d\theta = \frac{\partial R}{\partial \bar{s}} \tag{121}$$

$$\frac{\partial C_f}{\partial \sigma_\varepsilon} = -\mu\frac{\sigma_\eta^2}{\sigma_\varepsilon^3+\sigma_\varepsilon\cdot\sigma_\eta^2} = -\int_0^1 (1+\alpha\cdot M\cdot a(\theta) - \beta\cdot M\cdot b(\theta)) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}} \cdot \left(\bar{s}-\Phi^{-1}(\theta)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2} d\theta = \frac{\partial R}{\partial \sigma_\varepsilon} \tag{122}$$

That is one has to find σ_ε and $\bar{s}$ that jointly solves:

$$0 = \int_0^1 \left(1+\alpha\cdot M\cdot \int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}(\theta';\sigma_\eta);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot\int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}(\theta';\sigma_\eta);\sigma_\varepsilon\right)d\theta'\right) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon} d\theta \tag{123}$$

$$-\mu\frac{\sigma_\eta^2}{\sigma_\varepsilon^3+\sigma_\varepsilon\cdot\sigma_\eta^2} =$$

$$-\int_0^1 \left(1+\alpha\cdot M\cdot \int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}(\theta';\sigma_\eta);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot\int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}(\theta';\sigma_\eta);\sigma_\varepsilon\right)d\theta'\right) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}} \cdot \left(\bar{s}-\Phi^{-1}(\theta)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2} d\theta \tag{124}$$

Numerically one observes that if M is large enough, then in the symmetric equilibrium, welfare would go negative thus the payoff of all investors would be negative. On Figure 4 this means that the smooth continuation of the welfare curve calculated for low M would cross below zero for higher M. Clearly investors would prefer to stay out without learning and thus get zero if they would get negative payoff when entering. Denote the crossing point $\hat{M}$ at which $W = 0$ in the symmetric equilibrium. It is an equilibrium for $M > \hat{M}$ that $\hat{M}$ investors enter and the rest $(M - \hat{M})$ stay out. In this case the incentives and payoffs among entrants is exactly the same as it would be if the mass of investors was $\hat{M}$, thus they follow the same strategies as in that case and get zero payoffs. Since investors who stay out without learning also get zero payoff, they are ex ante indifferent between entering (with learning) and staying out (without learning). Thus this is an equilibrium, though it might or might not be unique. Thus for all $M > \hat{M}$ all aggregate quantities are the same as when there are only $\hat{M}$ investors. □

Proof of Proposition 10

Proof. We first set up the problem for general μ_d before setting the special case of $\mu_d \to \infty$. In equilibrium, the mass of lower types entering ("before" investor θ) becomes:

$$b(\theta) = M \cdot \int_0^\theta \omega \cdot m_c(\tilde{\theta}) + (1-\omega) \cdot m_d(\tilde{\theta}) \mathrm{d}\tilde{\theta} \tag{125}$$

the mass of higher types entering ("after" investor θ):

$$a(\theta) = M \cdot \int_\theta^1 \omega \cdot m_c(\tilde{\theta}) + (1-\omega) \cdot m_d(\tilde{\theta}) \mathrm{d}\tilde{\theta}. \tag{126}$$

Thus the problem the investors solve is the joint maximization of two equations $i \in [d, c]$:

$$\max_{m_i(\theta)} \int_0^1 \left(m_i(\theta) \cdot \Delta u(\theta) - \mu_i \cdot L(m_i)\right) \mathrm{d}\theta. \tag{127}$$

The optimal solution is characterized by the differential equation for m_c

$$(\alpha + \beta)\left(\omega\, \tilde{m}_c(\theta) + (1-\omega)\, \tilde{m}_d(\theta)\right) = -\frac{\mu_c\, m_c'(\theta)}{m_c(\theta)\left(1 - m_c(\theta)\right)} \tag{128}$$

with the boundary condition:

$$M \cdot \alpha \cdot \int_\theta^1 \left(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega) \cdot \tilde{m}_d(\tilde{\theta})\right) \mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \left(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega) \cdot \tilde{m}_d(\tilde{\theta})\right) \mathrm{d}\tilde{\theta} + 1 = \mu_c \cdot \left[\log\left(\frac{m_c(\theta)}{1 - m_c(\theta)}\right) - \log\left(\frac{p_c}{1 - p_c}\right)\right]. \tag{129}$$

