CENTRAL EUROPEAN UNIVERSITY

**DEPARTMENT OF ECONOMICS AND BUSINESS WORKING PAPERS**

economics.ceu.edu

# Learning in Crowded Markets

by

Peter Kondor[1]

Adam Zawadowski[2]

2018/4

[1] London School of Economics. Email: p.kondor@lse.ac.uk
[2] Department of Economics, Central European University, Budapest. Email: zawadowskia@ceu.edu

## Abstract

We study a capital reallocation problem in which investors can enter into a new market where they compete with each other in identifying the best deals. While ex ante investors are uncertain about their relative advantage in identifying the best deals, they can devote costly resources to learning about their relative advantage in a fully flexible way. We find that investors might allocate too much or too little capital to the new market. Increasing competition between investors induces them to learn more, shifting the distribution of entrants towards those with a relative advantage. However, competition does not change overall entry, so it does not affect the efficiency of capital allocation. Thus, learning induced by competition turns out to be wasteful and welfare decreasing. Allowing investors to learn in a fully flexible way - as opposed to requiring Gaussian signals - is what makes our argument transparent. We study several extension.

# Learning in Crowded Markets *

Péter Kondor[†] Adam Zawadowski[‡]

London School of Economics Central European University

April 3, 2018

**Abstract**

We study a capital reallocation problem in which investors can enter into a new market where they compete with each other in identifying the best deals. While ex ante investors are uncertain about their relative advantage in identifying the best deals, they can devote costly resources to learning about their relative advantage in a fully flexible way. We find that investors might allocate too much or too little capital to the new market. Increasing competition between investors induces them to learn more, shifting the distribution of entrants towards those with a relative advantage. However, competition does not change overall entry, so it does not affect the efficiency of capital allocation. Thus, learning induced by competition turns out to be wasteful and welfare decreasing. Allowing investors to learn in a fully flexible way – as opposed to requiring Gaussian signals – is what makes our argument transparent. We study several extension.

*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, Toulouse School of Economics. 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.hu/staff/Adam_Zawadowski/

# 1 Introduction

Global capital allocation is increasingly dominated by "smart money", such as hedge funds and venture capitalists. These investors devote vast resources to finding better investment opportunities than their competitors. Does increasing competition make markets more efficient and does it increase welfare?

We study a capital reallocation problem in which some investors have the opportunity to enter into a new market, for example, an emerging market or a new technology sector. Investors compete with each other in identifying the highest surplus deals in this market. However, ex ante they do not know their relative advantage in identifying the best deals. Each investor designs a signal structure to learn about its relative advantage and decides whether to enter accordingly. Depending on the technological parameters, potential shocks and investors characteristics, investors might allocate too much or too little capital to the new market. Increasing competition between investors induces them to learn more, shifting the distribution of entrants towards those with a relative advantage. However, competition does not change overall entry, thus it neither alleviates nor aggravates the inefficiency of capital allocation. Thus, incremental learning due to fiercer competition turns out to be wasteful and welfare decreasing. We show that allowing investors to precisely target learning to maximize the private payoff of information in a fully flexible way – as opposed to requiring Gaussian signals – is an important part of the argument.

Our model is a novel entry-game with heterogenous agents and rational inattention. Each agent (investor hereon) decides whether to enter a new market. Investors' pay-off from entering depends on their type relative to other entrants. In the baseline model we analyze, each entering investor's pay-off is decreasing in the mass of investors entering with a better type. We show that this is the natural assumption if the new market features a decreasing returns to scale aggregate technology. In general, each investor's pay-off might also be harmed or improved by the mass of entrants with a worse type. We show that this can be due to aggregate or idiosyncratic liquidity shocks. In our model, the market might get crowded because the median entrant is hurting other entrants' payoff.

To avoid entering crowded markets, investors can learn about their relative type. In equilibrium, this is similar to learning about the mass of investors entering with a better type. Following the rational inattention approach of Sims (1998, 2003); Woodford (2008); 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 about 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 about its type. However, entering only when very few other investors with a better type entered is costly, because this requires a large reduction in uncertainty. This approach has several advantages: it parsimoniously captures the joint choice of entry and learning, it allows for full flexibility in learning, and it also has an axiomatic foundation based on information theory.

Our main focus is to analyze how the allocation 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 marginal benefit of knowing your type more precisely before entering is increasing in competition because there are more investors with a better type. 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". We show that these two effects cancel out in the aggregate. Thus aggregate entry remains constant and allocative efficiency does not improve due to increased competition.

Whether there is over- or under-entry does not depend on the level of competition but instead on the characteristics of the new investment opportunity and the potential shocks. In particular, we find that markets are more prone to over-investment when: (1) idiosyncratic liquidity shocks are less likely, (2) aggregate shocks are more likely or more severe, and (3) learning is harder. Intuitively, 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. Also over-entry is more likely for investments with more likely and more severe aggregate shocks, since investors do not internalize the fire sale externalities. Finally, markets with higher cost of learning are more prone to over-investment because investors have a harder time conditioning on their relative type to avoid over-entry and thus the outcome resembles the tragedy of commons.

Nevertheless, as competition increases, welfare decreases. The key insight is that the "rat race effect" increases with competition, thus investors choose to learn more. While ceteris paribus more learning can alleviate inefficient over-entry, more learning due to increased competition does not change the amount of entry. In equilibrium, learning induced by competition simply makes better type investors more likely and worse type less likely to enter. However, as the best entrants always find the highest surplus deals, only the total mass of entrants matters for welfare, not their distribution. Thus the higher learning cost implied by more competition is socially wasteful.

The full flexibility of information acquisition plays a crucial role in our argument. It implies that investors can precisely tailor their learning to yield the highest private return. To illustrate the importance of flexibility, we show that if investors were restricted to obtaining Gaussian signals our argument would be obscured. Gaussian signals are suboptimal as they provide information that is unnecessary for the investor to decide whether to enter. This unused information is too costly, leading investors to learn less, implying that the crowding effect dominates the rat rate effect. Thus aggregate entry tends to increase with competition. However, the insight that over-learning decreases welfare is still true. Given that in todays professional market investors are pouring vast resources in optimizing their information gathering processes, we believe that assuming precisely targeted learning technology is realistic, and understanding its implications is important.

We analyze two extensions. First, we show that when better types find socially more valuable deals in the new market, then more competition often leads to a more efficient allocation of capital compared to the planner's solution. However, this is due to over-learning and thus it decreases welfare. Second, 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. Initially, increasing the fraction of sophisticated investors increases welfare since it raises the average sophistication of investors and this can alleviate over-entry. However, further increasing the fraction of sophisticated investors beyond a certain threshold, 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. Identifying the most sophisticated investors as high frequency traders connects this result to the policy debate on the social benefit of ultra high frequency trading.[1]

Our main contribution is to embed learning in a model of capital allocation. Our paper is connected to various branches of literature. First, a growing literature analyzes the consequences of limited information processing capacity based on the rational inattention approach pioneered by Sims (1998, 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,b). Typically, these papers focus on learning about common value uncertainty. Our paper introduces a novel entry game in which agents learn about relative values. Also, none of the above papers directly analyze capital allocation.

Second, 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)

[1] See, for example, Securities and Commission (2010).

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 in better types entering (higher speed) in the aggregate. Furthermore, our insights also work on longer time horizons. There is also a distinct literature on the social value of private learning: e.g. in Hirshleifer (1971) private information can change ex ante incentives for insurance, in Glode, Green, and Lowery (2012) learning effects ex-post trading opportunities. 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.

Third, 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.

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, the median entrant 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.

# 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. $\theta$ is characterizing the investor's ability to identify better deals in this new market than others. Lower $\theta$ implies a better type. The utility gain (or loss, if negative) from entry is given by

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

where $\alpha$ and $\beta$ are constant parameters. $b(\theta)$ denotes the equilibrium mass of entrants with a type better than $\theta$. $a(\theta)$ denotes the equilibrium mass of entrants whose type is worse than $\theta$. We show in the microfoundation that the following two assumptions are natural. We assume that, $\beta + \alpha > 0$, which is without loss of generality, it is simply consistent with the interpretation that a lower $\theta$ represents a better type. Second, $\beta - \alpha > 0$, such that the median entrant imposes a negative externality on others, that is, the market is prone to getting crowded from a social point of view. The two assumptions together imply that $\beta > 0$ while $\alpha$ could be positive or negative. When $\kappa > 0$, better investors have an absolute advantage, that is, better types derive more utility from entering regardless of the entry decision of others. Section 4.1.2 discusses this case, otherwise we analyze the simpler case of $\kappa = 0$.

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 $\alpha$, $\beta$ and $\kappa$ 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 reallocation: this is our leading microfoundation.[2] A richer model is presented in Section 4.1.1.

## 2.2 Microfoundation: Capital reallocation

There are two islands $A$ and $B$ indexed by $i \in \{A, B\}$. Each island is divided into infinitesimal farms owned by a farmer. On both islands, the farmer and her land are both indexed by $\vartheta$ uniformly distributed over $[0, 1]$. The quality of farms is heterogenous, lower $\vartheta$ means a better farm. Namely, on each island if there is a cow on farm $\vartheta$, it produces $\gamma - \delta \cdot \vartheta$ of the consumption good, where $\delta < \gamma$. Initially, on island $A$, each farmer $\vartheta \in [0, k_{A,0}]$ has a cow, while the rest of the farmers do not have any. On island $B$ there are no cows. Farmers cannot move and transfer cows across islands, however, there is a mass $M$ of investors uniformly distributed over types denoted by $\theta \in [0, 1]$ who can. Each can take a single cow from island $A$ and take it to island $B$. Of those who decide to do so, the better types (lower $\theta$) will be able to pick the cows on the worse quality (higher $\vartheta$) non-empty farms on $A$ not yet picked by a better a investors and take them to the best quality (lowest $\vartheta$) empty farms on island $B$. This is consistent with the interpretation that $\theta$ measures investors ability, as better investors are matched with higher surplus deals. Investors have a large endowment of the consumption good which they can use to pay the farmers for the cows. For simplicity, we assume that each farmer contacted by an investor engages in Nash bargaining over the surplus from transferring the cows. As a result, investors end up with $\xi \in (0, 1]$ share of the surplus.

In this abstract level, cows stands for capital, and island $B$ is representing a profitable investment opportunity. In fact, our analysis might apply in various contexts. One interpretation that fits well

[2] In an earlier version of this paper, Kondor and Zawadowski (2016), we provide microfoundations in various other contexts, including production with local spill-overs, consumption with externalities, and academic publications. The critical feature of all microfoundations is that each player's pay-off is lower if better types also enter, while worse entrants can either help or hurt.

is a real-economy decision in which global firms decide whether to enter a new market segment. In another interpretation, venture capitalists decide which start-up to support knowing that they will make a profit only if they identify the start-up which can produce the best product in that future market segment. Alternatively, we can interpret our variables in a financial context where hedge funds and other sophisticated investors are deciding whether to be active in a novel trading strategy like trading the CDS-bond basis. In this context, relative advantage might mean that they can identify a given profitable strategy earlier than others.[3] Finally, our results might also shed light on welfare effects of high-frequency trading where participants can make profit only if they can be the fastest traders among those who decide to enter. In this latter case, relative type is relative speed.

We show the following result. Note that all proofs are relegated to Appendix A.

**Lemma 1.** ***Microfoundation of reduced form parameters.*** *Choosing $k_{A,0} = \frac{1}{\xi \cdot \delta}$, the expected payoff of investor $\theta$ from transporting capital (given that investor $\theta$ can enter at time $t$) simplifies to (1) with $\alpha = 0$ and $\beta = 2\delta \cdot \xi$. This results in both our parametric assumptions being met as $\beta - \alpha = \alpha + \beta = 2\delta \cdot \xi > 0$.*

Note that both our parameter assumptions, $\beta - \alpha > 0$ and $\alpha + \beta > 0$, are driven by $\delta$ which measures the extent of decreasing return to scale. That is, entering with a better type is beneficial because there are more profitable investment opportunities available when not many better investors have entered. This shows that in this context our parameter restrictions are natural even in the absence of any externalities. While in this simple setup $\alpha = \kappa = 0$, in Section 4.1.1, we show how introducing shocks in this same model leads to $\alpha > 0$ or $\alpha < 0$, while in Section 4.1.2 we show that if the technology of transferring the cows depends on the investors' type, that leads to $\kappa > 0$.

To simplify the exposition in the rest of the paper we set $\xi = 1$. Setting $\xi < 1$ does not change any of our main insights and the analysis is relegated to Appendix B.

