

Bardhi, Arjada; Guo, Yingni; Strulovici, Bruno

Working Paper

Spiraling or self-correcting discrimination: A multi-armed bandit approach

CSIO Working Paper, No. 0147

Provided in Cooperation with:

Department of Economics - Center for the Study of Industrial Organization (CSIO),
Northwestern University

Suggested Citation: Bardhi, Arjada; Guo, Yingni; Strulovici, Bruno (2019) : Spiraling or self-correcting discrimination: A multi-armed bandit approach, CSIO Working Paper, No. 0147, Northwestern University, Center for the Study of Industrial Organization (CSIO), Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/274040>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Spiraling or Self-Correcting Discrimination: A Multi-Armed Bandit Approach

Arjada Bardhi* Yingni Guo[†] Bruno Strulovici[‡]

December 10, 2019

Abstract

Can workers from social groups of comparable productivities obtain comparable employment opportunities in the long run? We model dynamic hiring and employer learning via a general Poisson multi-armed bandit framework. Breakdown environments that reveal on-the-job mistakes rather than successes give a large advantage to marginally more productive groups. Breakthrough environments, in contrast, guarantee comparable payoffs to comparable groups. This insight is robust to various sources of across-groups heterogeneity, belief misspecification by employers, and varying degrees of labor demand scarcity. Equal access to productivity investment only enhances prior differences across groups.

JEL: D83, J71, C73

Keywords: breakdown learning, breakthrough learning, endogenous bandits, spiraling property, self-correcting property, ranking multiplicity

1 Introduction

Group belonging shapes employment opportunities in fundamental ways. Despite the substantial legal progress of Equal Employment Opportunity laws, evidence abounds that a

*Department of Economics, Duke University. Email: arjada.bardhi@duke.edu.

[†]Department of Economics, Northwestern University. Email: yingni.guo@northwestern.edu.

[‡]Department of Economics, Northwestern University. Email: b-strulovici@northwestern.edu.

worker’s race, gender, or age continue to loom large in the minds of employers (Pager and Karafin (2009), Bertrand and Duflo (2017), Carlsson and Eriksson (2019)). Models of statistical discrimination suggest that group belonging offers a useful statistical proxy for hard-to-observe characteristics, such as productivity, reliability, or cultural fit in the workplace. Hence one would expect that as the gap between different groups’ workplace performance gradually narrows, employers would treat all workers similarly, and as a result labor market outcomes of different groups would converge. Yet significant group differences in labor markets persist despite social progress. Such differences are particularly stark in higher-ranked occupations, for which workers need opportunities to prove their value in the workplace. Minority groups and women remain scarcely represented in the upper echelons of the political, academic, and corporate spheres.

This paper puts forth a theory of how group belonging shapes workers’ lifetime employment opportunities, from obtaining scarce first chances to being allowed extended periods to prove themselves. Do small differences in employers’ prior beliefs about the productivities of different groups possibly lead to wide disparities in their lifetime prospects? That is, does dynamic learning by employers correct or enhance initial differences across groups? Are certain jobs inherently prone to wider and more persistent disparities? Does equal access to investment in productivity have the potential to correct such disparities? Do ex-ante less productive groups ever get priority in hiring?

To address these questions, we model competition for scarce opportunities between workers of different social groups through a multi-armed *endogenous bandit* framework. An employer first hires workers based on their group’s average productivity. Such group productivity is determined endogenously through educational attainment and labor force participation decisions that group members make in equilibrium. Subsequently, the employer learns about workers’ productivities gradually through observed performance, and reallocates opportunities accordingly. We study a rich class of dynamic learning environments – given that the form and speed of employer learning varies widely across occupations and organizational ranks.

Our main insight is that an employer’s learning environment plays a fundamental role in the convergence of workers’ lifetime payoffs. In certain learning environments, a small

initial difference between groups has negligible long-run impact. The outcome of a group perceived to be slightly less productive than another is only marginally lower. In other environments a small initial difference triggers significant long-run divergence across groups: a slightly disfavored group fares substantially worse than a slightly favored one. Moreover, equal access to investing in productivity and to entering the labor force do not equalize workers' prospects.

Our stylized model features two long-lived workers, a and b , and a single employer. Players interact repeatedly over a long horizon. Each worker is labeled according to the distinct social group he belongs to. A worker's unobservable productivity is either high or low and is fully persistent. Workers from group a are ex ante more likely to have high productivity. In each period the employer decides which of the two workers to assign a scarce and disposable task to. Whereas workers benefit from being allocated the task despite their productivity, the employer benefits only from highly productive workers.

The employer learns gradually about a worker's productivity through his performance with the task. In modeling dynamic employer learning, we leverage the tractability of Poisson signals. Our benchmark analysis contrasts two canonical learning environments: conclusive-breakthrough learning and conclusive-breakdown learning. Under the former, a high-productivity worker generates a random and perfectly informative signal whereas a low-productivity worker generates no signal. The opposite is true with conclusive breakdowns: only a low-productivity worker generates a signal. Our benchmark insight generalizes to inconclusive environments as well, in which both types can generate a signal. Depending on which type generates it more frequently, we categorize inconclusive environments into breakdown and breakthrough environments.

In both conclusive environments, the employer assigns the first task to worker a since she believes that group a is on average more productive. However, the latter path of assignment decisions differs drastically across the two environments. In the absence of a breakthrough, the employer's belief that worker a has high productivity drops gradually until it reaches her prior belief about worker b . From that point on, the employer splits the task equally between workers until either one of them generates a breakthrough or the employer becomes sufficiently convinced that a productive worker is not to be found. The duration of the initial

interval over which the task is allocated exclusively to worker a reflects the difference in the workers' prior beliefs. The smaller this belief differential, the shorter the a -only interval is, hence the smaller the initial advantage that worker a obtains due to his group belonging. As the belief differential shrinks to zero, so does the starting advantage of worker a . Hence, the breakthrough environment is self-correcting.

This stands in marked contrast to the conclusive-breakdown environment. The absence of a breakdown from worker a makes the employer more optimistic about his productivity. She continues to allocate the task exclusively to worker a until the arrival of a breakdown. Therefore, worker b is granted a chance only if worker a is of low productivity and misperforms. As a result, worker b 's expected payoff is only a fraction of that of worker a 's. The starting advantage that worker a obtains for being from the group with higher average productivity does not depend on the belief differential. Even if group a is ever so slightly superior, this *spirals* into a large payoff advantage in a breakdown environment.

Our analysis endogenizes group differences in productivity by allowing workers to undertake costly one-shot investment in their productivity prior to facing the employer. That is, prior to the game described above low-productivity workers a and b simultaneously decide whether to pay a cost in exchange for a chance at becoming high-productivity workers. Investment is equally costly for both social groups.

We provide a careful taxonomy of all equilibria of this investment game – categorizing them into persistent, reversal, and equalizing equilibria – and compare workers' payoffs across them. Equal access to investment not only does not level the playing field for the two groups, but in fact it magnifies prior differences in both learning environments. In the presence of investment, the intuition provided above becomes more fragile for the breakthrough environment and more pronounced for the breakdown one. With conclusive breakthroughs, although there always exists one equilibrium which guarantees payoff convergence for a and b , there also exist other equilibria in which ex-ante similar workers end up with drastically different expected payoffs. With conclusive breakdowns, post-investment payoff difference between ex-ante comparable workers is generically even larger than in the absence of investment in any persistent or reversal equilibrium.

Our takeaway from contrasting the two environments is that whether learning in orga-

nizations focuses on tracking exemplary performance (i.e. breakthroughs) or poor performance (i.e. breakdowns) is highly consequential for the persistence of discrimination. In breakthrough-like jobs, great achievements by the worker are infrequent and benefit the organization greatly, whereas mistakes are inconsequential. Positions in real estate, sales, and academic research are examples of this. Workers in such positions have to meet a high performance standard in order to demonstrate their ability. Group identities are less likely to cast a long-run shadow in such environments. In contrast, in breakdown-like environments mistakes are detrimental for the organization but successes are not particularly beneficial. For instance, mistakes by a surgeon, an airline pilot, or a security guard are much more consequential than exceptional performance on the job. Other positions that display features of breakdown environments are corporate board seats and Article III federal judges, are closer to bad-news environments.¹ A worker is replaced only if serious mistakes or ethical breaches are confirmed: initial appointment often means a long tenure. This makes it harder for minorities and women to gain access to such appointments.² [Jacobs \(1981\)](#) coined the terms “star jobs” for breakthrough environments and “guardian jobs” for breakdown ones.

In terms of methodological contribution, our paper is the first to bring the relevance of the multi-armed bandit framework to models of discrimination. We model employer learning as an active process: employer acquires information about workers by dynamically allocating opportunities. This departs from the passive information acquisition at the heart of most models of discrimination. Our model has the structure of a *three-armed endogenous bandit problem*. Two of the arms are risky and the other is safe. The quality of the risky arms is determined endogenously in a pre-experimentation investment game.

Although the benchmark model is kept simple for sharpness of insight, its results are robust along key dimensions. First, they generalize to inconclusive learning environments. This bridges the rich continuum of learning environments between breakdown and breakthrough learning. Second, the results extend to more than two groups, more than one worker

¹Article III appointments include Supreme Court justices, court of appeals judges, and district court judges. They are nominated by the President, confirmed by the US Senate, and appointed for a life term under the clause of good behavior. They can be removed from office only through impeachment.

²For the slow progress of women towards corporate board and judicial appointments, see [“What’s Keeping More Women From Board Seats: Little Turnover”](#) (WSJ, 2019) and [“Women in Robes”](#) (Americas Quarterly, 2012).

from each group, and more than a single task. Crucial for our results is the scarcity of tasks relative to labor supply and the equal group size. Third, our results continue to hold in a world with misspecified beliefs by the employer. The breakdown environment magnifies even small amounts of prejudice by the employer. Fourth, results extend to a setting in which groups differ in the speed of learning that they generate rather than in ex-ante productivity.

Our results address a popular view that over time market forces will unequivocally correct discrimination. This view is too simplistic. Whether dynamic learning translates converging productivities among groups into converging opportunities depends on how employers learn about workers. By focusing on hiring dynamics exclusively, our model abstracts away from endogenous wage determination. Wages are fixed over a worker’s tenure and cannot be used as a strategic tool by the employer. Nonetheless, we acknowledge that wage discrimination is another important manifestation of differential treatment and wage flexibility could help in addressing discrimination. In this sense, our conceptual contribution is to clarify which environments are in more urgent need of wage flexibility and other policy interventions.

1.1 Related literature

Statistical discrimination. First and foremost, our paper contributes to the literature on statistical discrimination — an intellectual effort that begins with the seminal contributions of Phelps (1972) and Arrow (1973).³ Phelps (1972) and the literature that originates from it (e.g., Aigner and Cain (1977) and Cornell and Welch (1996)) assume some exogenous difference between social groups. Be that an intrinsic difference in groups’ productivities or a difference in the observability of such productivities, it nonetheless gives rise to outcome inequality among groups. On the other hand, Arrow (1973) and the literature that followed it (e.g., Coate and Loury (1993) and Foster and Vohra (1992)) assume no exogenous inter-group differences. Inequality across groups arises because groups fail to coordinate on the same equilibrium when multiple equilibria exist.⁴

³We refer the reader to Fang and Moro (2011) for an excellent survey of theories of statistical discrimination.

⁴Blume (2006) and Kim and Loury (2018) extend the static setup of Coate and Loury (1993) to incorporate generations of workers. Blume (2006) examines how learning dynamics select among equilibria in a static model. Kim and Loury (2018) examine cross-generational linkages in workers’ incentives to invest, and identify “reputational traps” from which a group cannot escape. In contrast, we examine a single generation

Our approach differs from both these strands of the literature. We follow Phelps (1972) in assuming that groups are exogenously different, but we instead zoom into equilibrium outcomes when such difference is vanishingly small. In Phelps (1972), across-group inequality trivially disappears as the difference between groups shrinks. In contrast, we identify environments in which a small difference across groups (be that either objective or subjectively perceived by the principal) generates significant and persistent inequality. This spiraling effect of arbitrarily small differences highlights the challenge faced by society in eliminating unequal outcomes among social groups.

A notable feature of most papers in these two strands is the lack of strategic interaction between groups: it is as if different groups act in parallel universes. In contrast, we are interested in explaining a world in which workers interact directly by competing for scarce opportunities and in which group identity shapes such competition. From this standpoint our paper is related to Cornell and Welch (1996) and Moro and Norman (2004). Cornell and Welch (1996) assume that one group can send more precise signals, so workers from this group are more likely to generate the best signals and obtain scarce jobs. Moro and Norman (2004) model group interaction by letting firms assign workers to two complementary tasks. They characterize asymmetric equilibria in which ex-ante symmetric groups are largely assigned to different tasks.

Because our focus is on the allocation of hiring (equivalently, promotion) opportunities, we abstract away from endogenous wages, similarly to Coate and Loury (1993) and Foster and Vohra (1992). This parallels a similar focus in recent empirical work on discrimination. Starting with Bertrand and Mullainathan (2004), a successful experimental literature has documented the presence of discrimination in hiring practices through correspondence studies, replicated over many countries and labor market contexts (Bertrand and Duflo (2017)).

The emphasis on dynamic processes naturally connects our paper to the notion of cumulative discrimination (e.g., Blank, Dabady and Citro (2004), Blank (2005)). We examine the impact of repeated interactions within a single social domain (i.e. labor market), as well as spillovers across domains (i.e. from education to labor market). We contribute two key insights to this agenda. First, the learning environment is critical in whether cumu-

of long-lived workers whose productivities are revealed gradually only when assigned tasks.

lative effects appear in labor markets. The juxtaposition of the bad-news and good-news environments shows that some dynamic processes perpetuate discriminatory effects whereas others mitigate them over time, as suggested by [Blank \(2005\)](#). Second, the prospect of future cumulative discrimination in labor markets casts a long shadow on pre-market educational investment. Differential educational attainment across groups further exacerbates inequality in labor market outcomes.