Where $p_s = \int_0^1 m_c(\tilde{\theta}) \mathrm{d}\tilde{\theta}$ is the average entry of the sophisticated. A symmetric set of equations hold for m_d which we omit for brevity. In equilibrium $\tilde{m}_d = m_d$ and $\tilde{m}_c = m_c$. In general such systems of interlinked differential equations for $m_d(\theta)$ and $m_c(\theta)$ cannot be solved, thus we set $\mu_d \to \infty$. This means that $m_d(\theta) \equiv m_d$

is a constant chosen such that the revenue of the dumb is exactly zero for interior solutions.

$$M \cdot \alpha \cdot \int_{\theta}^{1} \Big(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega) \cdot \tilde{m}_d(\tilde{\theta})\Big)\, \mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_{0}^{\theta} \Big(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega) \cdot \tilde{m}_d(\tilde{\theta})\Big)\, \mathrm{d}\tilde{\theta} + 1 = 0 \qquad (130)$$

$m_d = 0$ is chosen if the left hand side of the above equation is negative in such an equilibrium and $m_d = 1$ is chosen if it is positive. Equation 129 can be solved in implicit form and yields Equation in the proposition (evaluated at $\theta = 0$ to substitute out the constant). Equation 129 evaluated at $\theta = 0$ yields the equation for $m_c(0)$ given in Equation 38 in the Proposition. Equation 130 can be simplified to yield 37 in the Proposition. $\square$

B Online Appendix: More general microfoundations

B.1 Consumer/producer surplus

We show that our general model allows for a wide range of other applications differing in the source of the externalities and their welfare implications. For example, firms in industries with knowledge spillovers benefit from others following their location choices. Academics benefit from entering into fields early which later become very popular and thus increasing the number of their citations. Welfare consequences of crowding can be very different across these applications. The main reason is that, contrary to our baseline microfoundation, in most other applications, there is a group of investors whose welfare increases when others enter too much. For example, the over-entry of firms in an industry with knowledge spillover might decrease consumer prices in that industry below marginal cost. This harms firms but helps consumers. In the following, we show three examples where passive investors (consumers or producers) benefit from entry and thus the socially optimal level of entry changes. We then discuss the welfare implications in Section B.4.

The common theme in the models that follow is the presence of passive agents with a payoff (value) of

$$V^P \equiv \alpha_w \cdot \int_0^1 m(\theta) \cdot a(\theta) \cdot \mathrm{d}\theta = \alpha_w \cdot \frac{(M \cdot p)^2}{2} \tag{131}$$

with some parameter $\alpha_w \geq 0$. Thus the overall welfare in the whole economy can be computed as the sum of the payoff of active agents (investors in our baseline) and the payoff of passive investors:

$$W \equiv M \cdot V + V^P \tag{132}$$

One can interpret all our results up to now as a special case setting $\alpha_w = 0$. Thus the relevant payoff of entry from the social planner's perspective becomes:

$$\Delta u = (\alpha + \alpha_w) \cdot a(\theta) - \beta \cdot b(\theta) + 1 \tag{133}$$

Note that the investors' competitive payoff function is unchanged, i.e. it does not include α_w.

B.2 Product market with scarce resources

Consider a single good traded in a competitive market produced by heterogenous producers that compete for scarce resources. Producing the good is also subject to local spill-overs. In this example the producers are the active agents making the entry decision while the consumers are passive.