[3] We explicitly consider a dynamic interpretation of our static model in Section 3.4.

## 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.[4] 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 $\theta$ and $s$. As Sims (1998) argues, 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 $\theta$ in our model is linear in entry. Woodford (2008) derives the optimal signal structure and entry decision rule for such problems which we restate in the lemma below.

**Lemma 2.** ***Optimal signal choice.*** *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 $\theta$. The optimal entry decision conditional on the signal is: enter if $s = 1$, stay out if $s = 0$.*

Thus, similar to Yang (2015a) the conditional probability of entry $m(\theta)$, or equivalently, the conditional probability of getting a signal 1, is the only choice variable. The intuition for the binary signal structure is that the only reason investors want to learn about $\theta$ 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

[4]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).

in Lemma 2 and in $\theta$,

$$L(m) \equiv H(\theta) - H(\theta|s) = H(s) - H(s|\theta) = \tag{2}$$
$$\left(-p\,\log\left[\frac{1}{p}\right] - (1-p)\log\left[\frac{1}{1-p}\right]\right) - \int_0^1 \left(-m(\theta)\log\left[\frac{1}{m(\theta)}\right] - (1-m(\theta))\log\left[\frac{1}{1-m(\theta)}\right]\right) \mathrm{d}\theta$$

where and the first equation is a property of Shannon-entropy. $p$ denotes the unconditional probability of entry and is defined by:

$$p = \int_0^1 m(\tilde{\theta})\mathrm{d}\tilde{\theta} \tag{3}$$

The expression for learning (2) can be understood in the following way. There is no learning if the signal is uninformative about the state, that is, if it prompts the investor to enter with probability $p$ unconditional on its type $\theta$. 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 $\theta$ (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 the cost of learning induced by a signal defined in Lemma 2 is $\mu \cdot L(m)$ where $\mu$ 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.[5] Note that instead of explicitly considering the

[5]The interpretation of the size of $\mu$ also depends on the economic context. In the real economy context, it might measure the uncertainty about the most valued product characteristics in the given market segment. In the financial context, $\mu$ might depend on the extent prices are informative about the mass of early entrants into a strategy. Examples for a trade with low $\mu$ would be that of twin stocks or on-the-run-off-the-run bonds: a large price gap implies that the strategy is not yet very crowded. Another example is merger arbitrage, where the price offered by the bidder is known. On the other hand, with high $\mu$, it is very hard for the investor to determine whether to enter, e.g. because there is no clear price signal whether the trade is still profitable. Examples for such trades include: emerging markets, carry trade, January effect. In the language of Stein (2009), high $\mu$ represents unanchored strategies.

information content of equilibrium prices, we assume that prices are part of the set of inputs for the optimal signal construction. That is, we implicitly assume that inverting prices are just as costly as interpreting any other type of information.[6]

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 $\mu$ can be interpreted as the cost of building a marginally larger information gathering and evaluating machine or writing a longer "code".[7] We believe that given today's markets where professional traders invest vast resources in systems and practices of mapping large amount of data into investment strategies, this is an appropriate modelling approach.

## 2.4 Definition of allocative efficiency, median entrant 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 ("better" than investor $\theta$) becomes:

$$b(\theta) = M \cdot \int_0^\theta m(\tilde{\theta}) \mathrm{d}\tilde{\theta} \tag{4}$$

mass of higher types entering:

$$a(\theta) = M \cdot \int_\theta^1 m(\tilde{\theta}) \mathrm{d}\tilde{\theta} \tag{5}$$

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:

$$R \equiv \int_0^1 m(\theta) \cdot \Delta u(\theta) \cdot \mathrm{d}\theta \tag{6}$$

[6]While this assumption makes our problem very tractable, it is not without loss of generality. It might be that certain $\theta$'s are easier to learn from prices than others: we leave this theoretical question for future research.

[7]An alternative would be to think of capacity as limited and $\mu$ being the Lagrange multiplier of the capacity constraint. We choose to use a fixed $\mu$ 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.

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. 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:

$$V \equiv R - \mu \cdot L \tag{7}$$

which is what investors maximize. Using the same notation the overall welfare in the whole economy can be computed as

$$W \equiv M \cdot V. \tag{8}$$

We will be also interested in the type distribution of entrants. One simple measure of it is the median entrant type $\theta_{1/2}$ defined implicitly as

$$\int_0^{\theta_{1/2}} M \cdot m(\theta) \mathrm{d}\theta = \frac{M \cdot p}{2}.$$

# 3 Model Solution

In this section we present our main results. We allow the payoff defined by (1) using arbitrary $\beta$ and $\alpha$ satisfying our parameter restrictions but restrict $\kappa$ to 0. We analyze the $\kappa > 0$ case in Section 4.1.2. First, we formulate and solve the investors' and the planner's problem. Second, we analyze the efficiency of capital allocation. Third, we analyze welfare. Finally, we explore implications on the speed of capital reallocation under a dynamic interpretation.

## 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:

$$\max_{m(\theta)} \int_0^1 (m(\theta) \cdot \Delta u(\theta) - \mu \cdot L(m)) \, \mathrm{d}\theta. \tag{9}$$

where $\Delta u$ is the utility gain of entrants defined by (1). 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. i.e. that $\Delta u$ depends not only on $\theta$ but on the choice function $m$ of all other investors. The social planner chooses the symmetric function $m_s(\theta)$ to maximize

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

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:*

$$M \cdot \alpha \cdot \int_\theta^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \tilde{m}(\tilde{\theta}) d\tilde{\theta} + 1 = \mu \cdot \left[ \log\left(\frac{m(\theta)}{1 - m(\theta)}\right) - \log\left(\frac{p}{1-p}\right) \right]. \tag{11}$$

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

$$M \cdot (\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[ \log\left(\frac{m_s(\theta)}{1 - m_s(\theta)}\right) - \log\left(\frac{p}{1-p}\right) \right] \tag{12}$$

We define the equilibrium as a Nash equilibrium: all investors choose the $m(\theta)$ function that is their best response to others' $\tilde{m}(\theta)$. We look for a symmetric Nash equilibrium in which all agents choose the same $m(\theta)$ function. To achieve this, we differentiate the FOC with respect to $\theta$. That gives an ordinary differential equation for $m(\theta)$ where the original FOC at $\theta = 0$ (an integral equation) is the boundary condition. The following Proposition presents the solution of this ordinary differential equation up to a constant $m(0)$ (the boundary value).

**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:*

$$m(\theta) = \frac{1}{1 + W_0\left(e^{M \cdot \frac{\alpha+\beta}{\mu} \cdot \theta + \frac{1-m(0)}{m(0)} + \log\left(\frac{1-m(0)}{m(0)}\right)}\right)}, \tag{13}$$

*where $W_0$ denotes the upper branch of the Lambert function*[8] *and $m(0)$ is pinned down by the boundary condition ((11) evaluated at $\theta = 0$):*

$$M \cdot \alpha \cdot p + 1 = \mu \cdot \left[ \log\left(\frac{m(0)}{1 - m(0)}\right) - \log\left(\frac{p}{1-p}\right) \right]. \tag{14}$$

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

$$\frac{\bar{M} \cdot (\alpha + \beta)}{\mu} = e^{-\frac{1-\beta \cdot \bar{M}}{\mu}} - e^{-\frac{1+\alpha \cdot \bar{M}}{\mu}}. \tag{15}$$

Using a similar similar derivation, we obtain the following solution for the planner's problem.

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

**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 $\theta$:*

$$m_s(\theta) = \frac{1}{M \cdot (\beta - \alpha)}. \tag{16}$$

The first feature to note is 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 in the competitive equilibrium is that every investor wants to know whether (s)he is better than the other investors even if this is wasteful from the social planner's point of view.

## 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. To better understand the optimal strategies and aggregate entry, we look first 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 best $M \cdot p$ 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:*

$$M \cdot p|_{\mu=0} = \min\left(\frac{1}{\beta}, 1\right) \tag{17}$$

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

*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:*

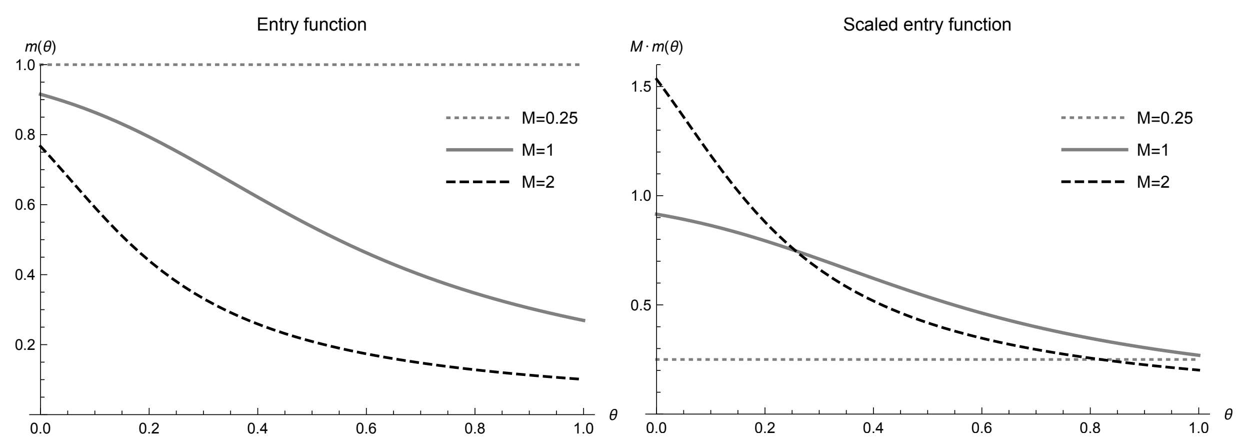
$$M \cdot p|_{\mu\to\infty} = \min\left(\frac{2}{\beta-\alpha}, 1\right) \tag{19}$$

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

Under full information, whether there is under- or over-entry, compared to the social planner's choice, depends on the sign of $\alpha$. There is competitive under-entry (over-entry) if $\alpha > 0$ ($\alpha < 0$), since investors with higher $\theta$ do not take into account the positive (negative) effect of their entry that accrues to entrants with lower $\theta$. Note that under full information without externalities $\alpha = 0$, private entry choice is socially optimal.

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, (s)he does not internalize that her own entry reduces the benefit of entry for everyone else. Note that there is over-entry even in our benchmark microfoundation without externalities ($\alpha = 0$). In this particular case, the "externality" comes from the decreasing returns to scale technology and the lack of information.

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: $\beta = 4$, $\alpha = 2$, $\mu = 0.5$, $\kappa = 0$.

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 $\theta$, 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. 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