Bandit problems. Our work contributes to the literature that applies the bandit model to labor market learning (e.g., [Jovanovic \(1979\)](#) and [Felli and Harris \(1996\)](#)). [Jovanovic \(1979\)](#) applies the bandit model to study the labor market matching and turnover phenomenon. Each period, a worker chooses to work for one of many firms. The worker’s productivity in a particular firm is learnt more precisely as his job tenure increases. We also make the natural assumption that a worker’s productivity is learnt only if he is assigned a job. However, unlike [Jovanovic \(1979\)](#), we focus on applications where multiple workers compete for scarce job opportunities. We contribute to the bandit and labor market literature by showing that different learning technologies have very different welfare implications for workers from various social backgrounds. Our results shed light on the discussions on labor market discrimination.

We explore competing workers’ incentives to invest in their productivities, so the qualities of the bandit arms are endogenously determined. Since we model arms as strategic players, our paper is closely related to [Bergemann and Valimaki \(1996\)](#), [Felli and Harris \(1996\)](#) and [Deb, Mitchell and Pai \(2019\)](#). [Bergemann and Valimaki \(1996\)](#) and [Felli and Harris \(1996\)](#) endogenize the cost of pulling a bandit arm. [Bergemann and Valimaki \(1996\)](#) considers the pricing decisions of firms which compete for a customer’s business. [Felli and Harris \(1996\)](#) considers the wage decisions of firms which compete for a worker’s employment. [Deb, Mitchell and Pai \(2019\)](#) analyze a reputation-building model cast as a one-arm bandit problem. The arm could be a privately informed worker who partially controls the experimentation and hence the information that the employer receives. Unlike our model, all these papers assume that the qualities of the arms are exogenously given. To the best of our knowledge, our model is the first in the economics literature to endogenize bandit arms’ qualities.

The arms’ endogeneity relates our work to [Ghosh and Hummel \(2012\)](#). They study the design of online learning algorithms, which endogenously determine the quality of costly user-generated contributions. The algorithm aims to identify quickly the best contribution for viewers, subject to incentivizing high-quality contributions. In contrast, our bandit operator does not commit to an allocation mechanism. Our focus is on the arms’ welfare rather than on the optimal allocation mechanism.

Our analysis leverages the tractability of Poisson bandits which have been studied extensively in strategic experimentation models (e.g., [Keller, Rady and Cripps \(2005\)](#), [Keller and Rady \(2010\)](#), [Keller and Rady \(2015\)](#), [Strulovici \(2010\)](#)). We embed this learning environment in a game among multiple strategic arms as opposed to multiple strategic bandit operators, as customary in strategic experimentation models. [Keller, Rady and Cripps \(2005\)](#), [Keller and Rady \(2010\)](#) and [Keller and Rady \(2015\)](#) demonstrate the difference between a good-news environment and a bad-news one in strategic experimentation models. We explore the difference between good-news and bad-news learning in the labor market. Moreover, we allow for the possibility of a mixed learning environment, in which learning proceeds via good news for certain arms and bad news for others.

Employer learning. Our paper also relates to the literature on dynamic employer learning ([Farber and Gibbons \(1996\)](#), [Altonji and Pierret \(2001\)](#)). A key difference is that in our model of scarce opportunities gradual learning is reflected in the dynamic evolution of the probability of a group being hired rather than that of wages. Yet similarly to these papers, as the employer learns over time she relies less on observable characteristics such as group belonging. As a result, initially disadvantaged groups might be allocated opportunities with a delay.

We think of our learning environments as capturing the dynamics of learning in a range of occupations and organizational ranks. On this point our work relates to [Mansour \(2012\)](#), [Altonji \(2005\)](#), [Lange \(2007\)](#), and [Antonovics and Golan \(2012\)](#). They all assume that the rate at which signals about a worker’s productivity arrive varies across occupations, but the learning environment is otherwise fixed. By leveraging the richness of Poisson bandits, we allow the nature of such signals — not merely the arrival rate — to differ across occupations. Poisson learning environments relate to the classification of jobs into “guardian jobs”, “star

jobs”, and “foot soldier jobs” proposed by [Jacobs \(1981\)](#) and [Baron and Kreps \(1999\)](#) in the organizational literature.⁵ Similarly to [Bose and Lang \(2017\)](#), we model these categories through Poisson signals and abstract away from wage determination and moral hazard. But in contrast to them, our focus is on workers’ lifetime payoffs from dynamic turnover rather than employers’ endogenous choice of learning environment.⁶

2 Model

Players and types. Consider a dynamic task-assignment game between a principal (“she”) and two agents (each “he”). The agents are from different social groups. We refer to the agent from group $i \in \{a, b\}$ simply as agent i . Time is continuous $t \in [0, \infty)$ and the players are long-lived. At time $t = 0$, agent i is born endowed with one of two types: $\theta_i = h$ (high type) or $\theta_i = \ell$ (low type). Agent i ’s type is observable to himself but not to other players. The common prior belief is $p_i := \Pr(\theta_i = h) \in (0, 1)$ for each $i \in \{a, b\}$. We assume $p_a > p_b$: that is, agent a is ex-ante more productive.

At $t = 0$, a low-type agent decides whether to undertake a costly investment to improve his type. We let $\theta_i^+ \in \{h, \ell\}$ denote agent i ’s post-investment type. If a low-type agent undertakes the investment, his type is improved to high with probability $\pi \in (0, 1)$; with complementary probability his post-investment type is low. Hence, the probability that agent i ’s post-investment type is high is at most $\bar{p}_i := p_i + (1 - p_i)\pi$. Agent i ’s investment decision and realized type are observed by himself only.

Timing and payoffs. At $t = 0$, a low-type agent decides whether to invest in his type at cost $c > 0$. Subsequently at each $t \geq 0$, the principal allocates a task either to one of the two agents or to a safe arm. Allocating the task to the safe arm can be interpreted as the principal resorting to her outside option.

Payoffs are discounted at common rate $r > 0$ and observed at the end of the horizon.⁷

⁵Even though [Jacobs \(1981\)](#) and [Baron and Kreps \(1999\)](#) do not offer a mathematical model of dynamic learning in these job categories, their definition of star/guardian jobs is analogous to our definition of conclusive-breakthrough/conclusive-breakdown environments. A “foot soldier” job corresponds to $\lambda_h = \lambda_\ell$ in our framework. [Bose and Lang \(2017\)](#) are the first to formalize this categorization as Poisson learning.

⁶Also, [Bose and Lang \(2017\)](#) focus exclusively on “guardian jobs”.

⁷In our formulation, learning proceeds through random signals rather than payoffs. This is equivalent to an alternative formulation in which learning proceeds through observable random payoffs. In this alternative

The principal obtains a flow payoff $v > 0$ if she allocates the task to a high-type agent and zero if she allocates the task to a low-type one. If she allocates the task to the safe arm, she earns a flow payoff $s \in (0, v)$.

An agent obtains a unit flow payoff for as long as he is assigned the task. Otherwise, he receives zero. His type is payoff-irrelevant from his perspective. In particular, this payoff structure for agents implies that wages are fixed over an agent's tenure and unalterable by the principal. We assume that both agents are available for the task at any given time and that there is no friction in switching among agents and the safe arm.⁸

Learning and the bandit formulation. Learning about an agent's type proceeds via Poisson signals. If agent i is allocated the task over interval $[t, t + dt)$ and his type is θ_i^+ , a public signal arrives with probability $\lambda_{\theta_i^+} dt$. With complementary probability $1 - \lambda_{\theta_i^+} dt$ no signal arrives. That is, learning is characterized by the pair of type-dependent arrival rates $(\lambda_h, \lambda_\ell) \in \mathbb{R}_+^2$. Based on whether the arrival of a signal is more likely to suggest a high or a low type, we distinguish two classes of learning environments:

- (i) a signal is a *breakthrough* if $\lambda_h > \lambda_\ell \geq 0$;
- (ii) a signal is a *breakdown* if $\lambda_\ell > \lambda_h \geq 0$.

The case of $\lambda_h = \lambda_\ell \geq 0$ corresponds to uninformative signals. Our analysis mostly focuses on two canonical learning environments: (i) conclusive breakthroughs, i.e. $\lambda_h > \lambda_\ell = 0$, and (ii) conclusive breakdowns, i.e. $\lambda_\ell > \lambda_h = 0$. A conclusive breakthrough perfectly reveals a high-type agent, whereas a conclusive breakdown perfectly reveals a low-type one.

We assume that the principal's prior belief about each agent is sufficiently high so that she prefers to experiment with both agents before turning to the safe arm.

Assumption 1 (Positive experimentation time). *For each $i \in \{a, b\}$, agent i is ex-ante preferred to the safe arm. That is $p_i > \underline{p}$ where \underline{p} is the belief threshold at which the principal*

formulation, in a breakthrough environment type θ_i^+ generates a random lump-sum benefit at arrival rate $\lambda_{\theta_i^+}$, where $\lambda_h > \lambda_\ell \geq 0$. In a breakdown environment, type θ_i^+ generates a random lump-sum cost at arrival rate $\lambda_{\theta_i^+}$, where $\lambda_\ell > \lambda_h \geq 0$. Our formulation allows greater flexibility in comparing welfare across learning environments and for varying the arrival rate across agents.

⁸Our result is robust to a small switching cost.

switches to the safe arm. For conclusive environments,

$$\underline{p} = \begin{cases} \frac{rs}{(r + \lambda_h)(v - s) + rs} & \text{if } \lambda_\ell = 0 \\ \frac{(r + \lambda_\ell)s}{r(v - s) + (r + \lambda_\ell)s} & \text{if } \lambda_h = 0. \end{cases}$$

The threshold \underline{p} is lower than the myopic threshold s/v . This shows that an employer begins with a risky worker even if her expected payoff $p_i v$ is lower than the sure payoff s from a safe worker. Learning has value for an employer who looks to hire a worker in many periods, so she begins with a risky worker even if he brings a slightly lower expected payoff than his counterpart. This stands in sharp contrast to static settings where a risk-neutral employer always chooses a worker with the highest expected payoff.

Given realizations (θ_a^+, θ_b^+) , the principal's subsequent problem is a standard three-armed bandit problem. At any $t \geq 0$, the principal decides between allocating the task to agent a , agent b , or the safe arm. The principal learns about θ_i^+ only for so long as she allocates the task to i . The quality of each agent — that is, his post-investment type — is determined endogenously at the investment stage preceding the bandit problem. We thus take an *endogenous bandit* approach to employer learning.

3 No-investment benchmark

This section analyzes agents' expected payoffs in the absence of an investment opportunity, i.e. when $\theta_i^+ = \theta_i$ for $i \in \{a, b\}$. Focusing on conclusive signals, we compare the agents' payoffs as their expected productivities become arbitrarily similar and address how this comparison depends on whether the signal takes the form of a breakthrough or a breakdown. The sharp insight we obtain from this comparison extends to inconclusive signals.

3.1 Conclusive-breakthrough learning

We start with conclusive breakthroughs: $\lambda_h > \lambda_\ell = 0$. Proposition 3.1 shows that in such a learning environment, an arbitrarily small difference in prior beliefs can only result in an arbitrarily small payoff difference for the two agents. We refer to this as the *self-correcting*

property of the conclusive-breakthrough environment.

Agent a , who is the first to be allocated the task, loses his preferential status if he does not generate a breakthrough within a given time window $[0, t^*)$. The duration of this time window is determined by prior beliefs: the further apart these beliefs are, the higher t^* is.⁹ If no breakthrough arrives until t^* , the principal splits the task equally between the two agents. Hence, for a small difference in prior beliefs, the players quickly enter the phase in which both agents are treated symmetrically, which corresponds to equal continuation payoffs.

Proposition 3.1 (Self-correcting property of breakthroughs). *Let $\lambda_\ell = 0$. As $p_b \uparrow p_a$, the two agents' limit expected payoffs are equal.*

Proof. The principal allocates the task first to agent a until the belief that agent a is a high type falls to p_b in the absence of a breakthrough. After that, the principal splits the task equally between the two agents until either (i) a first breakthrough arrives, or (ii) the belief hits \underline{p} . When the first breakthrough arrives, the principal sticks to the agent who generates that breakthrough. We let $U_i(p_a, p_b)$ be agent i 's payoff given the beliefs (p_a, p_b) . Note that $U_a(p, p) = U_b(p, p)$ for any $p \in (\underline{p}, 1)$. The time t^* it takes for p_a to fall to p_b satisfies:

$$\frac{p_a e^{-\lambda_h t^*}}{p_a e^{-\lambda_h t^*} + 1 - p_a} = p_b \implies t^* = \frac{1}{\lambda_h} \log \frac{p_a(1 - p_b)}{(1 - p_a)p_b}. \quad (1)$$

Over interval $[0, t^*)$, agent a generates a breakthrough with probability $p_a(1 - e^{-\lambda_h t^*})$. If a breakthrough arrives, agent a 's payoff is 1. If it doesn't arrive, agent a 's payoff consists of $1 - e^{-rt^*}$, the flow payoff from $[0, t^*)$, and $U_a(p_b, p_b)$, the continuation payoff from time t^* onward. Agent a 's total payoff is

$$p_a(1 - e^{-\lambda_h t^*}) + (1 - p_a + p_a e^{-\lambda_h t^*})(1 - e^{-rt^*} + e^{-rt^*} U_a(p_b, p_b)).$$

Agent b gets continuation payoff $U_b(p_b, p_b)$ at time t^* if and only if no breakthrough occurs over $[0, t^*)$:

$$(1 - p_a + p_a e^{-\lambda_h t^*}) e^{-rt^*} U_b(p_b, p_b).$$

⁹Duration t^* is also inversely proportional to λ_h : the faster learning is, the shorter the grace period granted to agent a .