The goods are purchased by a representative consumer with quadratic utility:

$$U(q) = q - \frac{\alpha_w}{2} q^2 \tag{134}$$

such that $MU = p$ yields linear demand for the good.

$$\text{price}(q) = 1 - \alpha_w \cdot q \tag{135}$$

A unit mass of firms indexed by θ decides to move to a specific area (e.g. the new Silicon Valley) where other firms might also move. Producers are heterogenous in their θ: lower θ producers have lower costs both because they have better technology and can also secure a better (heterogenous) input. The cost of building the plant is subject to weakly increasing marginal cost (and price of the building):

$$\text{cost}(\theta) = (\alpha + \beta)b(\theta) - (\alpha + \alpha_w)(a(\theta) + b(\theta)) \tag{136}$$

The $(\alpha + \beta)$ term multiplying $b(\theta)$ of the cost function comes from the price of the production input (e.g. land) which increases as more (and worse) type producers enter. Better types can choose a better quality input before the others at a fixed price (given by the value of external use). The $\alpha + \alpha_w$ term multiplying $(a(\theta) + b(\theta))$ captures network externalities that some resources might become cheaper if many producers use it because of economies of scale.

The payoff of firm θ conditional on moving is:

$$\Delta u = \text{price}(q) - \text{cost}(\theta) = 1 - \alpha_w \cdot q - (\alpha + \beta)b(\theta) + (\alpha + \alpha_w)(a(\theta) + b(\theta)) \tag{137}$$

Where in equilibrium the amount of goods produced depends on how many firm decide to produce: $q = a(\theta) + b(\theta)$ yielding the payoff function of Equation 1 in the reduced form.

The payoff of the passive agent in this example is the consumer surplus which can be computed as:

$$\text{CS} = \int_0^q (1 - \alpha_w \cdot \hat{q} - \text{price}(q))d\hat{q} = \int_0^q \alpha_w \cdot (q - \hat{q})d\hat{q} = \alpha_w \cdot \frac{q^2}{2} \tag{138}$$

A different way to compute the consumer surplus is to add a term of $\alpha_w \cdot a(\theta)$ (alternatively $\alpha_w \cdot b(\theta)$) to Δu

$$\text{CS} = \int_0^1 m(\theta) \cdot [\alpha_w \cdot a(\theta)] \, \mathrm{d}\theta = -\frac{\alpha_w}{2} a(\theta)^2 \Big|_0^1 = \frac{\alpha_w}{2} a(1)^2 = \alpha_w \cdot \frac{q^2}{2} \tag{139}$$

This means that the objective function for the social planner maximizing welfare is the same as (1), substituting α with $\alpha + \alpha_w$, yielding (133).

B.3 Academic publications

Consider an academic tournament: e.g. the strategic choice of field of an aspiring academic. The academic wants to choose a topic that has not yet been done and that will have many followers who cite him. Both reading through and understanding the previous literature is time consuming and also trying to figure out whether others will find the same topic interesting and cite you. The academic's payoff is the probability of publishing and being cited. θ can be interpreted as time. If lots of researchers finish their paper before (or other old but similar papers are discovered), the lower your chance of publication: this is captured by β. If lots of researchers write a paper on the same topic afterwards that increases citation and the academic's chance of publication, captured by α. Thus the payoff to the academic (producer of knowledge) is:

$$\Delta u = 1 - \beta \cdot b(\theta) + \alpha \cdot a(\theta) \tag{140}$$

where we assumed that the fixed payoff already incorporates the fixed cost of producing knowledge. This yields Equation 1 in the reduced form model. Adding a consumer of knowledge with quadratic utility:

$$U(q) = q - \frac{\alpha_w}{2} q^2 \tag{141}$$

yields a similar additional term in welfare as in Section B.2.

B.4 Welfare and opaqueness

Thus crowding is not necessarily welfare decreasing if there is a consumer or producer in the economy who benefits from higher consumption or production. While there is over-entry from the investors' point of view, from a welfare point of view this might be beneficial.

The level of socially optimal entry is increasing in α_w, i.e. how much passive investors value crowding. This is obvious from Lemma 4. For large enough M, the social level of aggregate entry in these more general models is

$$M \cdot p = \frac{1}{\beta - \alpha - \alpha_w}$$

which is increasing in α_w. Thus if there was private over-entry (crowding) in our baseline model, for high enough α_w, this over-entry from private point of view might just move entry closer to the social optimum. Thus over-entry is not wasteful taking into account all agents in the economy.