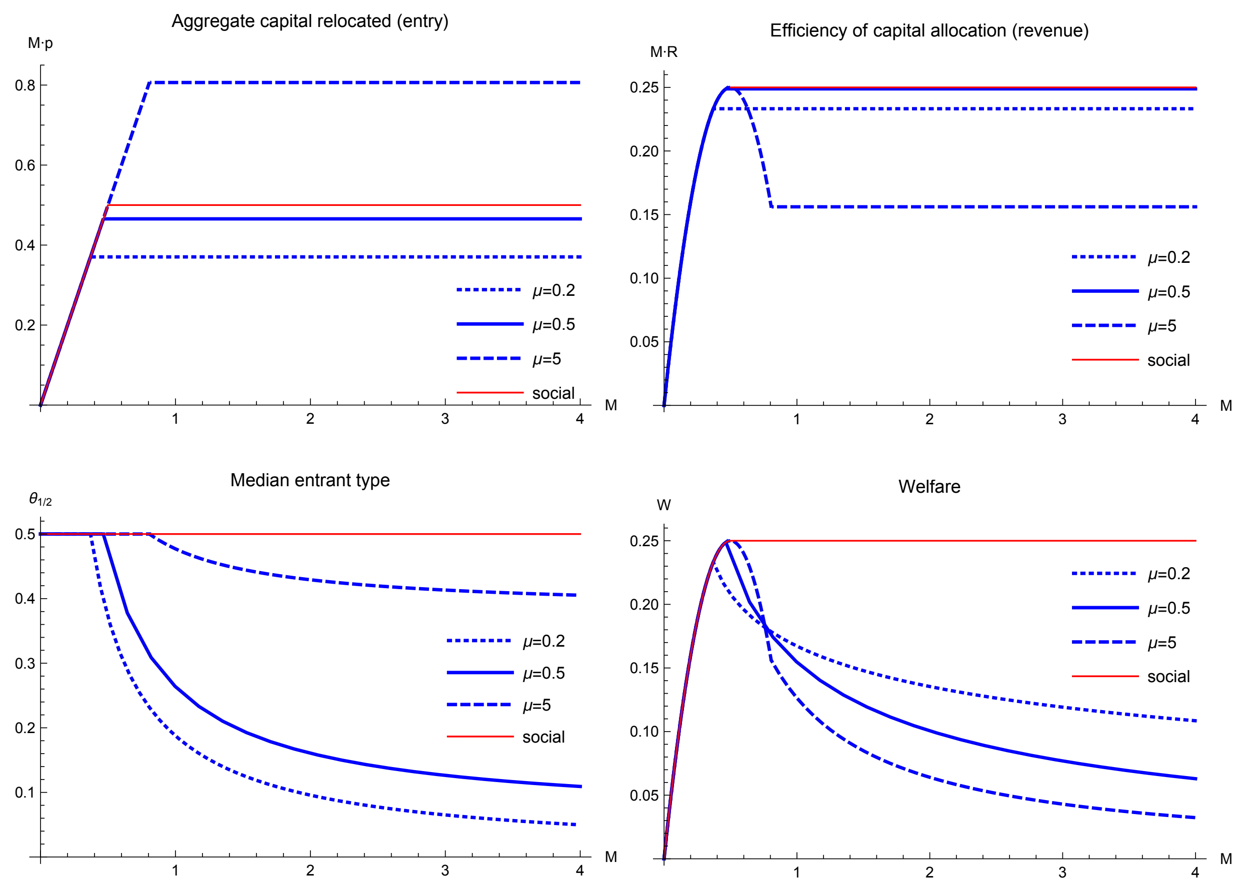
$$\frac{\partial \Delta u\left(\frac{1}{2}\right)}{\partial M} = \alpha \cdot \int_{\frac{1}{2}}^{1} \tilde{m}(\tilde{\theta}) \mathrm{d}\tilde{\theta} - \beta \cdot \int_{0}^{\frac{1}{2}} \tilde{m}(\tilde{\theta}) \mathrm{d}\tilde{\theta} < (\alpha - \beta)\,\frac{p}{2} < 0$$

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. We refer to this as the crowding effect that relies on $\beta - \alpha > 0$. While in equilibrium the strategy of other investors $\tilde{m}(\tilde{\theta})$ also changes, implying further adjustments, as Figure 1 demonstrates, the total effect is still driven by this intuition. In the next proposition we show that these two effects exactly cancel out.

**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 $M$.

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




Aggregate entry, revenue, type of the median entrant 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 $\mu$: $\mu = 0.2$ (dotted line), $\mu = 0.5$ (solid line), $\mu = 5$ (dashed line). Parameters: $\beta = 4$, $\alpha = 2$, $\kappa = 0$.

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. In Section 4.1.1 we extend our microfoundation to better understand the determinants of $\alpha$ and $\beta$ and analyze the effect of the underlying externalities on allocative inefficiency.

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.

### 3.3 Decoupling of welfare and allocative efficiency

Now we turn to the question of welfare as more and more 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 $\mu$ since no resources are spent on learning. The total mass of 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 $\theta$, inducing them to learn about it. A rat race ensues with increasing amounts invested in learning and reduced welfare. Thus an increasing mass of 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, and converges to zero from above as $M \to \infty$:*

$$W(\bar{M}) > \lim_{M\to\infty} W(M) = 0 \tag{21}$$

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

$$W_s(M) = \frac{1}{2\cdot(\beta-\alpha)}. \tag{22}$$

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 the latter 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.

Finally, it is instructive to see how the cost of learning $\mu$ affects welfare. While the cost of learning $\mu$ does not only influence the amount of entry for a given mass $M$ of investors but it does effect the welfare, see Figure 2. First, for high levels of $M$, easier learning (lower $\mu$) means higher welfare $W$. This holds irrespective of the fact that a very low $\mu$ 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 $\mu$ is high.

This also highlights that learning can be beneficial from a welfare point of view, especially if the cost is not that high. The intuition is that this is not a zero sum game, there is "production" in this game which might make information valuable, c.f. Hirshleifer (1971). 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 $\mu$ deters learning and thus helps avert under-entry. Thus if the policymaker can influence the transparency of the market (change $\mu$), it would want to make markets with lots of investors more transparent, while markets with fewer investors and high $\alpha$ might benefit from less transparency.

### 3.4 Speed of capital: a dynamic interpretation

While we analyze a static game, we can give the results a dynamic interpretation. We can interpret the type $\theta$ of an agent as the instant when she learns about the investment opportunity and can decide whether to enter. Thus, a lower $\theta$ agent can find a larger surplus deal, because she learns about the opportunity earlier. Under this interpretation, our analysis of the type distribution of entrants provides the endogenization of the speed of capital reallocation.[9] This interpretation is useful if we think of our model in the context of hedge funds looking for profitable trading strategies. In particular, as the median type of entrants, $\theta_{1/2}$ is getting smaller, capital reallocation is getting faster.

As the lower left panel on Figure 2 illustrates, increasing competition implies faster capital reallocation in our baseline model. The right panel of Figure 1 gives the intuition why this is so. As more learning implies a steeper scaled entry function $M \cdot m(\theta)$, better types who decide earlier, enter with higher probability, while worse types enter with lower probability. Therefore, in the aggregate, each unit of capital is reallocated earlier. We state the formal result in the next proposition.

[9] The dynamic interpretation connects our work to 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 frictionlessly 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 lead to better, e.g. faster, investors entering, markets do not converge to efficiency and welfare deteriorates.

**Proposition 5.** ***Median entrant.*** *The median entrant in the social planner's optimum is $\theta_{1/2} = \frac{1}{2}$. If the mass of investors is small $M \leq \bar{M}$, in the decentralized solution $\theta_{1/2} = \frac{1}{2}$. If the mass of investors is large $M > \bar{M}$, the median type is lower $\theta_{1/2} < \frac{1}{2}$, that is, capital reallocation is too fast. As the number of investors increases, in the limit it converges to:*

$$\lim_{M\to\infty} \theta_{1/2} = \frac{1}{1 + e^{\frac{\bar{M}\cdot(\alpha+\beta)}{2\mu}}} \tag{23}$$

In the dynamic interpretation, increasing the number of investors increases the speed at which capital is reallocated. Regardless whether too much or too little capital is reallocated in equilibrium, it is always moves too fast from a social perspective. However, 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. Contrary to what one might expect, the equilibrium speed is not infinite even with a large amount of investors, i.e. $\theta_{1/2}$ converges to a positive constant. 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.

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. We further discuss how the equilibrium speed depends on deep parameters of our micro-foundation in Section 4.1.1.

# 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 consider the case of type-dependent capital reallocation technologies leading to $\kappa > 0$.

### 4.1.1 Introducing shocks in the microfoundation

Remember from the baseline microfoundation in Section 2.2, that on each island if there is a cow on farm $\vartheta$, it produces $\gamma - \delta \cdot \vartheta$ of the consumption good. In the present extension, we allow $\delta$ to be state and island-contingent and denote it by $\delta_i$. For simplicity we assume that investors have full bargaining power ($\xi = 1$).

To explore the effect of various direct externalities in our problem, we consider the possibility that the reallocation of cows (capital) is subject to shocks. There are two time periods $t = 0, 1$. At $t = 0$ the investor decides whether to transfer a cow. The transfer is concluded at $t = 1$, by which time a shock could hit the investor or the island. 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 cow 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 cows in a fire sale.[10] In case of a fire sale the relative advantage of investors does not matter, they sell their cows on island $B$ randomly, i.e. the best investors are not matched with the best farmers. 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 cows from the farmers at its marginal product on island $A$ and sell capital at its marginal product on island $B$. Whenever this activity is profitable, it also decreases the difference between the marginal return on cows 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.

[10] For technical reasons we assume $\frac{\nu}{1-\nu} < \eta \cdot \frac{\delta_c}{2\delta}$, which ensures that not more than $k_{A,0}$ cows are transported from island $A$.

We interpret our deep parameters as follows. $\delta$ 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. $\delta_c$ characterizes the depth of the financial market where claims on these firms and sectors are traded. When $\delta_c$ is large, a sudden selling pressure of the participating investors drives the price down significantly. In contrast, $\eta$ and $\nu$ characterize the investors as opposed to the markets. A large $\eta$ 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 $\nu$ 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:*

$$\alpha = \left(\frac{1}{2} + (1-\nu)^2\left(\frac{(1-\eta)}{2} - 1\right)\right)\cdot\delta - \frac{1}{2}(1-\nu)^2\cdot\eta\cdot\delta_c \tag{24}$$

$$\beta = \left(\frac{1}{2} + (1-\nu)^2\left(\frac{(1-\eta)}{2} + 1\right)\right)\cdot\delta + \frac{1}{2}(1-\nu)^2\cdot\eta\cdot\delta_c \tag{25}$$

*These satisfy the parametric assumptions $\beta-\alpha>0$ and $\beta+\alpha>0$.*

To interpret $\alpha$ and $\beta$, 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, $\eta$, and the illiquidity of the market, $\delta_c$. 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 $\alpha$ is negative without idiosyncratic shock, because of the same logic: better entrants harm worse (higher $\theta$) entrants because they aggravate crowding. Without idiosyncratic shock, the effect of more worse entrants in a liquidity crisis is the same as the effect of more entrants, worse or better.

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

In Proposition 3 we showed that whether there is under- or over-entry is independent of the mass of investors $M$ for $M > \bar{M}$. In the following Proposition 6 we show that the level of under- or over-entry is determined by other characteristics of the capital reallocation problem.

**Proposition 6.** ***Comparative statics of crowding.*** *Under weak assumptions specified in the Appendix, 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 $\mu$, the marginal cost of information*
2. *decreasing in $\delta$, the rate of decreasing returns to scale of the technology in the presence of aggregate shocks, i.e. if $\eta > 0$*
3. *increasing in $\delta_c$, i.e. decreasing in market depth in crisis in the presence of aggregate shocks, i.e. if $\eta > 0$*
4. *increasing in $\eta$, the probability of aggregate shocks*
5. *decreasing in $\nu$, the probability of idiosyncratic shocks*

More frequent aggregate liquidity shocks (larger $\eta$) and less market depth (higher $\delta_c$) 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 $\delta$) 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 $\nu$) makes the market less crowded since it leads to under-entry due to worse entrants not internalizing the positive effect they have on earlier entrants.

Using the dynamic interpretation in Section 3.4, we showed in Proposition 5 that the speed of capital allocation converges to a constant as we increase the mass of investors. In the following Proposition 7, we use the dynamic interpretation to 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 weak assumptions specified in the Appendix, in the limit of a large mass of investors ($M \to \infty$), the equilibrium speed of entry $\frac{1}{\theta_{1/2}}$ is*

1. *decreasing in $\mu$, the marginal cost of information*
2. *increasing in $\delta$, the rate of decreasing returns to scale of the technology in the presence of aggregate shocks, i.e. if $\eta > 0$*
3. *decreasing in $\delta_c$, i.e. increasing in market depth in crisis*
4. *decreasing in $\eta$, the probability of aggregate shocks*
5. *increasing in $\nu$, 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 $\mu$), 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 $\delta$, the higher revenues from capital reallocation and thus more can be spent for learning. Markets with more severe (higher $\delta_c$) or more likely (higher $\eta$) aggregate shocks offer less revenues in expectation, again decreasing the amount spent on learning and thus equilibrium speed.

### 4.1.2 Socially more efficient types

Until now we analyzed the case where payoff of an investor $\theta$ depended only on its rank among those who entered. As a result, the type-dependent part of utility was simply redistributive. That is, the planner was not interested in which type enters only in aggregate entry. In this part, we consider the case of $\kappa > 0$, where better types are more efficient in reallocating capital both in a social and in a private sense. In our microfoundation, this would be the result of the assumption that worse types deliver less healthy cows to island $B$ resulting in a loss of marginal productivity of $-\kappa\theta$.

The next proposition characterizes the equilibrium in this case.