As $p_b \uparrow p_a$, $t^* \rightarrow 0$. Therefore, the two agents' payoffs are equal in the limit. ■

The reasoning generalizes to inconclusive breakthroughs. Even though the principal does not assign the task to agent a forever upon the realization of the first breakthrough, there is still a window $[0, t^*)$ over which a should generate a first breakthrough in order to continue being allocated the task exclusively. If no breakthrough arrives over this time window, the belief that $\theta_a = h$ drops down to p_b , at which point both agents receive equal continuation payoffs. It continues to be the case that as $p_b \uparrow p_a$, t^* shrinks to zero. Therefore, the two agents' limit payoffs are equal. Proposition A.1 provides the formal argument.

3.2 Conclusive-breakdown learning

In a conclusive-breakdown environment, $\lambda_\ell > \lambda_h = 0$. The principal allocates the task to agent a for as long as no breakdown is realized. The absence of a breakdown makes her increasingly optimistic that a is a high type. On the other hand, the realization of a breakdown by a prompts the principal to switch to b immediately. Again, b is used for as long as no breakdown is observed. The principal resorts to the safe arm otherwise.

Proposition 3.2 below establishes that conclusive-breakdown learning leads to a *spiraling effect*: even if p_b is just shy of p_a , agent a has a substantially higher payoff than agent b . In fact, agent a obtains the same payoff as if agent b did not exist. Agent a is the first to be hired and remains so until his first breakdown. This stands in contrast to the breakthrough environment, where agent a loses his preferential status if he fails to generate a breakthrough within a given window.

Proposition 3.2 (Spiraling property of breakdowns). *As $p_b \uparrow p_a$, the ratio of agent b 's to agent a 's expected payoff approaches*

$$(1 - p_a) \frac{\lambda_\ell}{\lambda_\ell + r} < 1. \tag{2}$$

Proof. If agent a is a high type, his payoff is 1. If he is a low type and the first breakdown arrives at t , his payoff is $(1 - e^{-rt})$. The arrival time t follows density $\lambda_\ell e^{-\lambda_\ell t}$. Hence, agent a 's expected payoff is:

$$p_a + (1 - p_a) \frac{r}{\lambda_\ell + r}. \tag{3}$$

Agent b 's payoff if the principal begins with him at time t is:

$$e^{-rt} \left(p_b + (1 - p_b) \frac{r}{\lambda_\ell + r} \right).$$

Conditional on agent a being a low type, this time t is distributed according to density $\lambda_\ell e^{-\lambda_\ell t}$. Hence, agent b 's expected payoff is:

$$(1 - p_a) \frac{\lambda_\ell}{\lambda_\ell + r} \left(p_b + (1 - p_b) \frac{r}{\lambda_\ell + r} \right). \quad (4)$$

■

The payoff ratio in proposition 3.2 has two components: (i) $(1 - p_a)$ captures the fact that agent b obtains a chance only if $\theta_a = \ell$, and (ii) $\lambda_\ell / (\lambda_\ell + r)$ reflects the expected time it takes for a 's low type to be revealed. Moreover, even as learning becomes almost instantaneous – i.e., as $\lambda_\ell \rightarrow +\infty$ – this payoff ratio does not approach unity. Agent b never obtains a chance if agent a is a high type.

The spiraling property generalizes to inconclusive breakdowns if the players are sufficiently impatient. As long as $p_a > p_b$, a is the first to be hired and stays employed in the absence of a breakdown. For impatient players, this advantage already leads to a significant payoff advantage for a . The departure from conclusive breakdowns brings a complication that the principal might revisit agents who have generated breakdowns in the past. Proposition A.2 presents the details.

3.3 Generalizations

The spiraling property of breakdown learning and the self-correcting property of breakthrough learning are robust to a number of alternative modeling choices. They continue to hold when (i) agents are objectively identical ex-ante but the principal perceives them to have different expected productivities, (ii) the agents' welfare is calculated differently, (iii) there are many tasks, groups, and agents available to the principal, and (iv) agents have the same expected productivity but differ in the speed of learning that they generate.

Misspecified beliefs. Suppose that the two agents have the same probability p_{true} of being

a high type, but the principal believes that agent b is a high type with a lower probability $p_{\text{mis}} < p_{\text{true}}$. Proposition 3.2 continues to hold, since agent a 's and agent b 's payoffs are still given by equations (3) and (4) (with p_a and p_b being replaced by p_{true}).

Proposition 3.1 continues to hold as well. Duration t^* — analogous to that in (1) — now measures how long it takes for p_{true} to fall to p_{mis} . We let $\hat{U}_a(p_{\text{true}}, p_{\text{mis}})$ and $\hat{U}_b(p_{\text{true}}, p_{\text{mis}})$ be the payoffs of agent a and b if (i) the principal splits the task equally, and (ii) agent a 's and b 's true probabilities of being a high type are p_{mis} and p_{true} , respectively. The payoff $\hat{U}_b(p_{\text{true}}, p_{\text{mis}})$ is higher than $\hat{U}_a(p_{\text{true}}, p_{\text{mis}})$, and converges to it as p_{mis} converges to p_{true} . To extend the proof of proposition 3.1 to the misspecified-belief case, we only need to replace t^* with the new definition and replace $U_i(p_b, p_b)$ with $\hat{U}_i(p_{\text{true}}, p_{\text{mis}})$ for agents' payoffs.

Belief misspecification is quite relevant in discussions of statistical discrimination. Lang and Lehmann (2012) document evidence of “either strong prejudice in only a small portion of the population or widespread mild prejudice” (instead of strong prejudice in a large portion of the population). Our results show that prejudice, even when very small, has very different implications under breakdown and breakthrough learning environments.

Other well-being criteria. The self-correcting and spiraling properties continue to hold if we use other measures of an agent's well-being. For instance, consider the probability that an agent is *eventually hired*, that is, the probability that after some finite time the principal always allocates the task to this agent. In conclusive-breakdown environment this probability is p_a for agent a and $(1 - p_a)p_b$ for agent b . In the conclusive-breakthrough environment it is given by

$$p_a (1 - e^{-\lambda_h t^*}) + (1 - p_a (1 - e^{-\lambda_h t^*})) \text{FB}(p_b),$$

for agent a , and

$$(1 - p_a (1 - e^{-\lambda_h t^*})) \text{FB}(p_b),$$

for agent b . Here, $\text{FB}(p_b)$ is the probability that an agent generates the first breakthrough before the principal switches to the safe arm when the prior belief for both agents is p_b . The ratio of these two probabilities goes to one as $p_b \uparrow p_a$.

Multiple tasks, groups, and agents. The absolute scarcity of agents, groups, and tasks is not essential to the arguments in subsections 3.1 and 3.2. What matters is the scarcity of

tasks relative to available agents and the equal size of groups. The proof of proposition C.1 establishes the details. We illustrate here the argument with more than one agent from each group.

Suppose there are four agents, two from each group: agents a_1, a_2 from group a and b_1, b_2 from group b . Fix $p_a > p_b$. Agents are iid draws from their respective social groups; hence, conditional on the group identity, an agent's productivity is uninformative of that of another agent from the same group.

Under conclusive breakdowns the principal's optimal strategy is as follows. She initially mixes equally between a_1 and a_2 . Once a breakdown is generated by one of them, the principal shifts her resources to the remaining a -agent. Once the other a -agent generates a breakdown as well, the principal allocates the task equally among b_1 and b_2 . Under this strategy, the ratio between the expected payoff of an a -agent to that of a b -agent remains bounded away from 1 as $p_b \uparrow p_a$, by a similar argument to that in Proposition 3.2. In contrast, in the conclusive-breakthrough environment the principal mixes equally between a_1 and a_2 over an interval $[0, t^{**})$ if no breakthrough occurs, where

$$t^{**} = \frac{2}{\lambda_h} \log \left(\frac{p_a(1-p_b)}{p_b(1-p_a)} \right).$$

At time t^{**} , she switches to equally allocating the task among all agents. As $p_b \uparrow p_a$, $t^{**} \rightarrow 0$. Similar to Proposition 3.1, the expected payoffs of b -agents converge to those of a -agents.

Heterogeneity in speed of learning. Suppose that the agents differ in arrival rates of signals rather than the prior beliefs. This would be the case, for instance, if the principal belongs to the same social group as one of the agents, which facilitates learning about his type. We let λ_θ^i denote agent i 's arrival rate when his type is θ . In conclusive-breakthrough (resp., conclusive-breakdown) environment $\lambda_h^a > \lambda_h^b$, $\lambda_\ell^a = \lambda_\ell^b = 0$ (resp., $\lambda_\ell^a > \lambda_\ell^b$, $\lambda_h^a = \lambda_h^b = 0$), and $p_a = p_b = p$. As λ_h^b approaches λ_h^a (resp., λ_ℓ^b approaches λ_ℓ^a) analogous statements to those in Proposition 3.1 and 3.2 hold true.

Agent a has initially a higher Gittins index due to the higher arrival rate; the principal starts experimentation with him. In the absence of a breakthrough in the conclusive-breakthrough environment, belief about agent a 's type drops gradually until the Gittins indices for the two agents are equal. At that point, the principal mixes between the two

agents until one of them generates a breakthrough or their indices drop below the safe arm’s index.¹⁰ As $\lambda_h^b \uparrow \lambda_h^a$ the first interval during which only agent a is allocated the task becomes arbitrarily short and the mixing probabilities in the subsequent phase become arbitrarily close to $1/2$. Therefore, the two agents’ payoffs get arbitrarily close. In contrast, in the conclusive-breakdown environment the limit payoff ratio remains bounded away from one: $(1 - p)\lambda_\ell^a / (\lambda_\ell^a + r) < 1$.

3.4 Breakthroughs vs. breakdowns

This discussion clarifies what is driving the difference between the two classes of learning environments. We focus on conclusive signals in order to make the point sharply, but the discussion applies to inconclusive signals more generally.

Let us entertain the following thought experiment with a single agent. Suppose that the principal stops assigning tasks to the agent once the belief that his type is high drops below some $\underline{p} \in (0, 1)$. Under conclusive breakthroughs a prior p is split over time interval $[t, t + dt)$ into a high posterior of 1 and a low posterior of $\text{BT}(p)$, where:

$$\text{BT}(p) := \frac{p(1 - \lambda_h dt)}{1 - p\lambda_h dt} < p.$$

The agent “secures his job” – that is, he is allocated the task indefinitely – if the posterior belief jumps to one. This is wasteful from the agent’s perspective, since he continues to be allocated the task for as long as the posterior belief is above \underline{p} . Securing the job at this high posterior implies that the low posterior $\text{BT}(p)$ realizes rather frequently so as to add up to prior p . Hence, the sequence of realized bad shocks needed for the posterior to drop below \underline{p} also occurs frequently.

In contrast, in the breakdown environment over each small time interval $[t, t + dt)$ a prior belief p is split into a low posterior of 0 and a high posterior of $\text{BD}(p)$, where:

$$\text{BD}(p) := \frac{p}{1 - (1 - p)\lambda_\ell dt} > p.$$

¹⁰The mixing probabilities are such that in the absence of breakthroughs the Gittins indices remain equal across the two arms. The principal puts a higher mixing probability on agent b due to the lower arrival rate.

The agent is abandoned only if the realized posterior belief is 0 (i.e., only when his type is revealed to be low). Given that the prior is assumed to be above \underline{p} , the optimal information structure for the agent in our thought experiment is to reveal no information, in which case he is allocated the task forever. The conclusive-breakdown environment does not perform as well as revealing no information, but it is rather favorable for the agent: because the low posterior is zero, the high posterior $BD(p)$ has to be realized with sufficiently high probability so as to add up to the prior p .

We think of the two learning environments as different approaches to job performance evaluation. Employers keep track of instances in which a worker either overperforms or underperforms. Breakdowns describe workplace environments that focus more on tracking mistakes rather than successes relative to expected performance. In such environments, the worker is guaranteed his job for as long as not many mistakes are observed. Breakthroughs, on the other hand, describes environments which track successes rather than mistakes in order to determine contract renewal and promotions. A breakthrough learning environment grants a limited time window within which the worker should prove his skills. Which approach is a better approximation of a particular workplace is a matter of both the nature of the occupation and the organizational rank.

4 Unobservable investment in types

This section analyzes the implications of a costly one-shot investment opportunity available to the low types of each agent prior to entering the labor market. Such implications are a priori unclear: access to investment might level the playing field for the ex-ante less productive agent or it might further exacerbate the initial productivity gap. We will show that conclusive-breakthrough and conclusive-breakdown environments continue to have different implications.

The benefit from investment for a low-type agent is the payoff difference between a high type and a low type, weighted by the investment success rate π . An agent invests if this benefit is greater than the cost c . Let (q_a, q_b) denote the post-investment belief pair, which dictates the principal's allocation strategy. Hence, it determines each agent's benefit from investment, which in turn drives his investment decision. In equilibrium, the

post-investment belief pair (q_a, q_b) is consistent with agents' investment decisions. This self-fulfilling mechanism is similar to that in [Coate and Loury \(1993\)](#) and not surprisingly, it leads to multiple equilibria. A notable difference is that in our analysis the benefit from investment for an agent i depends on the post-investment belief of both himself and his competitor $-i$.