**Proposition 8.** ***Equilibrium with socially more efficient types.*** *If better types are socially more efficient, ($\kappa > 0$), the competitive optimal strategy $m(\theta)$ in the symmetric equilibrium solves the differential equation*

$$(\alpha + \beta) \cdot m(\theta) + \kappa = -\mu \cdot \frac{m'(\theta)}{m(\theta) \cdot (1 - m(\theta))} \tag{26}$$

*with the boundary condition*

$$\alpha \cdot p + 1 = \mu \cdot \left[ \log \left( \frac{m(0)}{1 - m(0)} \right) - \log \left( \frac{p}{1 - p} \right) \right]. \tag{27}$$

*With socially more efficient types ($\kappa > 0$), the socially optimal strategy $m_s(\theta)$ in the symmetric equilibrium solves the differential equation*

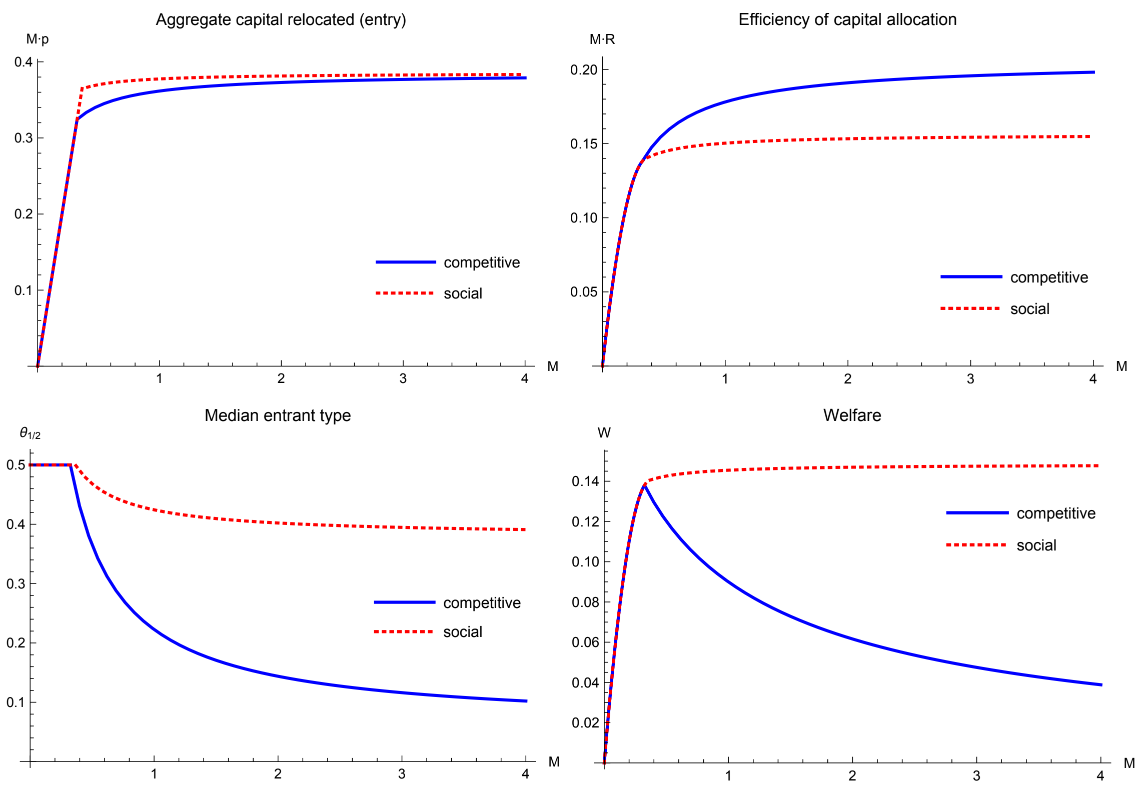
$$\kappa = -\mu \cdot \frac{m_s'(\theta)}{m_s(\theta) \cdot (1 - m_s(\theta))} \tag{28}$$

*subject to the boundary condition*

$$(\alpha - \beta) \cdot p_s + 1 = \mu \cdot \left[ \log \left( \frac{m_s(0)}{1 - m_s(0)} \right) - \log \left( \frac{p_s}{1 - p_s} \right) \right]. \tag{29}$$

From the differential equation (28) 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 (26) since private incentives include the rat race effect: every investor wants to know whether it is ahead of the others. No closed form can be attained in general for the private solution,[11] 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 $\kappa > 0$. As competition increases, more learning makes the median entrant a better type even in the social solution as better (lower $\theta$) entrants are socially beneficial. Note that more investors imply higher entry in both the competitive and the social optimum because better type entry has increasing benefit as the mass of investors increases and

[11]While (28) can be solved in closed form up to constant, to our best knowledge, (26) can only be solved for the special case $\alpha + \beta = \kappa$.

there are more potential good 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 investors than entry because most of the additional revenue comes from the better entrants, not simply more entry. Investors are also motivated by the rat race so there is still overlearning, though the welfare loss is partially offset by the welfare gain from the shifting type distribution of entry (lower $\theta$ entry). Nevertheless, the revenue gains from better entrants, 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

We now show that full flexibility of information acquisition plays a crucial role in our argument. We contrast our framework with fully flexible learning with the perhaps more standard Gaussian formalization (see e.g. Hellwig and Veldkamp (2009)). We show that if investors were restricted to obtain Gaussian signals our argument would be confounded. Gaussian signals are suboptimal as they provide information that is unnecessary for the investor to decide whether to enter. This suboptimality of signals results in new effects that obscure our basic argument.

Suppose that each investor observes a Gaussian signal about its type $\theta$ 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 $\sigma$ standard deviation. $\Phi^{-1}(.;\sigma)$ denotes the inverse of $\Phi(.;\sigma)$.

First, we introduce the transformed type variable $\zeta_i = \Phi^{-1}(\theta_i;\sigma_\zeta)$. Clearly, as $\theta_i$ is uniform on the unit interval, $\zeta_i \sim N\left(0,\sigma_\zeta^2\right)$. Investor $i$ with type $\zeta_i$ can, at a cost $C(\sigma_{\varepsilon_i})$, choose the standard deviation $\sigma_{\varepsilon_i}$ of a signal $s_i = \zeta_i + \varepsilon_i$ about its type where $\varepsilon_i \sim N\left(0, \sigma^2_{\varepsilon_i}\right)$ and $\varepsilon_i$ is independent of $\zeta_i$.[12] 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 full cost, we specify the cost of learning $C_f\left(\sigma_{\varepsilon_i}\right)$ as the reduction in entropy in knowledge after the observation of the signal $s_i$, which, by the property of the normal distribution is:

$$C_f\left(\sigma_{\varepsilon_i}\right) = \frac{1}{2} \cdot \log\left(1 + \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{\zeta}}\right). \tag{30}$$

This is the standard cost function employed in the literature and means that the agent has to pay for all the information acquired through learning the Gaussian signal, irrespective of whether it is then used in the entry decision.

In the second specification, which we denote partial cost $C_p\left(\sigma_{\varepsilon_i}\right)$, we assume the cost is identical to our baseline specification (2) with the only exception that the entry function $m\left(\theta\right)$ 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. Since the investor does not have to pay for unused information, we call this partial cost learning. This second specification is somewhat arbitrary but it allows us to analyze what leads to the difference in results between the our baseline and the Gaussian specification: whether it is the constraints on $m(\theta)$ or the fact that investors have to pay for unnecessary information. The following proposition describes the equilibrium in both cases.

**Proposition 9.** ***Equilibrium with Gaussian learning.*** *In a symmetric equilibrium with Gaussian learning, the optimal strategy of the investor can be fully described by a choice of the signal noise $\sigma_\varepsilon$ 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*

$$m_G\left(\theta\right) = \Phi\left(\bar{s} - \Phi^{-1}\left(\theta; \sigma_\zeta\right); \sigma_\varepsilon\right), \tag{31}$$

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

*and the probability of unconditional entry is:*

$$p = \Phi\left(\bar{s}; \sqrt{\sigma_\zeta^2 + \sigma_\varepsilon^2}\right). \tag{32}$$

*The equilibrium $\sigma_\varepsilon$ and $\bar{s}$ are pinned down by their respective first-order conditions for both cost specification (see the proof).*

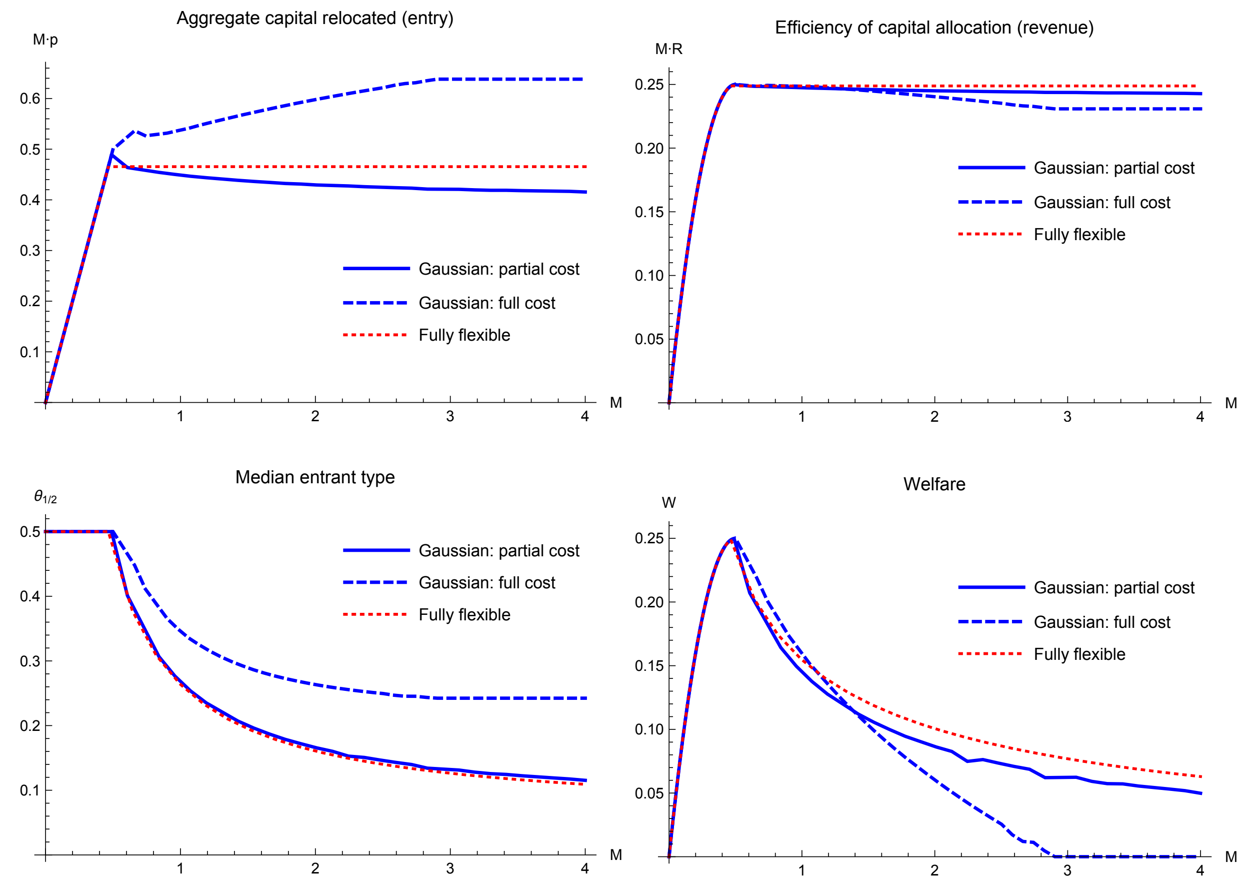
*With full cost learning, if $M$ is large enough no symmetric equilibrium exists, but there is an equilibrium with some investors learning and entering, other investors not entering at all. In this mixed-strategy equilibrium, all investors achieve zero profits.*

Note that $m_G(.)$ can be fully described using two parameters: the standard deviation of signal noise $\sigma_\varepsilon$ 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.

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 (31) 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

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: $\beta = 4$, $\alpha = 2$, $\mu = 0.5$, $\kappa = 0$.

information, that investors cannot learn enough and the equilibrium looks more and more like that in the tragedy of commons game. Another interesting observation is that if $M$ is large enough, learning is 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 shows that the main reason the results in the Gaussian case are different from those in our baseline specification is that investors are forced to pay for information they do not use.[13] Our analysis 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. In the financial context, one can think of high frequency traders and certain hedge funds as more sophisticated than pension funds. We analyze how changing the composition of investors, instead of their total mass $M$, influences allocative efficiency and welfare. This allows us to draw conclusions on e.g. the increasing presence of high frequency traders in markets.

We consider two groups of investors: $\omega \cdot M$ mass of investors is sophisticated and faces a lower learning cost of $\mu_L$, 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 $\theta$ 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 $\theta$. Otherwise the solution would be a set of two joint differential equations which cannot be easily solved.[14]

**Proposition 10.** ***Equilibrium with heterogenous investors.*** *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*

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

$$\omega \cdot \log\left(\frac{(1-\omega)\cdot m_H + \omega \cdot m_L(0)}{m_L(0)}\right) + (1-\omega)\cdot m_H \cdot \log\left(\frac{1-m_L(0)}{m_L(0)}\right) \tag{33}$$

[13] Yang (2015a) makes a related point in coordination games.
[14] See the proof of Proposition 10 for the full problem.