A feature common to both the breakthrough and the breakdown environments is that if the principal believes that agent i is better than $-i$ post-investment, then i 's benefit from investment is strictly higher than that of $-i$. Agent i has a stronger incentive to invest, which in equilibrium rationalizes the principal's ranking of i over $-i$. This gives rise to three classes of equilibria which are defined by agents' ranking and referred to as *ranking multiplicity*. In particular, (i) *persistent* equilibria preserve the advantage of agent a , that is $q_a > q_b$; (ii) *reversal* equilibria reverse the ranking, that is $q_a < q_b$; (iii) *equalizing* equilibria put both agents on an equal footing post-investment, that is $q_a = q_b$. Agent i is said to be post-investment favored (resp., disfavored) if $q_i > q_{-i}$ (resp., $q_i < q_{-i}$)

Under conclusive breakdowns, only the post-investment ranking determines the principal's allocation strategy, and the post-investment favored agent's benefit from investment does not depend on (q_a, q_b) . As a result, ranking multiplicity is the only type of multiplicity present, in the sense that each class of equilibria contains at most one equilibrium. In contrast, under conclusive-breakthrough environment, each class sustains possibly multiple equilibria. This is because belief levels also affect the principal's allocation strategy. We call this *belief-level multiplicity*, which in the breakthrough environment co-exists along ranking multiplicity. We characterize all equilibria in both environments, and then analyze agents' limit expected payoffs as their pre-investment productivities get arbitrarily close.

An observation similar to that in the no-investment benchmark continues to hold. With conclusive breakthroughs there always exists at least one equilibrium in which, as the agents' pre-investment probabilities of being a high type converge, their payoffs and post-investment probabilities converge as well. In contrast, with conclusive breakdowns such an equilibrium exists only for a small region of cost levels. Across all other equilibria, a small difference in pre-investment beliefs translates into a substantial payoff difference.

Another side of the no-investment benchmark result emerges as well. The availability

of investment tends to increase inequality among agents. Under conclusive-breakthrough learning there now exist equilibria in which the agents' payoffs diverge as p_b approaches p_a . We show that in the limit these take the form of reversal equilibria or persistent equilibria with asymmetric investment. Similarly, under conclusive breakdowns all persistent and reversal equilibria lead to an even higher limit payoff discrepancy than that identified in the no-investment benchmark.

We begin our analysis of investment equilibria with the case of conclusive breakdowns. This case is a natural starting point for two reasons: (i) benefit from investment takes a very simple form, and (ii) generically, there is a unique equilibrium of each class of equilibria.

4.1 Conclusive-breakdown environment

Given the post-investment belief (q_a, q_b) , let $\beta_i(q_a, q_b)$ denote the benefit from investment for agent i . He invests if this benefit outweighs the cost c . Lemma 4.1 describes this benefit supposing that agent a is post-investment favored. It will be clear that the post-investment favored agent has a higher benefit from investment than the disfavored one does. Moreover, the favored agent's benefit does not depend on the belief levels. The disfavored agent's benefit decreases in the probability that his opponent is a high type. Intuitively, the disfavored agent has no chance to perform if the favored agent is a high type. The benefit from investment when agent b is post-investment favored is analogous.

Lemma 4.1. *Suppose that $q_a > q_b$. Then:*

$$\beta_a(q_a, q_b) = \pi \frac{\mu_\ell}{\mu_\ell + 1} > \beta_b(q_a, q_b) = \pi \frac{\mu_\ell^2(1 - q_a)}{(\mu_\ell + 1)^2},$$

where $\mu_\ell = \lambda_\ell/r$.

If the two agents have the same post-investment belief (i.e., $q_a = q_b$), we assume that the

principal treats them equally:¹¹

$$\beta_a(q_a, q_b) = \beta_b(q_a, q_b) = \pi \frac{\mu_\ell(1 + \mu_\ell(2 - q))}{2(\mu_\ell + 1)^2}, \text{ if } q_a = q_b = q.$$

For an equalizing equilibrium to exist for some cost level, it is necessary that $p_a \leq \bar{p}_b$. The pre-investment beliefs for the two agents should be sufficiently close — or the investment success rate π sufficiently high — so that between an agent b who invests and an agent a who does not, the former has a higher chance of being a high type.

In an equalizing equilibrium, both agents have the same benefit from investment and the ex-ante less productive agent must invest more often. Therefore, both agents must be indifferent between investing and not. The resulting post-investment belief q is uniquely determined by this indifference condition and does not depend on the pre-investment productivities (p_a, p_b) . This post-investment belief is at least p_a and at most \bar{p}_b . This in return determines the cost region under which an equalizing equilibrium can be sustained. The further apart is an agent's ex-ante probability from the other's, the smaller is the set of costs for which the equalizing equilibrium can be sustained.

Proposition 4.1 (Equalizing equilibrium). *In an equalizing equilibrium, the post-investment belief q is given by:*

$$\frac{c}{\pi} = \frac{\mu_\ell(1 + \mu_\ell(2 - q))}{2(\mu_\ell + 1)^2}.$$

Such an equilibrium exists if and only if $p_a \leq \bar{p}_b$ and

$$\frac{\mu_\ell(1 + \mu_\ell(2 - \bar{p}_b))}{2(\mu_\ell + 1)^2} \leq \frac{c}{\pi} \leq \frac{\mu_\ell(1 + \mu_\ell(2 - p_a))}{2(\mu_\ell + 1)^2}. \quad (5)$$

As μ_ℓ increases — i.e. as learning becomes faster — the post-investment belief q is higher in an equalizing equilibrium.¹² Uncertainty about agents' types is resolved more quickly. A high type's payoff increases because his opponent's low type is revealed more quickly, whereas

¹¹We focus on equalizing equilibria with equal treatment to rule out artificial post-investment payoff asymmetries across agents. But there also exist other equalizing equilibria with unequal treatment. In any such equilibrium, at $t = 0$ the principal splits the task in favor of agent b . Once we allow for unequal treatment, an equalizing equilibrium exists if and only if a reversal equilibrium exists.

¹²Condition (5) implies that $\frac{c}{\pi} > \frac{\mu_\ell}{2(\mu_\ell + 1)}$, under which the equilibrium post-investment belief q increases in μ_ℓ . Hence, for both agents, the equilibrium probability of investment increases in μ_ℓ .

a low type's payoff decreases because his own type is revealed more quickly. Hence investment into becoming a high type becomes more valuable, and therefore both agents invest more often.

Our next result characterizes the unique reversal equilibrium. When the investment cost is moderate the ex-ante less productive agent can in fact not only catch up, but surpass the ex-ante more productive agent.

Proposition 4.2 (Reversal equilibrium). *There exists a reversal equilibrium if and only if $p_a < \bar{p}_b$ and*

$$\frac{\mu_\ell^2(1 - \bar{p}_b)}{(1 + \mu_\ell)^2} \leq \frac{c}{\pi} \leq \frac{\mu_\ell}{1 + \mu_\ell}. \quad (6)$$

Generically there is a unique reversal equilibrium in which agent b invests and agent a does not. Moreover, type by type, agent a's pre-investment payoff is at most $\frac{\mu_\ell(1 - \bar{p}_b)}{1 + \mu_\ell}$ times that of agent b.

The lower bound on the cost depends on the prior belief through \bar{p}_b : the higher \bar{p}_b is, the cost region for which a reversal equilibrium can be sustained is strictly larger. Moreover, a reversal equilibrium exists whenever an equalizing equilibrium exists: the cost region in (6) strictly nests that in (5).

Proposition 4.3 characterizes persistent equilibria as the investment cost varies. Persistent equilibria vary in the number of agents who invest, but for any cost level there is a unique persistent equilibrium. The ex-ante more productive agent continues to be in an advantaged position even in the presence of an investment opportunity, in the sense that a persistent equilibrium exists for any investment cost, whereas a reversal or equalizing equilibrium exists only when the cost is moderate.

Proposition 4.3 (Persistent equilibria).

- (i) *For any cost $c > 0$, a persistent equilibrium exists and is generically unique.*
- (ii) *If the cost is sufficient low $\left(\frac{c}{\pi} < \frac{\mu_\ell^2(1 - \bar{p}_a)}{(1 + \mu_\ell)^2}\right)$, both invest. If cost is sufficiently high $\left(\frac{c}{\pi} > \frac{\mu_\ell}{1 + \mu_\ell}\right)$, neither invests. For intermediate cost $\left(\frac{\mu_\ell^2(1 - \bar{p}_a)}{(1 + \mu_\ell)^2} < \frac{c}{\pi} < \frac{\mu_\ell}{1 + \mu_\ell}\right)$, only agent a invests.*

- (iii) If $\frac{c}{\pi} > \frac{\mu_\ell}{1+\mu_\ell}$, type by type agent b 's pre-investment payoff is $\frac{\mu_\ell(1-p_a)}{1+\mu_\ell}$ times that of agent a . If $\frac{c}{\pi} \leq \frac{\mu_\ell}{1+\mu_\ell}$, type by type agent b 's pre-investment payoff is at most $\frac{\mu_\ell(1-\bar{p}_a)}{1+\mu_\ell}$ times that of agent a .

Figure 1 summarizes our characterization of three classes of equilibria.¹³ Each type of equilibrium arises for at least some cost levels, and a persistent equilibrium always exists.

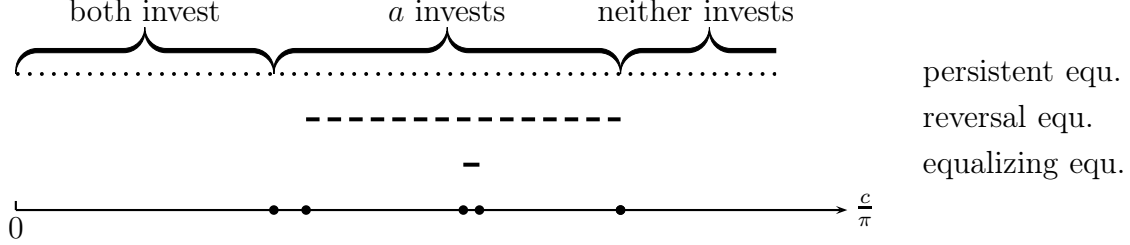


Figure 1: Equilibria across investment cost levels under breakdowns

Convergence of prior beliefs. As $p_b \uparrow p_a$, the lower bound of the cost region for which a reversal equilibrium exists converges to the lower bound of the cost region for a persistent equilibrium with asymmetric investment. The lower end of the equalizing equilibria region expands as well, albeit not all the way to the lower bound of the cost region for which a reversal equilibrium exists. Propositions 4.1 to 4.3 taken together imply the following result on the limit payoffs of the two agents.

Corollary 4.2 (Diverging payoffs in reversal or persistent equilibria). *Suppose that $p_b \uparrow p_a$. In a reversal equilibrium, agent a 's payoff is at most $\frac{\mu_\ell(1-\bar{p}_a)}{1+\mu_\ell}$ times agent b 's payoff. In a persistent equilibrium, agent b 's payoff is at most $\frac{\mu_\ell(1-p_a)}{1+\mu_\ell}$ times agent a 's payoff. Only in an equalizing equilibrium do the two agents' payoffs converge.*

In the limit as $p_b \uparrow p_a$, all equilibria except the equalizing equilibrium lead to weakly greater inequality between agents than the no-investment benchmark, where inequality is defined as the ratio of the higher payoff among the two agents to the lower one. Not only do the agents' payoffs not converge to each other — akin to the no-investment benchmark

¹³Parameters are $\mu_\ell = 4, p_a = 1/3, p_b = 1/4, \pi = 1/5$.

— but in fact they get further apart than in the benchmark due to the existence of an investment opportunity (cf. lemma B.2 in appendix B.1.1).

The inequality among agents is *strictly* higher in any non-equalizing equilibrium in which at least one agent invests. In any such equilibrium the post-investment favored agent is strictly more productive than the more productive agent before investment. Intuitively, the more likely the favored agent is a high type, the larger the payoff ratio between the two agents is. Even after taking into account the investment cost, the expected payoff of the post-investment favored agent is strictly higher than agent a 's payoff in the no-investment benchmark, and that of the disfavored agent is strictly lower than agent b 's. This leads to a higher post-investment payoff ratio. In the persistent equilibrium with no investment by either agent, the ratio stays the same as in the no-investment benchmark.

Finally, there does exist one equilibrium in which the two agents' payoffs converge as the prior beliefs get arbitrarily close: in the equalizing equilibrium, the two agents invest up to the point that they look identical to the principal upon entering the labor market. Such a possibility exists only for a small region of cost levels (given by (5) as $\bar{p}_b \rightarrow \bar{p}_a$ in its LHS).

4.2 Conclusive-breakthrough environment

We first characterize agents' benefit from investment given the post-investment belief (q_a, q_b) . Lemma 4.3 shows that (i) the post-investment favored agent has a higher benefit from investment than the disfavored one; (ii) for the disfavored agent, his benefit from investment increases in his own belief and decreases in his opponent's belief; (iii) the favored agent's benefit is single-peaked in his own belief. Moreover, if the principal believes that the post-investment favored agent is for sure a high type, this favored agent has no benefit from investment.

Lemma 4.3. *Suppose $\underline{p} < q_b < q_a < 1$. Then,*

$$(i) \beta_a(q_a, q_b) > \beta_b(q_a, q_b);$$

$$(ii) \beta_b(q_a, q_b) \text{ increases in } q_b \text{ and decreases in } q_a;$$

(iii) $\beta_a(q_a, q_b)$ is single-peaked in q_a over the domain $q_a \in (q_b, 1)$. Moreover,

$$\lim_{q_a \rightarrow 1} \beta_a(q_a, q_b) = 0.$$

This result highlights a fundamental difference with the conclusive-breakdown environment, in which the principal's optimal strategy depends only on the ranking of (q_a, q_b) but not on their levels. In the conclusive-breakthrough environment, the levels of (q_a, q_b) determine the deadline before which the favored agent has to generate a breakthrough in order to be allocated the task indefinitely. For a fixed q_b , the higher q_a is, the more distant this deadline is. A longer deadline initially encourages the favored agent to invest, but eventually it dampens his investment incentives as it becomes sufficiently distant.¹⁴ This non-monotonicity in the benefit of investment for the post-investment favored agent is similar to that in [Coate and Loury \(1993\)](#). In this sense, the argument in [Coate and Loury \(1993\)](#) is closer to our breakthrough environment.

The explicit expressions for the benefit of investment in [appendix B.2.1](#) allow us to fully characterize all equilibria. Let us start with equalizing equilibria. Suppose that there is an equalizing equilibrium with post-investment belief q . This belief must be larger than agent a 's pre-investment belief p_a and smaller than agent b 's highest attainable belief \bar{p}_b . The benefit from investment is the same for both agents. In [Lemma B.3](#) in the appendix, we show that either this benefit is monotone in q or it is single-peaked in q , so there are at most two equalizing equilibria.

Proposition 4.4 (Equalizing equilibria). *An equalizing equilibrium exists if and only if $p_a \leq \bar{p}_b$ and*

$$\min_{q \in [p_a, \bar{p}_b]} \beta_i(q, q) \leq c \leq \max_{q \in [p_a, \bar{p}_b]} \beta_i(q, q). \quad (7)$$

Under this condition, there exist at most two equalizing equilibria.

In any reversal equilibrium, the ex-ante less productive agent becomes favored after the investment stage by investing more frequently. Such equilibria can induce investment by either both agents or just the ex-ante less productive agent b . It is to be expected that as

¹⁴In [lemma B.3](#) in the appendix, we show that in the domain $q_b \in [p, q_a]$, either $\beta_a(q_a, q_b)$ decreases in q_b or it first increases and then decreases in q_b .

the agents become arbitrarily similar ex ante, i.e. $p_b \uparrow p_a$, the intermediate cost region (8) for which a reversal equilibrium exists does not vanish.

Proposition 4.5 (Reversal equilibrium). *Let $\hat{p} := \arg \max_{p \in [p_a, \bar{p}_b]} \beta_b(p_a, p)$.¹⁵ A reversal equilibrium exists if and only if $\bar{p}_b > p_a$ and*

$$\beta_a(p_a, \bar{p}_b) \leq c \leq \max \{ \beta_i(\bar{p}_b, \bar{p}_b), \beta_b(p_a, \hat{p}) \}. \quad (8)$$

Similar to the breakdown analysis, a persistent equilibrium is guaranteed to exist for any cost level. For sufficiently high cost or sufficiently low cost, the level of investment is the same for both agents. For moderate cost, agent a invests more frequently, thus preserving his favored position. Figure 1 summarizes our characterization of three classes of equilibria.¹⁶ An equilibrium of each class exists for some cost level.

Proposition 4.6 (Persistent equilibria). *A persistent equilibrium always exists, but it is not necessarily unique. In any such equilibrium, agent a invests weakly more often.*

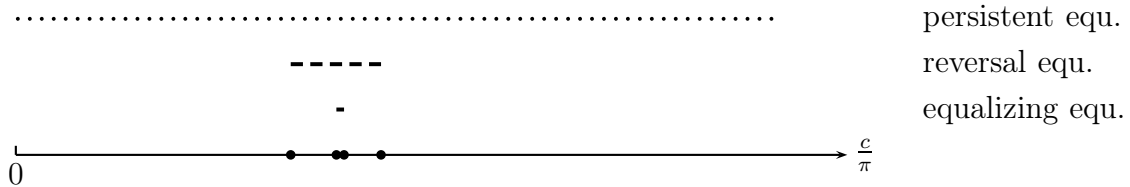


Figure 2: Equilibria across investment cost levels under breakthroughs

Convergence of prior beliefs. Consider the limit case $p_b \uparrow p_a$. Our next result shows that there always exists an equilibrium in which the two agents' payoffs and post-investment probabilities of being a high type converge as $p_b \uparrow p_a$. That is, the benchmark self-correcting property of proposition 3.2 continues to survive for any given level of investment cost in at least some equilibrium. This follows immediately from propositions 4.4 and 4.6.

Corollary 4.4 (Converging payoffs in persistent and equalizing equilibria). *Let $p_b \uparrow p_a$. For any $c > 0$, there exists at least one equilibrium in which the two agents' payoffs as well as their post-investment probabilities of being a high type converge.*

¹⁵By lemma 4.3, β_b is single-peaked in his own belief q_b because agent b is the post-investment favored agent.

¹⁶Parameters are $\mu_h = \lambda_h/r = 4, p_a = 1/3, p_b = 1/4, \pi = 1/5, \underline{p} = 1/10$.

Proof. From the proof of Proposition 4.6, for $c < \beta_i(\bar{p}_a, \bar{p}_a)$, there is a persistent equilibrium in which both agents invest; for $c > \beta_i(p_a, p_a)$, there is a persistent equilibrium in which neither agent invests. From proposition 4.4, for $c \in [\min_{q \in [p_a, \bar{p}_a]} \beta_i(q, q), \max_{q \in [p_a, \bar{p}_a]} \beta_i(q, q)]$, there is an equalizing equilibrium. The union of these parameter regions spans the real line. In any persistent equilibrium with symmetric investment and equalizing equilibria, the agents' payoffs converge as $p_b \rightarrow p_a$. ■

In any persistent equilibrium with symmetric investment, as $p_b \uparrow p_a$ post-investment beliefs converge and therefore continuation payoffs from interacting with the principal converge. Symmetric investment costs the same to both agents. On the other hand in any equalizing equilibrium posterior beliefs are equal, i.e., $q_a = q_b$, and the probability of investment across agents converges as $p_b \uparrow p_a$. Hence, the cost of investment becomes equal across agents.

Taken together, corollaries 4.2 and 4.4 highlight a difference in the equilibrium payoff set between breakthrough and breakdown environments. In both environments, if $p_a = p_b$ then for any cost $c > 0$ there exists an equilibrium in which the two agents obtain the same payoff. In the breakthrough environment, for any cost $c > 0$ there exists an equilibrium in which the two agents' payoffs converge as $p_b \uparrow p_a$. In contrast, in the breakdown environment, the two agents' payoffs diverge in any equilibrium for all cost levels except for a small cost region.

A No-investment benchmark with general $(\lambda_h, \lambda_\ell)$

Proposition A.1 (Self-correcting property of inconclusive breakthroughs). *For any $\lambda_h > \lambda_\ell$, the two agents' payoffs converge as $p_a \downarrow p_b$.*

Proof. Let $U_i(p_a, p_b)$ be agent i 's payoff given the belief pair (p_a, p_b) . For any $p_a > p_b$, the principal first uses agent a for a period of length t^* . If no breakthrough occurs in $[0, t^*)$, the principal's belief toward agent a drops to p_b . Let $f(s)$ for $s \in [0, t^*)$ be the density of the random arrival time of the first breakthrough from agent a . We let $p_a(s)$ be the belief that $\theta_a = h$ if there is no breakthrough up to time s , and let $j(p_a(s))$ be the belief that $\theta_a = h$ right after the first breakthrough at time s . Agent a 's payoff is given by

$$\int_0^{t^*} f(s) \{1 - e^{-rs} + e^{-rs} U_a(j(p_a(s)), p_b)\} ds + \left(1 - \int_0^{t^*} f(s) ds\right) \{1 - e^{-rt^*} + e^{-rt^*} U_a(p_b, p_b)\}.$$

Agent b 's payoff is given by

$$\int_0^{t^*} f(s) e^{-rs} U_b(j(p_a(s)), p_b) ds + \left(1 - \int_0^{t^*} f(s) ds\right) e^{-rt^*} U_b(p_b, p_b).$$

As $p_a \downarrow p_b$, t^* converges to zero. Both players' payoffs converge to $U_a(p_b, p_b) = U_b(p_b, p_b)$. ■

Proposition A.2 (Spiraling property of inconclusive breakdowns). *For any $\lambda_h < \lambda_\ell$, the two agents' payoffs do not converge if $r^2 - (1 - 2p_a)r(\lambda_\ell - \lambda_h) - \lambda_h\lambda_\ell > 0$ or equivalently:*

$$\frac{\lambda_h}{\lambda_h + r} p_a + \frac{\lambda_\ell}{\lambda_\ell + r} (1 - p_a) < \frac{1}{2}.$$

Proof. Let $U_i(p_a, p_b)$ be agent i 's payoff given the belief pair (p_a, p_b) . We let $p_a(s)$ be the belief toward agent a if there is no breakdown up to time s , and let $j(p_a(s))$ be the belief toward him right after the first breakdown at time s .

For any $p_a > p_b$, the principal begins with agent a , and uses agent a exclusively if no breakdown occurs. We let $f(s) = p_a \lambda_h e^{-\lambda_h s} + (1 - p_a) \lambda_\ell e^{-\lambda_\ell s}$ be the density of the arrival time $s \in [0, \infty)$ of the first breakdown from agent a . Correspondingly, $1 - F(t)$ is the probability that no breakdown occurs before time t . For any $t > 0$, we can write agent a 's payoff as

follows:

$$\int_0^t f(s) \{1 - e^{-rs} + e^{-rs} U_a(j(p_a(s)), p_b)\} ds + (1 - F(t)) (1 - e^{-rt} + e^{-rt} U_a(p_a(t), p_b)).$$

We can write agent b 's payoff as follows:

$$\int_0^t f(s) \{e^{-rs} U_b(j(p_a(s)), p_b)\} ds + (1 - F(t)) e^{-rt} U_a(p_a(t), p_b).$$

The payoff difference between a and b is:

$$\begin{aligned} \int_0^t f(s) \{1 - e^{-rs} + e^{-rs} (U_a(j(p_a(s)), p_b) - U_b(j(p_a(s)), p_b))\} ds \\ + (1 - F(t)) (1 - e^{-rt} + e^{-rt} (U_a(p_a(t), p_b) - U_b(p_a(t), p_b))). \end{aligned}$$

We claim that $U_a(j(p_a(s)), p_b) - U_b(j(p_a(s)), p_b) \geq -1$ since $U_i(p_a, p_b)$ is in the range $[0, 1]$ for any i, p_a, p_b . Also, $U_a(p_a(t), p_b) - U_b(p_a(t), p_b) \geq 0$ since $p_a(t) > p_b$ for any t . Therefore, the payoff difference is at least:

$$\begin{aligned} G(t) &:= \int_0^t f(s) (1 - 2e^{-rs}) ds + (1 - F(t)) (1 - e^{-rt}) \\ &= \frac{p_a(\lambda_h - r)e^{t(-\lambda_h - r)}}{\lambda_h + r} - \frac{2\lambda_h p_a}{\lambda_h + r} + \frac{(1 - p_a)(\lambda_\ell - r)e^{t(-\lambda_\ell - r)}}{\lambda_\ell + r} - \frac{2\lambda_\ell(1 - p_a)}{\lambda_\ell + r} + 1. \end{aligned}$$

Since we can choose any t , the payoff difference is at least $\sup_t G(t)$. Note that $G(0) = 0$, and $G(\infty) > 0$ if and only if $r^2 - (1 - 2p_a)r(\lambda_\ell - \lambda_h) - \lambda_h \lambda_\ell > 0$. \blacksquare

B Proofs for section 4

An agent's benefit from investment depends only on the posterior belief (q_a, q_b) and not on the prior belief (p_a, p_b) , since the posterior belief pair pins down the principal's allocation. Recall that $\beta_i(q_a, q_b)$ denote agent i 's benefit from investment given (q_a, q_b) . The low type of agent i invests if and only if this benefit is greater than the cost c . We let $\alpha_i \in [0, 1]$ denote the probability that a low-type agent i invests.

Let $U_i(\theta_i^+; q_a, q_b)$ denote agent i 's post-investment payoff given his post-investment type

θ_i^+ and belief pair (q_a, q_b) . Agent i 's benefit from investment is given by:

$$\beta_i(q_a, q_b) = \pi (U_i(h; q_a, q_b) - U_i(\ell; q_a, q_b)).$$

B.1 Conclusive breakdowns

B.1.1 A (low-type) agent's benefit from investment

We begin with the case that $q_a > q_b$ (i.e., agent a is post-investment favored). Then,

$$U_a(\theta_a^+; q_a, q_b) = \begin{cases} 1 & \text{if } \theta_a^+ = h \\ \frac{1}{\mu_\ell + 1} & \text{if } \theta_a^+ = \ell \end{cases}, \quad U_b(\theta_b^+; q_a, q_b) = \begin{cases} \frac{\mu_\ell(1-q_a)}{\mu_\ell + 1} & \text{if } \theta_b^+ = h \\ \frac{\mu_\ell(1-q_a)}{(\mu_\ell + 1)^2} & \text{if } \theta_b^+ = \ell \end{cases}$$

where $\mu_\ell := \lambda_\ell/r$. It follows that:

$$\beta_a(q_a, q_b) = \pi \frac{\mu_\ell}{\mu_\ell + 1},$$

$$\beta_b(q_a, q_b) = \pi \frac{\mu_\ell^2(1-q_a)}{(\mu_\ell + 1)^2}.$$

If $q_b > q_a$ (i.e., agent b is post-investment favored), then $\beta_a(q_a, q_b) = \pi \frac{\mu_\ell^2(1-q_b)}{(\mu_\ell + 1)^2}$ and $\beta_b(q_a, q_b) = \pi \frac{\mu_\ell}{\mu_\ell + 1}$.