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

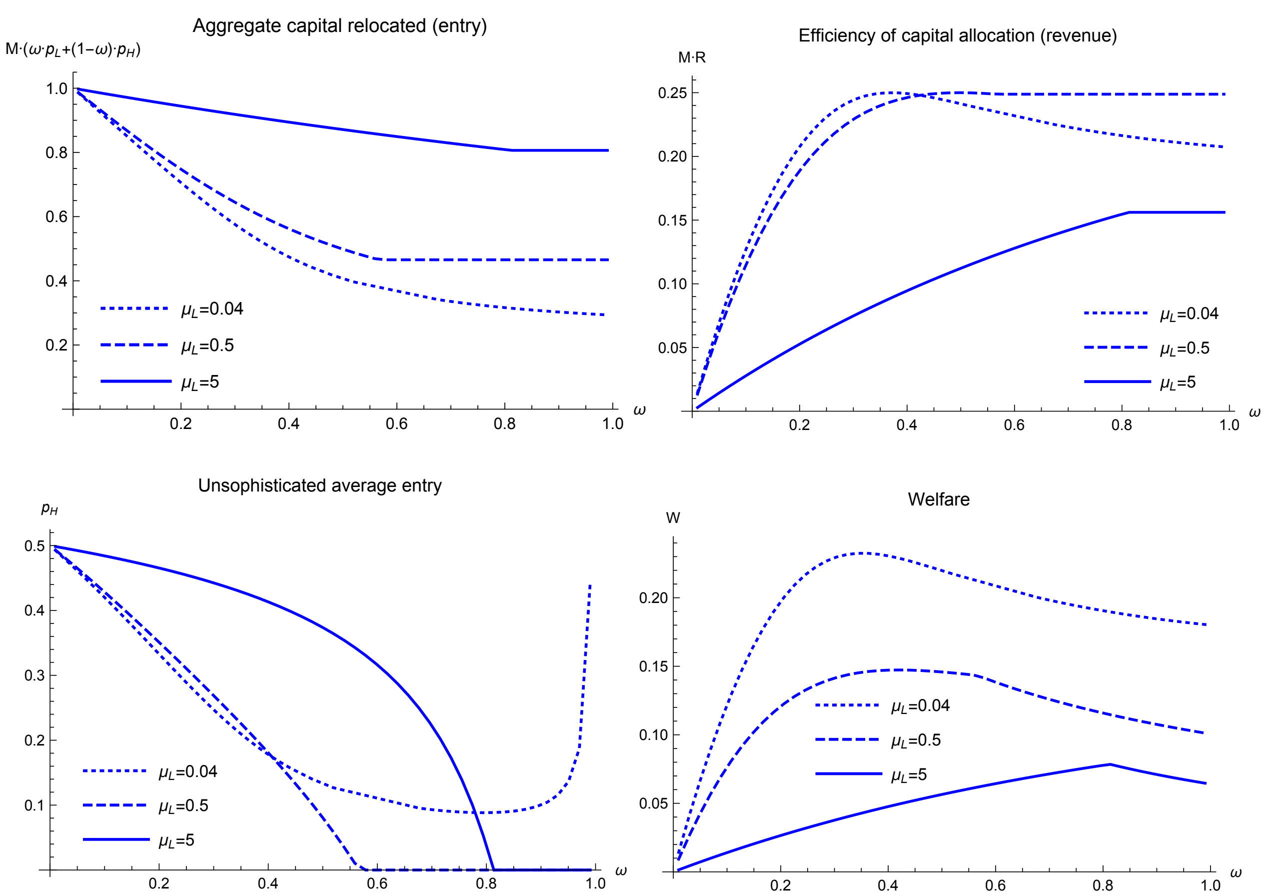
$$1 - M \cdot (\beta - \alpha) \cdot (1-\omega) \cdot m_H + M \cdot \omega \cdot \int_0^1 \left[ \alpha \cdot \int_\theta^1 m_L(\tilde{\theta}) \mathrm{d}\tilde{\theta} - \beta \cdot \int_0^\theta m_L(\tilde{\theta}) \mathrm{d}\tilde{\theta} \right] \mathrm{d}\theta = 0 \tag{34}$$

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

$$M \cdot \alpha \cdot [\omega \cdot p_L + (1-\omega) \cdot m_H] + 1 = \mu_L \cdot \left[ \log \left( \frac{m_L(0)}{1 - m_L(0)} \right) - \log \left( \frac{p_L}{1 - p_L} \right) \right] \tag{35}$$

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

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




Here we change the portion $\omega$ of sophisticated investors who can learn with low cost $\mu_L$, while $1-\omega$ cannot learn. The mass of investors $M$ is kept constant. Parameters: $\beta = 4$, $\alpha = 2$, while $\mu$ 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 $\frac{1}{\beta - \alpha} = \frac{1}{2}$.

We solve the above set of equations numerically since it is analytically intractable. In Figure 5 we vary the portion $\omega$ of sophisticated investors who can learn with cost $\mu_L$. 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 $\omega$ initially increases, welfare increases since the average investor is 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 $\mu$ decreases). There are two effects leading to decreasing welfare as $\omega$ increases further. First, if the sophisticated are very sophisticated (low $\mu_L$) 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 $\mu$ is welfare reducing at low levels of $\mu$ since it aggravates under-entry. Second, and more interestingly, welfare can be decreasing in the share of sophisticated investors $\omega$ even for high $\mu_L$ in the absence of under-entry. The reason is that as $\omega$ 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 $\omega$. 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 $\omega$ of sophisticated increases because there is more and more aggregate entry at low $\theta$, 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 $\omega$ 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 ($\mu_L$ 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 $\theta$ knows it can get positive payoff if its $\theta$ 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 $\theta$ higher than the cutoff since there are no other entrants with higher $\theta$ in equilibrium. Figure 5 shows that in this case with $\omega$ 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. In the dynamic interpretation, this excessive learning leads to inefficiently fast moving capital. Overall, our analysis cautions against using market efficiency or speed of capital allocation as proxies for 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.

Kondor, Peter, and Adam Zawadowski, 2016, Learning in crowded markets, The Paul Woolley Centre Paper Series.

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.

Securities, and Exchange Commission, 2010, Concept release on equity market structure; proposed rule, *Federal Register* 75, 3593–3614.

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.

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.* $b(\theta)$ is the mass of investors who chose to enter (i.e. engage in transporting a cow) with a type better (lower) than $\theta$. Thus investor of type $\theta$ will transfer the cow from the farm indexed by $\vartheta = k_{A,0} - b(\theta)$ on island A and move it to the farm indexed by $\vartheta = b(\theta)$ on island B. The revenue of investor $\theta$ is

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

Choosing $k_{A,0} = \frac{1}{\delta \cdot \xi}$ yields $\alpha$ and $\beta$ in the Lemma.

To show that the amount of cows transferred is less than the amount of cows $k_{A,0}$ available, we use the independently proven Lemma 4 that the maximum entry is $\frac{2}{\beta - \alpha}$ when $\mu \to \infty$ and the payoff to every entrant is zero. In the case of this microfoundation, using $\alpha$ and $\beta$ in the Lemma, $\frac{2}{\beta-\alpha} = \frac{2}{2 \cdot \delta \cdot \xi} = k_{A,0}$, thus at most all cows will be moved to the other island. $\square$

**Proof of Lemma 2**

*Proof.* The proof follows that of Lemma 1 in Woodford (2008). $\square$

**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 $m(\theta) + \chi \cdot \epsilon(\theta)$ as $m(\theta)$, while we keep the entry decision of the others $\tilde{m}$ fixed:

$$\int_0^1 \left( (m(\theta) + \chi \cdot \epsilon(\theta)) \cdot \Delta u(\tilde{m}, \theta) - \mu \cdot L(m(\theta) + \chi \cdot \epsilon(\theta)) \right) \mathrm{d}\theta. \tag{36}$$

We then take derivative w.r.t. $\chi$ and then set $\chi = 0$ yielding the FOC:

$$\int_0^1 \epsilon(\theta) \cdot \left( \Delta u(\tilde{m}, \theta) - \mu \cdot \left[ \log\left( \frac{m(\theta)}{1 - m(\theta)} \right) - \log\left( \frac{\int_0^1 m(\tilde{\theta}) d\tilde{\theta}}{1 - \int_0^1 m(\tilde{\theta}) d\tilde{\theta}} \right) \right] \right) \mathrm{d}\theta = 0. \tag{37}$$

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 $\theta$. Setting $\tilde{m} = m$ we arrive at the symmetric solution we get (11). For the social FOC we also use a perturbation method similar to the proof in Yang (2015a). In the first order perturbation we set $m_s(\theta) + \chi \cdot \epsilon(\theta)$ as $m_s(\theta)$, take derivative w.r.t. $\chi$ and then set $\chi = 0$ in order to arrive at the following equation that has to hold for any function $\epsilon(\theta)$:

$$\int_0^1 \epsilon(\theta) \cdot \left( M \cdot \alpha \cdot \int_\theta^1 m_s(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta m_s(\tilde{\theta}) d\tilde{\theta} - \mu \cdot \left[ \log\left( \frac{m_s(\theta)}{1 - m_s(\theta)} \right) - \log\left( \frac{\int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}}{1 - \int_0^1 m_s(\tilde{\theta}) d\tilde{\theta}} \right) \right] \right) \mathrm{d}\theta + \tag{38}$$

$$+ \int_0^1 m_s(\theta) \cdot \left( M \cdot \alpha \cdot \int_\theta^1 \epsilon(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \epsilon(\tilde{\theta}) d\tilde{\theta} \right) \mathrm{d}\theta = 0 \tag{39}$$

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:

$$M \cdot \alpha \cdot \int_\theta^1 m_s(\tilde{\theta})\mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta m_s(\tilde{\theta})\mathrm{d}\tilde{\theta} + 1 - \mu \cdot \left[\log\left(\frac{m_s(\theta)}{1-m_s(\theta)}\right) - \log\left(\frac{\int_0^1 m_s(\tilde{\theta})\mathrm{d}\tilde{\theta}}{1-\int_0^1 m_s(\tilde{\theta})\mathrm{d}\tilde{\theta}}\right)\right] + \tag{40}$$

$$+ M \cdot \alpha \cdot \int_0^\theta m_s(\tilde{\theta})\mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_\theta^1 m_s(\tilde{\theta})\mathrm{d}\tilde{\theta} = 0 \tag{41}$$

which simplifies to (12). ☐

**Proof of Proposition 1**

*Proof.* Differentiating the FOC (11) we arrive at the following differential equation:

$$(M \cdot \alpha + M \cdot \beta) \cdot \tilde{m}(\theta) = -\frac{\mu \cdot m'(\theta)}{m(\theta) \cdot (1 - m(\theta))}. \tag{42}$$

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 (14). If there is an interior solution (s.t. $m(\theta) \neq 1$), it can be written in the form

$$\frac{\frac{1}{m(\theta)} + \log\left(\frac{1-m(\theta)}{m(\theta)}\right)}{M(\alpha+\beta)} = C + \frac{\theta}{\mu} \tag{43}$$

for an appropriate constant $C$. Setting $\theta = 0$ above and subtracting from the above we can eliminate $C$ and thus arrive at (44).

$$\frac{1}{m(\theta)} + \log\left(\frac{1-m(\theta)}{m(\theta)}\right) - \frac{M(\alpha+\beta)}{\mu} \cdot \theta = \frac{1}{m(0)} + \log\left(\frac{1-m(0)}{m(0)}\right), \tag{44}$$

Taking logs and using the definition of the Lambert function (upper branch if $z > 0$) yields (13). 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:

$$\bar{M} = \lim_{\mu \to \infty} (M \cdot p) \tag{45}$$

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 $\theta$. Thus the implicit equation (44) for $m(\theta)$ can be approximated by

$$\frac{1}{m} - M(\alpha+\beta)\left(C + \frac{\theta}{\mu}\right) = 0 \tag{46}$$

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:

$$m(\theta) = \frac{\mu}{M(\alpha+\beta)(C\mu+\theta)} \tag{47}$$

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

$$M \cdot p = M \cdot \int_0^1 m(\theta)\mathrm{d}\theta = \frac{\mu}{\alpha+\beta} \cdot \log\left(\frac{1}{C\mu}+1\right). \tag{48}$$