B.1.2 Proofs for equilibrium characterization

Proof for proposition 4.1. Let $q_a = q_b = q$ be the post-investment belief. Agent b must be investing, for otherwise $q_b = p_b < p_a \leq q_a$ which contradicts $q_a = q_b$. So agent b 's benefit from investment exceeds the cost:

$$\pi \frac{\mu_\ell(1 + \mu_\ell(2 - q))}{2(\mu_\ell + 1)^2} \geq c.$$

If this inequality is strict, both agents have a strict incentive to invest which contradicts the presumption that $q_a = q_b$ and $p_a > p_b$. Therefore, it holds with equality. The post-investment

belief q is given by:

$$q = \frac{\mu_\ell \pi (2\mu_\ell + 1) - 2c(\mu_\ell + 1)^2}{\mu_\ell^2 \pi}.$$

On the other hand, $q = p_a + (1 - p_a)\alpha_a\pi = p_b + (1 - p_b)\alpha_b\pi$. There exist $\alpha_a, \alpha_b \in [0, 1]$ such that both equalities hold if and only if $p_a \leq q \leq \bar{p}_b$. Substituting the value of q into the inequality above, we obtain condition (5). \blacksquare

Proof for proposition 4.2. There exist $\alpha_a, \alpha_b \in [0, 1]$ such that $q_b > q_a$ if and only if $\bar{p}_b > p_a$. Since $p_b < p_a$ and $q_b > q_a$, it must be true that $\alpha_b > \alpha_a$. Hence, agent b 's benefit from investing must be weakly higher than the cost, while agent a 's benefit must be weakly lower:

$$\pi \frac{\mu_\ell}{\mu_\ell + 1} \geq c \geq \pi \frac{\mu_\ell^2 (1 - q_b)}{(\mu_\ell + 1)^2}.$$

Since $q_b \leq \bar{p}_b$, a necessary condition for the inequality above is:

$$\pi \frac{\mu_\ell}{\mu_\ell + 1} \geq c \geq \pi \frac{\mu_\ell^2 (1 - \bar{p}_b)}{(\mu_\ell + 1)^2}.$$

Therefore, condition (6) is necessary for a reversal equilibrium to exist. If (6) holds, $\alpha_b = 1$ and $\alpha_a = 0$ sustain a reversal equilibrium, so the condition is also sufficient. If $\pi \frac{\mu_\ell}{\mu_\ell + 1} > c > \pi \frac{\mu_\ell^2 (1 - \bar{p}_b)}{(\mu_\ell + 1)^2}$, any reversal equilibrium must feature $\alpha_b = 1$ and $\alpha_a = 0$, so generically there is a unique reversal equilibrium.

A pre-investment high-type agent b obtains payoff 1. A pre-investment high-type agent a obtains payoff $\frac{\mu_\ell(1 - \bar{p}_b)}{\mu_\ell + 1}$. A pre-investment low-type agent b obtains $\frac{1}{\mu_\ell + 1}$ if he deviates to no investment. Therefore, a pre-investment low-type agent b must obtain a weakly higher payoff if he invests. A pre-investment low-type agent a obtains

$$\frac{\mu_\ell(1 - \bar{p}_b)}{(\mu_\ell + 1)^2}.$$

Hence, type by type, agent a 's pre-investment payoff is at most $\frac{\mu_\ell(1 - \bar{p}_b)}{\mu_\ell + 1}$ times agent b 's pre-investment payoff. \blacksquare

Proof for proposition 4.3. Given that $q_a > q_b$, agent a 's benefit from investment is $\pi \frac{\mu_\ell}{\mu_\ell + 1}$. In any persistent equilibrium, agent a invests if $c < \pi \frac{\mu_\ell}{\mu_\ell + 1}$ and doesn't invest if $c > \pi \frac{\mu_\ell}{\mu_\ell + 1}$. The

rest of the equilibrium characterization follows immediately from

$$\beta_a(q_a, q_b) = \pi \frac{\mu_\ell}{\mu_\ell + 1}, \quad \beta_b(q_a, q_b) = \pi \frac{\mu_\ell}{\mu_\ell + 1} (1 - q_a)$$

derived in section B.1.1. Part (i) then follows immediately from (ii).

For all three types of equilibria, the payoff ratio between a pre-investment high-type agent b and a pre-investment high-type agent a equals $\frac{\mu_\ell}{\mu_\ell + 1}(1 - x)$, where x is the post-investment belief that agent a is a high type. If $c > \pi \frac{\mu_\ell}{\mu_\ell + 1}$ the payoff ratio between the pre-investment low types is $\frac{\mu_\ell}{\mu_\ell + 1}(1 - p_a)$. If $\pi \frac{\mu_\ell}{\mu_\ell + 1} > c > \pi \frac{\mu_\ell^2(1 - \bar{p}_a)}{(\mu_\ell + 1)^2}$, the payoff ratio between the pre-investment low types is $\frac{\mu_\ell}{\mu_\ell + 1}(1 - \bar{p}_a)$ if a pre-investment low-type agent a deviates to no investment. Hence, the payoff ratio must be lower when a pre-investment low-type agent a invests. If $c < \pi \frac{\mu_\ell^2(1 - \bar{p}_a)}{(\mu_\ell + 1)^2}$, the payoff ratio between the pre-investment low types is $\frac{\mu_\ell}{\mu_\ell + 1}(1 - \bar{p}_a)$ if neither invests. Once we subtract the cost from both the numerator and the denominator, the payoff ratio must be smaller. ■

B.1.3 Additional results

In a reversal equilibrium, agent b is the first to be allocated the task. The payoff of each type of agent b is higher in a reversal equilibrium than in the equilibrium of the no-investment benchmark. The high type's payoff is higher because he is allocated the task forever. The low type's payoff is higher as well. Moreover, for a fixed type agent b 's expected payoff is strictly higher than that of agent a in a reversal equilibrium. In fact, the payoff gap widens as p_b approaches p_a because the expected payoff of agent b remains fixed whereas that of agent a goes down as he faces a stronger opponent.

Lemma B.1. *For any $p_b < p_a$, both types of agent a prefer the no-investment benchmark to the equalizing equilibrium and the reversal equilibrium when investment is available.*

Proof for lemma B.1. We show that both types prefer the no-investment benchmark to the equalizing equilibrium. The proof for the reversal equilibrium is similar and thus omitted. In the no-investment benchmark, a pre-investment high type's payoff of the favored agent is 1 and a pre-investment low type's payoff is $1/(\mu_\ell + 1)$. In the equalizing equilibrium with posterior belief q , a pre-investment high type's payoff is $\frac{1}{2} \left(\frac{\mu_\ell}{\mu_\ell + 1}(1 - q) + 1 \right) < 1$. A pre-

investment low type is indifferent between investing and not, so his payoff is the same as the payoff with no investment: $\frac{1}{2(\mu_\ell+1)} \left(\frac{\mu_\ell}{(\mu_\ell+1)}(1-q) + 1 \right) < \frac{1}{\mu_\ell+1}$. ■

Lemma B.2 (Pre-investment inequality).

(i) *The ratio of the expected payoff of agent a to that of agent b in the persistent equilibrium with asymmetric investment and in the persistent equilibrium with $\alpha_a = \alpha_b = 1$ is strictly greater than that in the no-investment benchmark.*

(ii) *As $p_b \uparrow p_a$, the ratio of the expected payoff of agent b to that of agent a in the reversal equilibrium is strictly greater than the ratio of the payoff of agent a to that of agent b in the no-investment benchmark.*

Proof for lemma B.2.

(i) For the persistent equilibrium with $\alpha_a = \alpha_b = 1$, the expected cost of investment for agent i is given by $(1-p_i)c$. The expected payoff of agent a is strictly higher than that under the no-investment benchmark because

$$\bar{p}_a + (1 - \bar{p}_a) \frac{1}{\mu_\ell + 1} - (1 - p_a)c = \underbrace{\left(p_a + (1 - p_a) \frac{1}{\mu_\ell + 1} \right)}_{\text{no-investment benchmark payoff}} + (1 - p_a) \left(\pi \frac{\mu_\ell}{\mu_\ell + 1} - c \right)$$

and the second term in the RHS is strictly positive by proposition 4.3. By a similar reasoning, the expected payoff of agent b is strictly lower than in the no-investment benchmark because

$$\bar{p}_b \frac{\mu_\ell(1 - \bar{p}_a)}{\mu_\ell + 1} + (1 - \bar{p}_b) \frac{\mu_\ell(1 - \bar{p}_a)}{(\mu_\ell + 1)^2} - (1 - p_b)c$$

is strictly lower than the no-investment benchmark payoff for $c < \pi \frac{\mu_\ell^2(1 - \bar{p}_a)}{(\mu_\ell + 1)^2}$, the condition that sustains a persistent equilibrium with symmetric investment. Hence, the ratio of the payoff of agent a to that of agent b is greater than the no-investment payoff ratio.

In a persistent equilibrium with asymmetric investment, agent a has the same payoff as in that with symmetric investment discussed above, hence his payoff is again greater

than in the no-investment benchmark. The expected payoff of agent b on the other hand is just

$$p_b \frac{\mu_\ell(1 - \bar{p}_a)}{\mu_\ell + 1} + (1 - p_b) \frac{\mu_\ell(1 - \bar{p}_a)}{(\mu_\ell + 1)^2} < p_b \frac{\mu_\ell(1 - p_a)}{\mu_\ell + 1} + (1 - p_b) \frac{\mu_\ell(1 - p_a)}{(\mu_\ell + 1)^2},$$

where the RHS is his payoff from the no-investment benchmark. Hence, the ratio of the payoff of agent a to that of agent b is strictly greater than in the benchmark.

- (ii) The expected payoffs of agent b and agent a in a reversal equilibrium are respectively given by

$$p_b + (1 - p_b) \left(\pi + (1 - \pi) \frac{1}{\mu_\ell + 1} - c \right), \quad p_a \frac{\mu_\ell(1 - \bar{p}_b)}{\mu_\ell + 1} + (1 - p_a) \frac{\mu_\ell(1 - \bar{p}_b)}{(\mu_\ell + 1)^2}.$$

At $\bar{p}_b = \bar{p}_a$, agent b 's payoff is strictly greater. At this limit case, the ratio of agent b 's payoff to agent a 's payoff is

$$\frac{(\mu_\ell + 1)(\mu_\ell \bar{p}_a + 1 - c(1 - p_a)(\mu_\ell + 1))}{\mu_\ell(1 - p_a)(1 - \pi)(\mu_\ell p_a + 1)} > 1,$$

whereas in the no-investment benchmark the ratio of agent a 's payoff to agent b 's payoff is

$$\frac{\mu_\ell + 1}{\mu_\ell(1 - p_a)} > 1.$$

The condition for which the ratio in the reversal equilibrium is greater than that in the no-investment equilibrium reduces to $c < \pi/(1 - p_a)$. The RHS is greater than π , whereas the upper bound on c that guarantees a reversal equilibrium is $\pi\mu_\ell/(\mu_\ell + 1) < \pi$. Hence, the condition holds for any such c . ■

Consider first the persistent equilibrium with asymmetric investment. Compared to the no-investment benchmark, agent b 's payoff decreases since he now faces a stronger, post-investment favored agent a . On the other hand, agent a 's equilibrium payoff is higher than in the no-investment benchmark because his low type finds it strictly optimal to invest. Without investment, he would obtain the no-investment benchmark payoff. Given that

agent a 's payoff increases and b 's payoff decreases, the payoff ratio is higher than in the no-investment benchmark. In the persistent equilibrium in which both agents invest, agent a is even more advantaged vis-à-vis agent b and the expected cost of investment for agent a is smaller than that for agent b . The two forces taken together lead to a higher payoff ratio.

In a reversal equilibrium, agent a has a lower payoff than what agent b obtains in the no-investment benchmark because agent a now faces a stronger opponent after agent b invests. Moreover, as $p_b \uparrow p_a$, the post-investment favored agent b has the same payoff as agent a does in the no-investment benchmark even if he deviates to no investment. Hence, agent b 's equilibrium payoff is higher than that of the pre-investment favored agent a . As $p_b \uparrow p_a$, the payoff ratio is higher than in the benchmark as well.

B.2 Conclusive breakthroughs

B.2.1 A (low-type) agent's benefit from investment

Suppose $q_a > q_b$. We first calculate $\beta_a(q_a, q_b)$ and $\beta_b(q_a, q_b)$: agent a 's and agent b 's benefit from investment respectively. Recall that $U_i(\theta_i^+; q_a, q_b)$ denotes agent i 's post-investment payoff given his type θ_i^+ and the principal's post-investment belief (q_a, q_b) .

The principal uses agent a exclusively for a period of length $t^* = \frac{1}{\lambda_h} \log \frac{q_a(1-q_b)}{(1-q_a)q_b}$ and then splits the task equally among the two agents for a subsequent period of length $t_s := \frac{2}{\lambda_h} \log \frac{q_b(1-p)}{(1-q_b)p}$. Agent i 's benefit from investment is given by

$$\beta_i(q_a, q_b) = \pi (U_i(h; q_a, q_b) - U_i(\ell; q_a, q_b)).$$

The post-investment payoff for each agent and type are:

$$\begin{aligned} U_a(h; q_a, q_b) &= 1 - e^{-\lambda_h t^*} + e^{-\lambda_h t^*} (1 - e^{-rt^*} + e^{-rt^*} S(h, q_b)), \\ U_a(\ell; q_a, q_b) &= 1 - e^{-rt^*} + e^{-rt^*} S(\ell, q_b), \\ U_b(h; q_a, q_b) &= e^{-rt^*} (1 - q_a + q_a e^{-\lambda_h t^*}) S(h, q_b), \\ U_b(\ell; q_a, q_b) &= e^{-rt^*} (1 - q_a + q_a e^{-\lambda_h t^*}) S(\ell, q_b). \end{aligned}$$

Here, $S(h, q_b)$ and $S(\ell, q_b)$ denote the payoffs to a high type and a low type, respectively,

if (i) his opponent is a high type with probability q_b ; (ii) the principal splits the task equally between the two agents until the belief drops to \underline{p} . To characterize $S(h, q_b)$ and $S(\ell, q_b)$, let t_1 be the arrival time of a high type's breakthrough and t_2 the arrival time of his opponent's breakthrough. For a low type, a breakthrough never arrives. In the absence of any breakthroughs, the principal experiments with the agents until the belief hits \underline{p} . The length of this experimentation period is given by t_s as defined above. The CDFs of t_1 and t_2 for $t_1, t_2 \leq t_s$ are:

$$F_1(t_1) = 1 - e^{-\frac{\lambda_h t_1}{2}}, \quad F_2(t_2) = q_b(1 - e^{-\frac{\lambda_h t_2}{2}}),$$

with corresponding density functions f_1 and f_2 respectively. Therefore,

$$\begin{aligned} S(h, q_b) &= \int_0^{t_s} f_1(t_1) \left(\int_0^{t_1} f_2(t_2) \frac{1 - e^{-rt_2}}{2} dt_2 + (1 - F_2(t_1)) \left(\frac{1 - e^{-rt_1}}{2} + e^{-rt_1} \right) \right) dt_1 \\ &\quad + (1 - F_1(t_s)) \left(\int_0^{t_s} f_2(t_2) \frac{1 - e^{-rt_2}}{2} dt_2 + (1 - F_2(t_s)) \frac{1 - e^{-rt_s}}{2} \right), \\ S(\ell, q_b) &= \int_0^{t_s} f_2(t_2) \frac{1 - e^{-rt_2}}{2} dt_2 + (1 - F_2(t_s)) \frac{1 - e^{-rt_s}}{2}. \end{aligned}$$

B.2.2 Proofs for equilibrium characterization

For the proofs of lemmata 4.3 and B.3, we use the following notation:

$$t^* := \frac{1}{\lambda_h} \log \frac{q_a(1 - q_b)}{(1 - q_a)q_b}, \quad t_s := \frac{2}{\lambda_h} \log \frac{q_b(1 - \underline{p})}{(1 - q_b)\underline{p}}, \quad \lambda_h/r =: \mu_h,$$

and apply the change of variables

$$\underline{p} = \frac{\underline{x}}{1 + \underline{x}}, \quad q_b = \frac{x_b}{1 + x_b}, \quad q_a = \frac{x_a}{1 + x_a}.$$

Given the assumption that $0 < \underline{p} < q_b < q_a$, it follows that $0 < \underline{x} < x_b < x_a$.

Lemma B.3.

- (i) Suppose $1 > q_a > q_b > \underline{p}$. Then, $\beta_a(q_a, q_b)$ either decreases or is single-peaked in q_b .
- (ii) Suppose $1 > q_a = q_b > \underline{p}$. Then, $\beta_i(q_b, q_b)$ either increases or is single-peaked in q_b .

Proof for lemma B.3.

- (i) We show that either $U_a(h; q_a, q_b) - U_a(\ell; q_a, q_b)$ decreases in q_b or it first increases and then decreases in q_b . The inequality $\frac{\partial(U_a(h; q_a, q_b) - U_a(\ell; q_a, q_b))}{\partial x_b} > 0$ is equivalent to:

$$(\mu_h + 2)\underline{x}^2 + \left(\frac{x_b}{\underline{x}}\right)^{2/\mu_h} (x_a(\mu_h + 2x_b + 2) - x_b(\mu_h(x_b + 2) + 2(x_b + 1))) + 2\underline{x}(\mu_h - x_a + 1) - (\mu_h + 2)x_a > 0.$$

The left-hand side equals zero when $x_b = \underline{x}$ and is strictly negative when $x_b = x_a$. It increases in x_b if and only if $(1 + \mu_h)x_b < x_a$. Hence, either the left-hand side is negative for all $x_b \in (\underline{x}, x_a)$ or it is first positive and then negative for $x_b \in (\underline{x}, x_a)$.

- (ii) We show that either $U_i(h; q_b, q_b) - U_i(\ell; q_b, q_b)$ is monotone in q_b or it first increases and then decreases in q_b . The inequality $\frac{\partial(U_i(h; q_b, q_b) - U_i(\ell; q_b, q_b))}{\partial x_b} > 0$ is equivalent to:

$$\underline{x}(\mu_h(\underline{x} + 2)(2x_b + 1) + 2(x_b + 1)(\underline{x} - x_b + 1)) - x_b((\mu_h + 2)x_b + 2) > 0.$$

The left-hand side is quadratic, and hence convex, in x_b . Moreover, it is positive when $x_b = \underline{x}$ and negative when $x_b = \infty$. Hence, in the domain that $q_b \in [p_a, \bar{p}_b]$, either $U_i(h; q_b, q_b) - U_i(\ell; q_b, q_b)$ is strictly increasing in q_b or it first increases and then decreases in q_b . ■

Proof for lemma 4.3.

- (i) We first show that $U_a(h; q_a, q_b) - U_a(\ell; q_a, q_b) > U_b(h; q_a, q_b) - U_b(\ell; q_a, q_b)$. This inequality follows directly once we apply the condition that $0 < \underline{x} < x_b < x_a$.
- (ii) Next, we show that $U_b(h; q_a, q_b) - U_b(\ell; q_a, q_b)$ decreases in q_a and increases in q_b by showing that it decreases in x_a and increases in x_b .

The inequality $\frac{\partial(U_b(h; q_a, q_b) - U_b(\ell; q_a, q_b))}{\partial x_a} < 0$ is equivalent to:

$$x_b \left(\mu_h(x_b + 1) \left(\frac{x_b}{\underline{x}}\right)^{2/\mu_h} + \mu_h + 2 \right) - \underline{x}(\mu_h(\underline{x} + 2) + 2(\underline{x} - x_b + 1)) > 0.$$

The left-hand side increases in x_b and equals to zero when $x_b = \underline{x}$. Hence, the inequality holds.

The inequality $\frac{\partial(U_b(h;q_a,q_b)-U_b(\ell;q_a,q_b))}{\partial x_b} > 0$ is equivalent to:

$$(\mu_h + 1)(\mu_h + 2)\underline{x}^2 + \mu_h x_b(\mu_h x_b + x_b + 1) \left(\frac{x_b}{\underline{x}}\right)^{2/\mu_h} + 2\underline{x}((\mu_h + 1)^2 - x_b) - (\mu_h + 2)x_b > 0.$$

The left-hand side is convex in x_b , increases in x_b when $x_b = \underline{x}$, and is positive when $x_b = \underline{x}$. Hence, the inequality holds.

(iii) Next, we show that $U_a(h; q_a, q_b) - U_a(\ell; q_a, q_b)$ is single-peaked in q_a by showing that it first increases in x_a and then decreases in x_a . The inequality $\frac{\partial(U_a(h;q_a,q_b)-U_a(\ell;q_a,q_b))}{\partial x_a} > 0$ is equivalent to:

$$\begin{aligned} \left(\frac{x_b}{\underline{x}}\right)^{2/\mu_h} & ((\mu_h + 1)x_b((\mu_h + 2)x_b + 2) - x_a(2(\mu_h + 1)x_b + \mu_h + 2)) + 2\underline{x}((\mu_h + 1)^2 - x_a) \\ & + (\mu_h + 1)(\mu_h + 2)\underline{x}^2 - (\mu_h + 2)x_a > 0 \end{aligned}$$

The left-hand side is linear and decreases in x_a , so the inequality holds only if x_a is small enough. When $x_a = x_b$, the inequality is equivalent to

$$(\mu_h + 1)(\mu_h + 2)\underline{x}^2 + \mu_h x_b(\mu_h x_b + x_b + 1) \left(\frac{x_b}{\underline{x}}\right)^{2/\mu_h} + 2\underline{x}((\mu_h + 1)^2 - x_b) - (\mu_h + 2)x_b > 0.$$

The left-hand side is convex in x_b , increases in x_b when $x_b = \underline{x}$, and is positive when $x_b = \underline{x}$. Hence, the inequality $\frac{\partial(U_a(h;q_a,q_b)-U_a(\ell;q_a,q_b))}{\partial x_a} > 0$ holds when x_a equals x_b . ■

Proof for proposition 4.4. Let $q_a = q_b = q$ be the post-investment belief. It must be that $p_a \leq \bar{p}_b$ and $\alpha_b > 0$. Based on B.2.1, it is immediate that

$$\beta_a(q, q) = \beta_b(q, q) = \pi \frac{\mu_h - \frac{\left(\frac{(1-p)q}{\underline{p}(1-q)}\right)^{-\frac{\mu_h+2}{\mu_h}} (\underline{p}(\mu_h(\underline{p}-2)-2) + (\mu_h+2)q)}{(\underline{p}-1)\underline{p}}}{2(\mu_h + 2)}.$$

Note that $\beta_i(\underline{p}, \underline{p}) = 0$ and $\beta_i(1, 1) = \frac{\mu_h}{2\mu_h+4}$. Moreover, $\beta_i(q, q)$ is single peaked in $q \in (\underline{p}, 1)$.

In equilibrium, $\beta_i(q, q) = c$. Otherwise, both agents either strictly prefer to invest or not to invest, which contradicts $p_a > p_b$ and $q_a = q_b$. On the other hand, $q = p_a + (1 - p_a)\alpha_a\pi = p_b + (1 - p_b)\alpha_b\pi$. There exist $\alpha_a, \alpha_b \in [0, 1]$ such that both equalities hold if and only if $p_a \leq q \leq \bar{p}_b$. Hence, there is an equalizing equilibrium if and only if (7) holds. ■

Proof for proposition 4.5. A reversal equilibrium is possible only if $\bar{p}_b > p_a$. Since $p_b < p_a$ and $q_b > q_a$, it must be that $\alpha_b > \alpha_a$. Let $\tilde{p} := \arg \max_{p \in (p_a, 1)} \beta_b(p_a, p)$. There are three subcases to consider:

- (i) $\alpha_b = 1$ and $\alpha_a \in (0, 1)$. Let $\bar{\alpha}_a$ be the solution to $\bar{p}_b = p_a + (1 - p_a)\pi\bar{\alpha}_a$. For $q_b > q_a$ to hold, it must be that $\alpha_a \in (0, \bar{\alpha}_a)$ and α_a solves $c = \beta_a(p_a + (1 - p_a)\pi\alpha_a, \bar{p}_b)$. Such an equilibrium exists if and only if $\beta_a(p_a, \bar{p}_b) < c < \beta_a(\bar{p}_b, \bar{p}_b)$.
- (ii) $\alpha_b = 1$ and $\alpha_a = 0$. This is an equilibrium if and only if $\beta_a(p_a, \bar{p}_b) \leq c \leq \beta_b(p_a, \bar{p}_b)$.
- (iii) $\alpha_b \in (0, 1)$ and $\alpha_a = 0$. Let $\underline{\alpha}_b$ be the solution to $p_b + (1 - p_b)\pi\underline{\alpha}_b = p_a$. Hence, α_b must be in $(\underline{\alpha}_b, 1)$.
 - (a) If $\bar{p}_b \leq \tilde{p}$, then this equilibrium exists iff $\beta_b(p_a, p_a) < c < \beta_b(p_a, \bar{p}_b)$.
 - (b) Suppose that $\bar{p}_b > \tilde{p}$. Based on Lemma 4.3, the minimal and maximal $\beta_b(p_a, x)$ for $x \in [p_a, \bar{p}_b]$ are $\min\{\beta_b(p_a, p_a), \beta_b(p_a, \bar{p}_b)\}$ and $\beta_b(p_a, \tilde{p})$, respectively. This equilibrium exists if and only if $\min\{\beta_b(p_a, p_a), \beta_b(p_a, \bar{p}_b)\} < c \leq \beta_b(p_a, \tilde{p})$.

We first observe that $\beta_a(p_a, \bar{p}_b) < \beta_a(p_a, p_a) = \beta_b(p_a, p_a)$, where the inequality follows from $\beta_a(q_a, q_b)$ being strictly decreasing in q_b for any $q_a < q_b$. From (ii) and (iii), a reversal equilibrium exists if $\beta_a(p_a, \bar{p}_b) \leq c \leq \beta_b(p_a, \min\{\bar{p}_b, \tilde{p}\})$. Hence, combining this with (i), a reversal equilibrium exists if and only if $\beta_a(p_a, \bar{p}_b) \leq c \leq \max\{\beta_a(\bar{p}_b, \bar{p}_b), \beta_b(p_a, \min\{\bar{p}_b, \tilde{p}\})\}$. Since $\beta_b(p_a, \min\{\bar{p}_b, \tilde{p}\}) = \beta_b(p_a, \hat{p})$, we obtain (8). ■

Proof for proposition 4.6. In a persistent equilibrium, agent a 's benefit from investing is strictly higher than agent b 's, so either $\alpha_b = 0$ or $\alpha_a = 1$. First, consider three subcases in which $\alpha_a = 1$:

- (i) $\alpha_b = 1$. This equilibrium exists if and only if $c \leq \beta_b(\bar{p}_a, \bar{p}_b)$. Since $\beta_a(\bar{p}_a, \bar{p}_b) > \beta_b(\bar{p}_a, \bar{p}_b)$, agent a 's unique best response is indeed $\alpha_a = 1$.
- (ii) $\alpha_b \in (0, 1)$. This equilibrium exists if and only if $\beta_b(\bar{p}_a, p_b) < c < \beta_b(\bar{p}_a, \bar{p}_b)$. Because $\beta_b(q_a, q_b)$ is strictly increasing in q_b , there exists a unique $\alpha_b \in (0, 1)$ such that $c = \beta_b(\bar{p}_a, p_b + (1 - p_b)\pi\alpha_b)$. Since agent a 's benefit is strictly higher, his unique best response is indeed $\alpha_a = 1$.
- (iii) $\alpha_b = 0$. This equilibrium exists if and only if $\beta_b(\bar{p}_a, p_b) \leq c \leq \beta_a(\bar{p}_a, p_b)$.