Substituting this into the boundary condition (14) yields:

$$\alpha\frac{\mu}{\alpha+\beta}\log\left(\frac{1}{C\mu}+1\right)+1 = \mu\left[\log\left(\frac{1}{CM(\alpha+\beta)-1}\right) - \log\left(\frac{M(\alpha+\beta)-\mu\log\left(\frac{1}{C\mu}+1\right)}{\mu\log\left(\frac{1}{C\mu}+1\right)}\right)\right] \tag{49}$$

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 $M \to \infty$:

$$\mu(-\alpha-\beta)\log\left(\frac{1}{C}\right) + (\alpha+\beta)\left(\mu\log\left(\mu\log\left(\frac{1}{C\mu}+1\right)\right)+1\right) + \alpha\mu\log\left(\frac{1}{C\mu}+1\right) = 0 \tag{50}$$

Using the relation between $C$ and $M \cdot p$ in (48), one can eliminate $C$:

$$(\alpha+\beta)\left(\mu\log(M\cdot p\cdot(\alpha+\beta)) - \mu\log\left(\mu\left(e^{\frac{M\cdot p\cdot(\alpha+\beta)}{\mu}}-1\right)\right) + \alpha M\cdot p + 1\right) = 0 \tag{51}$$

using (45) and rearranging yields equation (15) in the proposition. □

**Proof of Proposition 2**

*Proof.* The derivative of FOC (12) w.r.t. $\theta$ delivers the differential equation

$$0 = -\frac{\mu \cdot m_s'(\theta)}{m_s(\theta)\cdot(1-m_s(\theta))} \tag{52}$$

subject to the boundary condition (setting $\theta = 0$ in (12))

$$M\cdot(\alpha-\beta)\cdot p_s + 1 = \mu\cdot\left[\log\left(\frac{m_s(0)}{1-m_s(0)}\right) - \log\left(\frac{p_s}{1-p_s}\right)\right]. \tag{53}$$

This trivially yields

$$m_s(\theta) = C \tag{54}$$

for some constant $C$, implying $p_s = C$. The boundary condition (53) simplifies to

$$M \cdot (\alpha - \beta) p_s + 1 = 0 \tag{55}$$

implying (16). 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 investor with the highest $\theta = \bar{\theta}$ to enter is indifferent between entering and not:

$$- M \cdot \beta \cdot \bar{\theta} + 1 = 0 \tag{56}$$

yielding Eq. 17.

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.[15] $\bar{\theta}$ is given by maximizing:

$$\int_0^{\bar{\theta}} \big((M \cdot \alpha) \cdot (\bar{\theta} - \theta) - M \cdot \beta \cdot \theta + 1\big) \,\mathrm{d}\theta = \frac{M \cdot \alpha - M \cdot \beta}{2} \cdot \bar{\theta}^2 + 1 \cdot \bar{\theta} \tag{57}$$

yielding the interior optimum in Eq. 18 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 $\theta$ and use a uniform prior. Expected payoff to entering:

$$\int_0^1 (M \cdot \alpha \cdot (1 - \theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1) \,\mathrm{d}\theta = 0 \tag{58}$$

yielding the unconditional entry probability in Eq. 19. 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

$$\int_0^1 p \cdot (M \cdot \alpha \cdot (1 - \theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1) \,\mathrm{d}\theta \tag{59}$$

taking derivative w.r.t. $p$ and setting to zero, this implies the entry probability in Eq. 20. 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**

[15] In fact for the case $\kappa > 0$ this is the unique solution.

*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 (11) at $\theta = 0$ and $\theta = 1$.

$$p = \frac{\mu\left(\log\left(\frac{m(0)}{1-m(0)}\right) - \log\left(\frac{m(1)}{1-m(1)}\right)\right)}{M(\alpha+\beta)} \tag{60}$$

Second, the boundary condition (11) at $\theta = 0$

$$\alpha M p + 1 = \mu\left(\log\left(\frac{m(0)}{1-m(0)}\right) - \log\left(\frac{p}{1-p}\right)\right). \tag{61}$$

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

$$\log\left(\frac{m(0)}{1-m(0)}\right) - \log\left(\frac{m(1)}{1-m(1)}\right) = \frac{M(\alpha+\beta)}{\mu} + \frac{1}{m(0)} - \frac{1}{m(1)} \tag{62}$$

Substituting

$$x_0 = \log\left(\frac{m(0)}{1-m(0)}\right) \tag{63}$$

and

$$x_1 = \log\left(\frac{m(1)}{1-m(1)}\right) \tag{64}$$

the system of three equations can be written as:

$$p = \frac{\mu(x_0 - x_1)}{M(\alpha+\beta)} \tag{65}$$

$$\alpha M p + 1 = \mu\left(x_0 - \log\left(\frac{p}{1-p}\right)\right) \tag{66}$$

$$x_0 - x_1 = \frac{M(\alpha+\beta)}{\mu} + e^{-x_0} - e^{-x_1}. \tag{67}$$

Note that $p > 0$ by definition (3) which implies $x_0 > x_1$ by (65). Substituting out $p$ from (65), (66), (67) we arrive at a system of two equations:

$$F = \mu\left(x_0 - \log\left(\frac{\mu(x_0 - x_1)}{M(\alpha+\beta) + \mu(x_1 - x_0)}\right)\right) - \left(\frac{\alpha\mu(x_0 - x_1)}{\alpha+\beta} + 1\right) = 0 \tag{68}$$

$$G = \frac{M(\alpha+\beta)}{\mu} - (x_0 - x_1) + e^{-x_0} - e^{-x_1} = 0 \tag{69}$$

To prove $M \cdot p$ is constant, it is sufficient to prove $\frac{\partial(M \cdot p)}{\partial M} = 0$ which from (65) is equivalent to

$$\frac{\partial x_0}{\partial M} = \frac{\partial x_1}{\partial M} \tag{70}$$

We apply Cramer's rule both for $x_0$ and $x_1$ to the system of equations (68) and (69):

$$\frac{\partial x_0}{\partial M} = \frac{\begin{vmatrix} \frac{\partial F}{\partial x_0} & -\frac{\partial F}{\partial M} \\ \frac{\partial G}{\partial x_0} & -\frac{\partial G}{\partial M} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} \\ \frac{\partial G}{\partial x_0} & \frac{\partial G}{\partial x_1} \end{vmatrix}} \tag{71}$$

$$\frac{\partial x_1}{\partial M} = \frac{\begin{vmatrix} -\frac{\partial F}{\partial M} & \frac{\partial F}{\partial x_1} \\ -\frac{\partial G}{\partial M} & \frac{\partial G}{\partial x_1} \end{vmatrix}}{\begin{vmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} \\ \frac{\partial G}{\partial x_0} & \frac{\partial G}{\partial x_1} \end{vmatrix}} \tag{72}$$

First, we show that the denominator,

$$\frac{\partial F}{\partial x_0}\frac{\partial G}{\partial x_1} - \frac{\partial F}{\partial x_1}\frac{\partial G}{\partial x_0}$$

is always positive. For this, note that $\frac{\partial F}{\partial x_0} + \frac{\partial F}{\partial x_0} = \mu$. Hence, we can rewrite the denominator as

$$\frac{\partial F}{\partial x_0}\left(\frac{\partial G}{\partial x_0} + \frac{\partial G}{\partial x_1}\right) + \mu\frac{\partial G}{\partial x_0}.$$

As $\frac{\partial G}{\partial x_0} > 0$ and, using $x_0 > x_1$,

$$\frac{\partial G}{\partial x_0} + \frac{\partial G}{\partial x_1} = e^{-x_1} - e^{-x_0} > 0,$$

it is sufficent to show that $\frac{\partial F}{\partial x_0} > 0$. This implied by the observations of $\lim_{M\to 0} \frac{\partial F}{\partial x_0} = \frac{\beta\mu}{\alpha+\beta} > 0$ and

$$\frac{\partial \frac{\partial F}{\partial x_0}}{\partial M} = \frac{\mu^2(\alpha+\beta)}{(M(\alpha+\beta) + \mu(x_1 - x_0))^2} > 0.$$

Second, we show that the numerators of the Cramer rule for the two derivatives are equal, yielding the sufficient condition

$$\frac{(\alpha+\beta)e^{-x_0-x_1}\left(e^{x_0+x_1}(M(\alpha+\beta) + \mu(x_1 - x_0)) - \mu e^{x_0} + \mu e^{x_1}\right)}{M(\alpha+\beta) + \mu(x_1 - x_0)} = 0. \tag{73}$$

It follows from (67) that the denominator is non-zero if $x_0 \neq x_1$. Thus it is sufficient to prove that

$$\frac{M(\alpha+\beta)}{\mu} + (x_1 - x_0) + \frac{1}{e^{x_0}} - \frac{1}{e^{x_1}} = 0, \tag{74}$$

which is exactly the function $G = 0$ defined in (69). Thus the identity holds and we have proved that $M \cdot p$ is constant in $M$ for interior solutions. $\square$

**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

$$M \cdot R = W = M \cdot R_s = W_s = M \cdot \int_0^1 \left(M \cdot \alpha \cdot (1-\theta) - M \cdot \beta \cdot \theta + 1\right) \mathrm{d}\theta = M - \frac{M^2\,(\beta - \alpha)}{2}. \tag{75}$$

To arrive at a formula for $W(M)$ one can rearrange the aggregate learning from (2) to get:

$$M \cdot L = M \int_0^1 m(\theta) \cdot \mu \cdot \left(\log\left(\frac{m(\theta)}{1-m(\theta)}\right) - \log\left(\frac{p}{1-p}\right)\right) \cdot \mathrm{d}\theta - M \int_0^1 \mu \log\left(\frac{1-p}{1-m(\theta)}\right) \cdot \mathrm{d}\theta \tag{76}$$

where the interior part of the first integral multiplying $m(\theta)$ can be replaced using the FOC (11) to yield:

$$M \cdot L = M \int_0^1 m(\theta) \cdot \left[M \cdot \alpha \cdot \int_\theta^1 \tilde{m}(\tilde{\theta}) d\tilde{\theta} - M \cdot \beta \cdot \int_0^\theta \tilde{m}(\tilde{\theta}) d\tilde{\theta} + 1\right] \mathrm{d}\theta - M \int_0^1 \mu \log\left(\frac{1-p}{1-m(\theta)}\right) \cdot \mathrm{d}\theta \tag{77}$$

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:

$$W(M) = M \cdot \int_0^1 \log\left(\frac{1-p}{1-m(\theta)}\right) \mathrm{d}\theta \tag{78}$$

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 (78) 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

$$W_s = M \cdot \int_0^1 p \cdot \left(M \cdot \alpha \cdot (1-\theta) \cdot p - M \cdot \beta \cdot \theta \cdot p + 1\right) \mathrm{d}\theta = \frac{1}{2 \cdot (\beta - \alpha)}. \tag{79}$$

□

**Proof of Proposition 5**

*Proof.* For flat entry function $m(\theta)$, the median entrant is exactly at $\theta_{1/2} = \frac{1}{2}$. For $M > \bar{M}$, the privately optimal entry function is downward sloping. This can be seen by observing that (42) implies that $m'(\theta)$ is always strictly negative. A decreasing $m(\theta)$ implies that the median entrant is smaller than $\frac{1}{2}$. Finally, for $M \to \infty$ we use the result from the proof of Proposition 1 that in the limit $m(\theta)$ converges to (47) where $C$ converges to a finite constant. Solving for $\theta_{1/2}$ using the approximation in the limit:

$$\int_0^{\theta_{1/2}} M \cdot m(\theta) d\theta = \int_0^{\theta_{1/2}} \frac{\mu}{(\alpha + \beta)(C\mu + \theta)} d\theta = \frac{\bar{M}}{2} \tag{80}$$

Evaluating the integral and using the relationship between $C$ and $\bar{M}$ in (48) to substitute out $C$, one gets the expression for $\theta_{1/2}$ stated in the Lemma. $\square$

**Proof of Lemma 5**

*Proof.* With probability $\nu$ the investor is reverted to island $A$ and $\nu \cdot (a(t) + b(t))$ capital is sold by investors at $t = 1$ with random matching between the reverted investors and the best available empty farms. The farmer and the reverted investor engage in Nash bargaining with the investor getting all the surplus. In case of a crisis, $(1-\nu) \cdot (a(t) + b(t))$ capital is sold on island $B$ in a fire sale, with random matching between the investors and the best available farms. Overall, the revenue of an investor that chooses to transport capital at time $t$ is:

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

$$\nu\cdot\underbrace{\left[\gamma-\delta\cdot\left(k_{A,0}-[a(t)+b(t)]+\nu\cdot\frac{a(t)+b(t)}{2}\right)\right]}_{\text{sell price (idiosyncratic shock)}}-\underbrace{\Big[\gamma-\delta\cdot[k_{A,0}-b(t)]\Big]}_{\text{buy price}} \tag{81}$$

Choosing $k_{A,0} = \frac{1}{\delta\cdot(1-\nu)}$, the expected payoff of investor $\theta$ from transporting capital (given that investor $\theta$ can enter at time $t$) simplifies to (1) if $\alpha$ and $\beta$ are given by the equations in the lemma. Resulting in positive crowding and rat-race parameters of:

$$\beta-\alpha=(1-\nu)^2\cdot(\eta\cdot\delta_c+2\delta)>0 \tag{82}$$

$$\alpha+\beta=\left(1+(1-\nu)^2\,(1-\eta)\right)\cdot\delta>0 \tag{83}$$

for all parameter values.

We now check that not more than all the capital is transported from island $A$. We know from the independently proven Lemma 4 that the maximum entry is $\frac{2}{\beta-\alpha}$ when $\mu\to\infty$ and the payoff to every entrant is zero. Thus $\frac{2}{\beta-\alpha}$ is the maximum amount of capital that is moved from island $A$ and this has to be less than $k_{A,0}$:

$$\frac{2}{(1-\nu)^2\cdot(\eta\cdot\delta_c+2\delta)}=\frac{2}{\beta-\alpha}<k_{A,0}=\frac{1}{\delta\cdot(1-\nu)}$$

where we used the expression for $\beta-\alpha$ from above. The above condition is equivalent with

$$\frac{\nu}{1-\nu}<\eta\cdot\frac{\delta_c}{2\delta}$$

which is assumed in footnote 10.

$\square$

**Proof of Proposition 6**

*Proof.* Denote

$$A = e^{\frac{\bar{M}(\alpha+\beta)}{\mu}} - 2e^{\frac{\alpha\bar{M}+1}{\mu}} + 1 \tag{84}$$

and

$$B = \beta\left(e^{\frac{\alpha\bar{M}+1}{\mu}} - e^{\frac{\bar{M}(\alpha+\beta)}{\mu}}\right) + \alpha\left(e^{\frac{\alpha\bar{M}+1}{\mu}} - 1\right). \tag{85}$$

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 $\alpha$ and $\beta$, that is why we refer to them as weak assumptions. Remember that $\bar{M}$ is defined by the implicit equation 15:

$$F = \frac{\bar{M}\cdot(\alpha+\beta)}{\mu} - e^{-\frac{1-\beta\cdot\bar{M}}{\mu}} + e^{-\frac{1+\alpha\cdot\bar{M}}{\mu}} = 0. \tag{86}$$

Using the result from Eq. 16 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:

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial.} = \frac{\partial\bar{M}}{\partial.}\cdot(\beta-\alpha) + \frac{\partial(\beta-\alpha)}{\partial.}\cdot\bar{M} = -\frac{\frac{\partial F}{\partial.}}{\frac{\partial F}{\partial\bar{M}}}\cdot(\beta-\alpha) + \frac{\partial(\beta-\alpha)}{\partial.}\cdot\bar{M} \tag{87}$$

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

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\beta} = -\frac{A}{B}\cdot\alpha\cdot\bar{M} \tag{88}$$

the sign of which is the same as the sign of $\alpha$, and

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\alpha} = \frac{A}{B}\cdot\beta\cdot\bar{M} < 0 \tag{89}$$

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\mu} = \frac{B\cdot\bar{M}+e^{\frac{\bar{M}(\alpha+\beta)}{\mu}}-1}{B}\cdot\frac{\beta-\alpha}{\mu} > 0 \tag{90}$$

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:

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\delta} = \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\beta}\cdot\frac{\partial\beta}{\partial\delta} + \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\alpha}\cdot\frac{\partial\alpha}{\partial\delta} = -\left(-\frac{A}{B}\cdot\bar{M}\right)\cdot\frac{1}{2}\delta_c\cdot\eta\cdot(1-\nu)^2\cdot\big((1-\eta)(1-\nu)^2+1\big) \leq 0 \tag{91}$$

$$\frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\delta_c} = \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\beta}\cdot\frac{\partial\beta}{\partial\delta_c} + \frac{\partial\frac{\bar{M}}{\bar{M}_s}}{\partial\alpha}\cdot\frac{\partial\alpha}{\partial\delta_c} = \left(-\frac{A}{B}\cdot\bar{M}\right)\cdot\frac{1}{2}\delta\cdot\eta\cdot(1-\nu)^2\cdot\big((1-\eta)(1-\nu)^2+1\big) \geq 0 \tag{92}$$

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

$$\frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \eta} = \frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \eta} + \frac{\partial \frac{\bar{M}}{\bar{M}_s}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \eta} = \left(-\frac{A}{B} \cdot \bar{M}\right) \cdot \frac{1}{2}\delta \cdot (1-\nu)^2 \cdot \left((1-\nu)^2(2\delta + \delta_c) + \delta_c\right) > 0 \tag{93}$$

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

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 (85). Remember that $\bar{M}$ is defined by the implicit equation (86) in an implicit form as $F(\bar{M}) = 0$. Denote $\lim_{M\to\infty} \theta_{1/2} = \bar{\theta}_{1/2}$. Using the result that $\bar{\theta}_{1/2} = \frac{1}{e^{\frac{\bar{M}\cdot(\alpha+\beta)}{2\mu}}+1}$ and the implicit function theorem, the derivative of interest $\frac{d\bar{\theta}_{1/2}}{d.}$ for any parameter "." becomes:

$$\frac{d\bar{\theta}_{1/2}}{d.} = \frac{\partial \bar{\theta}_{1/2}}{\partial \bar{M}} \cdot \frac{\partial \bar{M}}{\partial .} + \frac{\partial \bar{\theta}_{1/2}}{\partial .} = \frac{\partial \bar{\theta}_{1/2}}{\partial \bar{M}} \cdot \left(-\frac{\frac{\partial F}{\partial .}}{\frac{\partial F}{\partial \bar{M}}}\right) + \frac{\partial \bar{\theta}_{1/2}}{\partial .} \tag{95}$$

Basic algebra yields:

$$\frac{\partial \bar{\theta}_{1/2}}{\partial \beta} = \frac{\alpha \bar{M} C}{2(-B)\mu} \tag{96}$$

the sign of which is the same as the sign of $\alpha$.

$$\frac{\partial \bar{\theta}_{1/2}}{\partial \alpha} = -\frac{\beta \bar{M} C}{2(-B)\mu} < 0 \tag{97}$$

$$\frac{\partial \bar{\theta}_{1/2}}{\partial \mu} = \frac{(\alpha+\beta) C}{2(-B)\mu^2} > 0 \tag{98}$$

where $C$ is defined by:

$$C = \frac{e^{\frac{\bar{M}(\alpha+\beta)}{2\mu}}}{\left(e^{\frac{\bar{M}(\alpha+\beta)}{2\mu}}+1\right)^2} \cdot \left(e^{\frac{\bar{M}(\alpha+\beta)}{\mu}} - 1\right). \tag{99}$$

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:

$$\frac{\partial \bar{\theta}_{1/2}}{\partial \delta} = \frac{\partial \bar{\theta}_{1/2}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \delta} + \frac{\partial \bar{\theta}_{1/2}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \delta} = \frac{C\delta_c \eta (1-\nu)^2 \bar{M}}{4B\mu}\left((1-\eta)(1-\nu)^2 + 1\right) < 0 \tag{100}$$

$$\frac{\partial \bar{\theta}_{1/2}}{\partial \delta_c} = \frac{\partial \bar{\theta}_{1/2}}{\partial \beta} \cdot \frac{\partial \beta}{\partial \delta_c} + \frac{\partial \bar{\theta}_{1/2}}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \delta_c} = -\frac{C\eta(1-\nu)^2 \bar{M}}{4B\mu}(\alpha+\beta) > 0 \tag{101}$$

Furthermore,

$$\frac{\partial\bar{\theta}_{1/2}}{\partial\eta} = \frac{\partial\bar{\theta}_{1/2}}{\partial\beta}\cdot\frac{\partial\beta}{\partial\eta} + \frac{\partial\bar{\theta}_{1/2}}{\partial\alpha}\cdot\frac{\partial\alpha}{\partial\eta} = -\frac{C(1-\nu)^2\bar{M}}{4B\mu}((\beta-\alpha)\cdot\delta + (\beta+\alpha)\cdot\delta_c) > 0 \tag{102}$$

$$\frac{\partial\bar{\theta}_{1/2}}{\partial\nu} = \frac{\partial\bar{\theta}_{1/2}}{\partial\beta}\cdot\frac{\partial\beta}{\partial\nu} + \frac{\partial\bar{\theta}_{1/2}}{\partial\alpha}\cdot\frac{\partial\alpha}{\partial\nu} = \frac{C\delta(1-\nu)\bar{M}(2\delta+\delta_c\eta)}{2B\mu} < 0 \tag{103}$$

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:

$$1 - \kappa\cdot\theta + \alpha\cdot\int_\theta^1 \tilde{m}(\tilde{\theta})d\tilde{\theta} - \beta\cdot\int_0^\theta \tilde{m}(\tilde{\theta})d\tilde{\theta} = \mu\cdot\left[\log\left(\frac{m(\theta)}{1-m(\theta)}\right) - \log\left(\frac{p}{1-p}\right)\right]. \tag{104}$$

Differentiating this we arrive at the differential equation:

$$(\alpha+\beta)\cdot\tilde{m}(\theta) + \kappa = -\mu\cdot\frac{m'(\theta)}{m(\theta)\cdot(1-m(\theta))}. \tag{105}$$

Imposing symmetry $\tilde{m}(\theta) = m(\theta)$ results in Equation 26. The boundary condition is given by the original integral-differential Equation 104 evaluated at any $\theta$: in Equation 27 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

$$\int_0^1 m_s(\theta)\cdot\Delta u(\theta, m_s)d\theta - \mu\cdot I(m_s) \tag{106}$$

where it takes into account that $\Delta u$ depends not only on $\theta$ 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 $m_s(\theta) + \chi\cdot\epsilon(\theta)$ as $m_s(\theta)$, take derivative w.r.t. $\chi$ and then set $\chi = 0$ in order to arrive at the following equation that has to hold for any function $\epsilon(\theta)$:

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

$$+\int_0^1 m_s(\theta)\cdot\left(\alpha\cdot\int_\theta^1 \epsilon(\tilde{\theta})d\tilde{\theta} - \beta\cdot\int_0^\theta \epsilon(\tilde{\theta})d\tilde{\theta}\right)d\theta = 0 \tag{107}$$

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:

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

$$+\alpha\cdot\int_0^\theta m_s(\tilde{\theta})d\tilde{\theta} - \beta\cdot\int_\theta^1 m_s(\tilde{\theta})d\tilde{\theta} = 0 \tag{108}$$

which simplifies to:

$$(\alpha-\beta)\cdot p_s+1-\kappa\cdot\theta-\mu\cdot\left[\log\left(\frac{m(\theta)}{1-m(\theta)}\right)-\log\left(\frac{p_s}{1-p_s}\right)\right]=0 \tag{109}$$

The derivative of Equation 109 w.r.t. $\theta$ delivers Equation 28, while setting $\theta=0$ in Equation 109 gives the boundary condition Equation 29. ☐

**Proof of Proposition 9**

*Proof.* First, note that the expected payoff is monotonously decreasing in either $\zeta_i$ or $\theta_i$. Also, if $s_i>s_j$ then $g\left(\zeta_i|s_i;\sigma_\zeta,\sigma_{\varepsilon_i}\right)$ first order stochastically dominates $g\left(\zeta_j|s_j;\sigma_\zeta,\sigma_{\varepsilon_i}\right)$ where $g\left(\cdot|\cdot\right)$ is the distribution of $\zeta_i$ 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 $\sigma_{\varepsilon_i}$ and $\bar{s}_i$ are symmetric and dropping the subscripts, each agent solves

$$\max_{\sigma_\varepsilon,\bar{s}}\int_{-\infty}^{\infty}\left[1+\alpha M\int_{\zeta}^{\infty}f\left(s<\tilde{\bar{s}}|\zeta';\tilde{\sigma}_\varepsilon,\sigma_\zeta\right)\phi\left(\zeta';\sigma_\zeta\right)d\zeta'-\beta M\int_{-\infty}^{\zeta}f\left(s<\tilde{\bar{s}}|\zeta';\tilde{\sigma}_\varepsilon,\sigma_\zeta\right)\phi\left(\zeta';\sigma_\zeta\right)d\zeta'\right]\cdot f\left(s<\bar{s}|\zeta;\sigma_\varepsilon,\sigma_\zeta\right)\phi\left(\zeta;\sigma_\zeta\right)d\zeta-C\left(\sigma_\varepsilon\right) \tag{110}$$

where the tilde denotes the choice of others, and $f\left(s<\bar{s}|\zeta;\sigma_\varepsilon,\sigma_\zeta\right)$ is the conditional probability of $s<\bar{s}$ given $\zeta$ with choice $\sigma_\varepsilon$. Note that we can rewrite the expected revenue in (110) as follows