Next, we consider two subcases in which $\alpha_b = 0$:

- (iv) $\alpha_a = 0$. This equilibrium exists if and only if $\beta_a(p_a, p_b) \leq c$.
- (v) $\alpha_a \in [0, 1]$ and agent a is indifferent between investing and not. This equilibrium exists if and only if $\min_{q \in [p_a, \bar{p}_a]} \beta_a(q, p_b) \leq c \leq \max_{q \in [p_a, \bar{p}_a]} \beta_a(q, p_b)$.

By (iv) and (v), a persistent equilibrium exists for $c \geq \min \{\beta_a(p_a, p_b), \beta_a(\bar{p}_a, p_b)\}$. Combining this with (iii), a persistent equilibrium exists if $c \geq \beta_b(\bar{p}_a, p_b)$. Given (i) and $\beta_b(\bar{p}_a, p_b) \leq \beta_b(\bar{p}_a, \bar{p}_b)$, a persistent equilibrium always exists. ■

C Arbitrary number of agents, groups, and tasks

Suppose there are $K \geq 2$ social groups and n_i agents from group i , where $i = 1, \dots, K$. The frequency of high types in group i is $p_i \in (0, 1)$. Groups are indexed in decreasing order of average productivity:

$$1 > p_1 > p_2 > \dots > p_K > 0.$$

Agents are iid draws from their respective groups. We refer to each agent of group i as agent i . Without loss, all groups are assumed to be ex-ante preferred to the safe arm. An agent can complete at most one task per unit of time. The principal has $M \geq 1$ identical and perfectly divisible tasks available for assignment.

Assumption 2 (Relative scarcity of tasks). $M < n_1$.

Assumption 3 (Equal group size). $n_i = n$ for all $i = 1, \dots, K$.

Proposition C.1. *Suppose assumptions 2 and 3 hold.*

- (i) *(Spiraling property of breakdowns) Fix $i < j$. As $p_j \uparrow p_i$, the ratio of the expected payoff of agent i to that of agent j is bounded away from one.*
- (ii) *(Self-correcting property of breakthroughs) For any $i < j$, the expected payoff of a agent i converges to that of agent j as $p_j \uparrow p_i$.*

Proof.

- (i) The principal allocates M/n tasks to each agent 1 for as long as no breakdowns are observed from him. After $x \leq n$ breakdowns are realized from group 1, the principal allocates $M/(n - x)$ tasks to agents 1 who have not generated a breakdown yet. After all agents 1 have generated a breakdown, principal starts splitting the task equally among all agents 2. For as long as at least one agent 2 has not generated a breakdown yet, principal splits M equally among agents 2 who have not generated a breakdown yet. This assignment rule subsequently continues with all groups $i = 3, \dots, K$ until all agents have generated breakdowns.

Let $U(p)$ denote the expected payoff of an agent with likelihood p of being a high type when the principal starts with his group. Per this definition, the expected payoff of an agent 1 is $U(p_1)$. The expected payoff of an agent $i \geq 1$ is:

$$\prod_{k=1}^{i-1} (1 - p_k)^n D_i U(p_i)$$

where D_i is the expected discounted delay until it is the turn of agents i to be allocated tasks. Because agents i 's turn is earlier than that of agents j , $D_j < D_i$. Therefore, the ratio of the expected payoff of i to that of j is

$$\frac{1}{\prod_{k=i}^{j-1} (1 - p_k)^n} \frac{D_i U(p_i)}{D_j U(p_j)}.$$

As $p_j \uparrow p_i$, $U(p_j) \rightarrow U(p_i)$. But D_j does not converge to D_i : the expected delay for agents j is independent of p_j for as long as $p_j < p_i$. Therefore, the ratio of the expected

payoff of i to that of j as $p_j \uparrow p_i$ converges to

$$\prod_{k=i}^{j-1} (1 - p_k)^{-n} \frac{D_i}{D_j} > 1.$$

- (ii) The tasks are initially split equally among agents 1 (i.e. each is assigned M/n tasks) until

$$t_1^0 = \frac{n}{M\lambda} \log \left(\frac{\phi(p_1)}{\phi(p_2)} \right)$$

in the absence of any breakthroughs from agents 1. If a breakthrough occurs at time $t \leq t_1^0$, the principal assigns one task to the agent 1 that generated the breakthrough forever after; she then splits $M - 1$ tasks equally among the rest of agents 1 and assigns new deadline

$$t_1^1 = \frac{n - 1}{(M - 1)\lambda} \log \left(\frac{\phi(p_{1,t})}{\phi(p_2)} \right).$$

If $m < M$ agents 1 generate breakthroughs and $n - m$ others do not generate breakthroughs within the most recent deadline, principal splits equally $M - m$ remaining tasks to agents 1 who did not generate a breakthrough yet and all agents 2, i.e. for each $(M - m)/(2n - m)$. The deadline for agents 2 when zero breakthroughs have occurred among them is $t_2^0 = \frac{2n - m}{(M - m)\lambda} \log \left(\frac{\phi(p_2)}{\phi(p_3)} \right)$. Principal proceeds optimally in this fashion until either (i) all tasks are allocated forever to agents who generated breakthroughs, or (ii) all groups are tried but deadlines are over before all tasks are assigned forever.

Fix group i . Upon reaching group i , suppose m agents from groups $1, \dots, i - 1$ have not generated breakthroughs yet. Therefore,

$$t_i^0 = \frac{m + n}{(M - n(i - 1))\lambda} \log \left(\frac{\phi(p_i)}{\phi(p_{i+1})} \right)$$

As $p_{i+1} \uparrow p_i$, $t_i^0 \rightarrow 0$. Therefore the expected payoffs of agents i and $i + 1$ converge. By a similar reasoning, if $p_j \uparrow p_i$ for some $j > i$ then the expected payoffs of agents i and j converge. ■

References

- Aigner, Dennis J., and Glen G. Cain.** 1977. “Statistical Theories of Discrimination in Labor Markets.” *Industrial and Labor Relations Review*, 30(2): 175–187. 6
- Akerlof, George, and Rachel Kranton.** 2000. “Economics and Identity.” *The Quarterly Journal of Economics*, 115(3): 715–753.
- Altonji, Joseph G.** 2005. “Employer Learning, Statistical Discrimination and Occupational Attainment.” *American Economic Review: Papers and Proceedings*, 95(2): 112–117. 9
- Altonji, Joseph G., and Charles R. Pierret.** 2001. “Employer Learning and Statistical Discrimination.” *The Quarterly Journal of Economics*, 116(1): 313–350. 9
- Antonovics, Kate, and Limor Golan.** 2012. “Experimentation and job choice.” *Journal of Labor Economics*, 30(2): 333–366. 9
- Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo.** 2010. “Beyond Signaling and Human Capital: Education and the Revelation of Ability.” *American Economic Journal: Applied Economics*, 2(4): 76–104.
- Arrow, Kenneth.** 1973. “The Theory of Discrimination.” In *Discrimination in Labor Markets*, ed. O. Ashenfelter and A. Rees, 3–33. Princeton University Press. 6
- Baron, James N., and David M. Kreps.** 1999. *Strategic Human Services: Frameworks for General Managers*. New York:Wiley. 10
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova.** 2009. “Powerful Women: Does Exposure Reduce Bias?” *The Quarterly Journal of Economics*, 124(4): 1497–1540.
- Bergemann, Dirk, and Juuso Valimaki.** 1996. “Learning and Strategic Pricing.” *Econometrica*, 64(5): 1125–1149. 8
- Bertrand, Marianne, and Esther Duflo.** 2017. “Handbook of Field Experiments.” , ed. Esther Duflo and Abhijit Banerjee Vol. 1, Chapter Field Experiments on Discrimination. Elsevier. 2, 7

- Bertrand, Marianne, and Sendhil Mullainathan.** 2004. “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review*, 94(4): 991–1013. 7
- Blank, Rebecca.** 2005. “Tracing the Economic Impact of Cumulative Discrimination.” *American Economic Review: Papers and Proceedings*, 95(2): 99–103. 7, 8
- Blank, Rebecca, Marilyn Dabady, and Connie Citro.** 2004. “Measuring racial discrimination.” National Research Council The National Academies Press, Washington DC. 7
- Blume, Lawrence E.** 2006. “The Dynamics of Statistical Discrimination.” *The Economic Journal*, 116: F480–F498. 6
- Bose, Gautam, and Kevin Lang.** 2017. “Monitoring for Worker Quality.” *Journal of Labor Economics*, 35(3): 755–785. 10
- Carlsson, Magnus, and Stefan Eriksson.** 2019. “Age discrimination in hiring decisions: Evidence from a field experiment in the labor market.” *Labour Economics*, Forthcoming. 2
- Coate, Stephen, and Glenn C. Loury.** 1993. “Will Affirmative-Action Policies Eliminate Negative Stereotypes?” *American Economic Review*, 83(5): 1220–1240. 6, 7, 20, 26
- Cornell, Bradford, and Ivo Welch.** 1996. “Culture, Information, and Screening Discrimination.” *Journal of Political Economy*, 104(3): 542–571. 6, 7
- Deb, Rahul, Matthew Mitchell, and Mallesh M. Pai.** 2019. “Our Distrust is Very Expensive.” *working paper*. 8
- Fang, Hanming.** 2001. “Social Culture and Economic Performance.” *American Economic Review*, 91(4): 924–937.
- Fang, Hanming, and Andrea Moro.** 2011. “Chapter 5 - Theories of Statistical Discrimination and Affirmative Action: A Survey.” In . Vol. 1 of *Handbook of Social Economics*, 133 – 200. North-Holland. 6

- Farber, Henry S., and Robert Gibbons.** 1996. “Learning and Wage Dynamics.” *The Quarterly Journal of Economics*, 111(4): 1007–1047. 9
- Felli, Leonardo, and Christopher Harris.** 1996. “Learning, Wage Dynamics, and Firm-Specific Human Capital.” *Journal of Political Economy*, 104(4): 838–868. 8
- Foster, Dean P., and Rakesh V. Vohra.** 1992. “An Economic Argument for Affirmative Action.” *Rationality and Society*, 4(2): 176–188. 6, 7
- Fryer, Roland G.** 2007. “Belief flipping in a dynamic model of statistical discrimination.” *Journal of Public Economics*, 91: 1151–1166.
- Ghosh, Arpita, and Patrick Hummel.** 2012. “Learning and Incentives in User-Generated Content: Multi-Armed Bandits with Endogenous Arms.” 9
- Habermalz, Steffen.** 2011. “The speed of employer learning and job market signalling revisited.” *Applied Economic Letters*, 18: 607–610.
- Jacobs, David.** 1981. “Toward a Theory of Mobility and Behavior in Organizations: An Inquiry into the Consequences of Some Relationships Between Individual Performance and Organizational Success.” *American Journal of Sociology*, 87(3): 684–707. 5, 10
- Jovanovic, Boyan.** 1979. “Job Matching and the Theory of Turnover.” *Journal of Political Economy*, 87(5): 972–990. 8
- Kamenica, Emir, and Matthew Gentzkow.** 2011. “Bayesian Persuasion.” *American Economic Review*, 101(6): 2590–2615.
- Keller, Godfrey, and Sven Rady.** 2010. “Strategic experimentation with Poisson bandits.” *Theoretical Economics*, 5: 275–311. 9
- Keller, Godfrey, and Sven Rady.** 2015. “Breakdowns.” *Theoretical Economics*, 10(1): 175–202. 9
- Keller, Godfrey, Sven Rady, and Martin Cripps.** 2005. “Strategic Experimentation with Exponential Bandits.” *Econometrica*, 73(1): 39–68. 9

- Kim, Young-Chul, and Glenn C. Loury.** 2018. “Collective Reputation and the Dynamics of Statistical Discrimination.” *International Economic Review*, 59(1): 1–16. 6
- Kremer, Ilan, Yishay Mansour, and Motty Perry.** 2014. “Implementing the “Wisdom of the Crowd”.” *Journal of Political Economy*, 122(5): 988–1012.
- Lange, Fabian.** 2007. “The Speed of Employer Learning.” *Journal of Labor Economics*, 25(1): 1–35. 9
- Lang, Kevin, and Jee-Yeon K. Lehmann.** 2012. “Racial Discrimination in the Labor Market: Theory and Empirics.” *Journal of Economic Literature*, 50(4): 959–1006. 16
- Lang, Kevin, and Michael Manove.** 2011. “Education and Labor Market Discrimination.” *American Economic Review*, 101(4): 1467–96.
- Mansour, Hani.** 2012. “Does Employer Learning Vary by Occupation?” *Journal of Labor Economics*, 30(2): 415–444. 9
- Moro, Andrea, and Peter Norman.** 2004. “A general equilibrium model of statistical discrimination.” *Journal of Economic Theory*, 114: 1–30. 7
- Pager, Devah, and Diana Karafin.** 2009. “Bayesian Bigot? Statistical Discrimination, Stereotypes, and Employer Decision Making.” *Annals of the American Academy of Political and Social Sciences*, 621: 70–93. 2
- Peski, Marcin, and Balázs Szentes.** 2013. “Spontaneous Discrimination.” *American Economic Review*, 103(6): 2412–2436.
- Phelps, Edmund S.** 1972. “The Statistical Theory of Racism and Sexism.” *American Economic Review*, 62(4): 659–661. 6, 7
- Riach, Peter, and Judith Rich.** 2013. “Discrimination in the Labor Market.” In *The Oxford Handbook of Post-Keynesian Economics*. Vol. 1: Theory and Origins, , ed. G. C. Harcourt and Peter Kriesler. Oxford University Press.
- Strulovici, Bruno.** 2010. “Learning While Voting: Determinants of Collective Experimentation.” *Econometrica*, 78(3): 933–971. 9