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left(1+\alpha M\int_{\zeta}^{\infty}f\left(s<\tilde{\bar{s}}|\zeta';\tilde{\sigma}_\varepsilon,\sigma_\zeta\right)\phi\left(\zeta';\sigma_\zeta\right)d\zeta'-\beta M\int_{-\infty}^{\zeta}f\left(s<\tilde{\bar{s}}|\zeta';\tilde{\sigma}_\varepsilon,\sigma_\zeta\right)\phi\left(\zeta';\sigma_\zeta\right)d\zeta'\right)\cdot\\
&\qquad f\left(s<\bar{s}|\zeta;\sigma_\varepsilon,\sigma_\zeta\right)\phi\left(\zeta;\sigma_\zeta\right)d\zeta=\\
&=\int_{-\infty}^{\infty}\left[1+\alpha M\int_{\zeta}^{\infty}\Phi\left(\tilde{\bar{s}}-\zeta';\tilde{\sigma}_\varepsilon\right)\phi\left(\zeta';\sigma_\zeta\right)d\zeta'-\beta M\int_{-\infty}^{\zeta}\Phi\left(\tilde{\bar{s}}-\zeta';\tilde{\sigma}_\varepsilon\right)\phi\left(\zeta';\sigma_\zeta\right)d\zeta'\right]\cdot\\
&\qquad\Phi\left(\bar{s}-\zeta;\sigma_\varepsilon\right)\phi\left(\zeta;\sigma_\zeta\right)d\zeta=\\
&=\int_0^1\left[1+\alpha M\underbrace{\int_{\theta}^{1}\Phi\left(\tilde{\bar{s}}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\tilde{\sigma}_\varepsilon\right)d\theta'}_{a(\theta)}-\beta M\underbrace{\int_0^{\theta}\Phi\left(\tilde{\bar{s}}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\tilde{\sigma}_\varepsilon\right)d\theta'}_{b(\theta)}\right]\cdot\\
&\qquad\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta.
\end{aligned} \tag{111}$$

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

$$m_G\left(\theta\right)=\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\zeta\right);\sigma_\varepsilon\right), \tag{112}$$

The probability of unconditional entry is $p=\Phi\left(\bar{s};\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}\right)$ because the standard deviation of the signal is $\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}$ due to the independence of $\zeta_i$ and $\varepsilon_i$. The solution can be obtained by setting marginal cost and benefit of both parameters $\sigma_\varepsilon$ 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:

$$C_p(\sigma_{\varepsilon_i}) = \mu \cdot \left[\left(p\log\left[\frac{1}{p}\right] + (1-p)\log\left[\frac{1}{1-p}\right]\right) - \int_0^1 \left(m_G(\theta)\log\left[\frac{1}{m_G(\theta)}\right] + (1-m_G(\theta))\log\left[\frac{1}{1-m_G(\theta))}\right]\right)\mathrm{d}\theta\right] \tag{113}$$

This implies the two first order conditions:

$$\frac{\partial C_p}{\partial \bar{s}} = \mu \cdot \left[\log\left(\frac{1-p}{p}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\zeta^2+\sigma_\varepsilon^2)}}}{\sqrt{2\pi}\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}} - \int_0^1 \log\left(\frac{1-m_G(\theta)}{m_G(\theta)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta\right] = \int_0^1 \left(1+\alpha\cdot M\cdot a\left(\theta\right) - \beta\cdot M\cdot b\left(\theta\right)\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta = \frac{\partial R}{\partial \bar{s}} \tag{114}$$

$$\frac{\partial C_p}{\partial \sigma_\varepsilon} = \mu \cdot \left[-\log\left(\frac{1-p}{p}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\zeta^2+\sigma_\varepsilon^2)}}\cdot\sigma_\varepsilon}{\sqrt{2\pi}(\sigma_\zeta^2+\sigma_\varepsilon^2)^{\frac{3}{2}}} + \int_0^1 \log\left(\frac{1-m_G(\theta)}{m_G(\theta)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}\cdot\left(\bar{s}-\Phi^{-1}\left(\theta\right)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2}d\theta\right] = -\int_0^1 \left(1+\alpha\cdot M\cdot a\left(\theta\right) - \beta\cdot M\cdot b\left(\theta\right)\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}\cdot\left(\bar{s}-\Phi^{-1}\left(\theta\right)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2}d\theta = \frac{\partial R}{\partial \sigma_\varepsilon} \tag{115}$$

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

$$\mu \cdot \left[\log\left(\frac{1-\Phi\left(\bar{s};\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}\right)}{\Phi\left(\bar{s};\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}\right)}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\zeta^2+\sigma_\varepsilon^2)}}}{\sqrt{2\pi}\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}} - \int_0^1 \log\left(\frac{1-\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\zeta\right);\sigma_\varepsilon\right)}{\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\zeta\right);\sigma_\varepsilon\right)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta\right] = \int_0^1 \left(1+\alpha\cdot M\cdot\int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot\int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta'\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon}d\theta \tag{116}$$

$$\mu\cdot\left[-\log\left(\frac{1-\Phi\left(\bar{s};\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}\right)}{\Phi\left(\bar{s};\sqrt{\sigma_\zeta^2+\sigma_\varepsilon^2}\right)}\right)\frac{e^{-\frac{\bar{s}^2}{2(\sigma_\zeta^2+\sigma_\varepsilon^2)}}\cdot\sigma_\varepsilon}{\sqrt{2\pi}(\sigma_\zeta^2+\sigma_\varepsilon^2)^{\frac{3}{2}}} + \int_0^1 \log\left(\frac{1-\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\zeta\right);\sigma_\varepsilon\right)}{\Phi\left(\bar{s}-\Phi^{-1}\left(\theta;\sigma_\zeta\right);\sigma_\varepsilon\right)}\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}\cdot\left(\bar{s}-\Phi^{-1}\left(\theta\right)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2}d\theta\right] = -\int_0^1 \left(1+\alpha\cdot M\cdot\int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot\int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta'\right)\frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}\cdot\left(\bar{s}-\Phi^{-1}\left(\theta\right)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2}d\theta \tag{117}$$

Second, consider the full cost of learning, where the function $C_f(\sigma_{\varepsilon_i})$ is defined by (30). This implies the two first order conditions:

$$\frac{\partial C_f}{\partial \bar{s}} = 0 = \int_0^1 (1+\alpha\cdot M\cdot a(\theta) - \beta\cdot M\cdot b(\theta)) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon} d\theta = \frac{\partial R}{\partial \bar{s}} \tag{118}$$

$$\frac{\partial C_f}{\partial \sigma_\varepsilon} = -\mu \frac{\sigma_\zeta^2}{\sigma_\varepsilon^3 + \sigma_\varepsilon\cdot\sigma_\zeta^2} = -\int_0^1 (1+\alpha\cdot M\cdot a(\theta) - \beta\cdot M\cdot b(\theta)) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}} \cdot \left(\bar{s}-\Phi^{-1}(\theta)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2} d\theta = \frac{\partial R}{\partial \sigma_\varepsilon} \tag{119}$$

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

$$0 = \int_0^1 \left(1+\alpha\cdot M\cdot \int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot \int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta'\right) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}}}{\sqrt{2\pi}\sigma_\varepsilon} d\theta \tag{120}$$

$$\begin{aligned} &-\mu \frac{\sigma_\zeta^2}{\sigma_\varepsilon^3 + \sigma_\varepsilon\cdot\sigma_\zeta^2} = \\ &-\int_0^1 \left(1+\alpha\cdot M\cdot \int_\theta^1 \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta' - \beta\cdot M\cdot \int_0^\theta \Phi\left(\bar{s}-\Phi^{-1}\left(\theta';\sigma_\zeta\right);\sigma_\varepsilon\right)d\theta'\right) \frac{e^{-\frac{\left(\bar{s}-\Phi^{-1}(\theta)\right)^2}{2\sigma_\varepsilon^2}} \cdot \left(\bar{s}-\Phi^{-1}(\theta)\right)}{\sqrt{2\pi}\sigma_\varepsilon^2} d\theta \end{aligned} \tag{121}$$

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 $\mu_d$ before setting the special case of $\mu_d \to \infty$. In equilibrium, the mass of lower types entering ("before" investor $\theta$) becomes:

$$b(\theta) = M\cdot \int_0^\theta \omega\cdot m_c(\tilde{\theta}) + (1-\omega)\cdot m_d(\tilde{\theta})\mathrm{d}\tilde{\theta} \tag{122}$$

the mass of higher types entering ("after" investor $\theta$):

$$a(\theta) = M \cdot \int_{\theta}^{1} \omega \cdot m_c(\tilde{\theta}) + (1-\omega) \cdot m_d(\tilde{\theta}) \mathrm{d}\tilde{\theta}. \tag{123}$$

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

$$\max_{m_i(\theta)} \int_0^1 (m_i(\theta) \cdot \Delta u(\theta) - \mu_i \cdot L(m_i))\, \mathrm{d}\theta. \tag{124}$$

The optimal solution is characterized by the differential equation for $m_c$

$$(\alpha+\beta)\,(\omega\, \tilde{m}_c(\theta) + (1-\omega)\, \tilde{m}_d(\theta)) = -\frac{\mu_c\, m_c'(\theta)}{m_c(\theta)\,(1-m_c(\theta))} \tag{125}$$

with the boundary condition:

$$M\cdot\alpha\cdot\int_{\theta}^{1} \left(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega)\cdot \tilde{m}_d(\tilde{\theta})\right) \mathrm{d}\tilde{\theta} - M\cdot\beta\cdot\int_0^{\theta} \left(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega)\cdot \tilde{m}_d(\tilde{\theta})\right) \mathrm{d}\tilde{\theta} + 1 = \mu_c\cdot\left[\log\left(\frac{m_c(\theta)}{1-m_c(\theta)}\right) - \log\left(\frac{p_c}{1-p_c}\right)\right]. \tag{126}$$

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.

$$M \cdot \alpha \cdot \int_{\theta}^{1} \left(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega)\cdot \tilde{m}_d(\tilde{\theta})\right) \mathrm{d}\tilde{\theta} - M \cdot \beta \cdot \int_0^{\theta} \left(\omega \cdot \tilde{m}_c(\tilde{\theta}) + (1-\omega)\cdot \tilde{m}_d(\tilde{\theta})\right) \mathrm{d}\tilde{\theta} + 1 = 0 \tag{127}$$

$m_d = 0$ 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 126 can be solved in implicit form and yields Equation 33 in the proposition (evaluated at $\theta = 0$ to substitute out the constant). Equation 126 evaluated at $\theta = 0$ yields the equation for $m_c(0)$ given in Equation 35 in the Proposition. Equation 127 can be simplified to yield 34 in the Proposition. □

# B Online Appendix: Farmers have bargaining power

Here we analyze the case in the microfoundation from Section 2.2 in which farmers have some bargaining power, i.e. $\xi < 1$. While this leaves the payoff function of investors intact, it does influence the welfare calculations. In this case only $\xi$ fraction of the surplus from the capital reallocation is captured by the investors, $1 - \xi$ is by the farmers. Note that the surplus captured by the investors equals $M \cdot R$, thus the total surplus is $\frac{M \cdot R}{\xi}$. Thus instead of (8), the overall welfare in the whole economy can be computed as

$$W \equiv M \cdot V + \frac{1-\xi}{\xi} \cdot M \cdot R.$$

Since for $M > \bar{M}$ total entry and thus $M \cdot R$ is constant, the welfare in equilibrium is the same as before with constant added. Thus welfare still decreases with $M$ but it does not converge to zero in the $M \to \infty$ limit as stated in Proposition 4 but instead to $\frac{1-\xi}{\xi} \cdot M \cdot R$.