IFN Working Paper No. 1454, 2023

Why Big Data Can Make Creative Destruction More Creative – But Less Destructive

Pehr-Johan Norbäck and Lars Persson

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Why Big Data can make Creative Destruction more Creative—but less Destructive

Pehr-Johan Norbäck*

Research Institute of Industrial Economics

Lars Persson

Research Institute of Industrial Economics, CEPR and CESifo

February 22, 2023

Abstract

The application of machine learning (ML) to big data has become increasingly important. We propose a model where firms have access to the same ML, but incumbents have access to historical data. We show that big data raises entrepreneurial barriers making the creative destruction process less destructive (less business-stealing) if the entrepreneur has weak access to the incumbent's data. It is also shown that this induces entrepreneurs to take on more risk and be more creative. Policies making data generally available may therefore be suboptimal. Supporting entrepreneurs' access to ML might be preferable since it stimulates creative entrepreneurship.

Keywords: *Machine Learning, Big Data, Creative Destruction; Entrepreneurship, Operational Data*

JEL classification: L1, L2, M13, O3

*Research Institute of Industrial Economics (IFN), P.O. Box 55665, SE-102 15 Stockholm, Sweden. Email: pehr-johan.norback@ifn.se, and lars.persson@ifn.se. We gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Research Foundation and the Marianne and Marcus Wallenberg Foundation (grant number 2020.0049). We have benefitted from the feedback provided by participants at numerous seminars.

1. Introduction

Firms today often collect vast amounts of data (big data) through their regular activities, such as data on sales transactions and on the production process—what we call "operational data". With the introduction of machine learning (ML), operational data has become much more informative and important.[1] Use of ML on increasing amounts of operational data is likely to create large efficiency gains. However, it is also likely to produce regulatory challenges. As is often emphasized, a fundamental feature of ML is that the more data that are available to train a system, the better the system becomes (see, e.g., Dutton, 2018). The introduction of ML applications together with operational data thus creates competitive advantages for incumbent firms due to their access to more operations data (see, e.g., Bessen, 2018). This may increase barriers to entrepreneurship, with severe implications for the dynamism of the economy since breakthrough innovations tend to come from smaller firms and startups rather than large incumbent firms (Cohen, 2010).

The purpose of this paper is to examine the impact of the application of ML to operational data on firm entrepreneurial activity. To this end, we develop a model in which firms use ML on operational data to improve their processes and provide more value for consumers. The model combines active machine-learning-by-doing mechanisms with entrepreneurial innovation mechanisms. As noted by Varian (2018), there are at least three types of returns to scale from applications of ML to operational data whereby incumbents can gain a competitive advantage: classical returns to scale in production, returns to scale due to demand-side network effects, and learning-by-doing, which leads to quality improvements or cost decreases. This model focuses on learning-by-doing aspects of ML applications to operational data by incorporating incumbent firms that employ ML on previously collected proprietary operational data and incoming sales data to enhance the consumer experience associated with their goods or services. However, incumbents also face potential competition from entrepreneurial firms that can invest in research to find more valuable goods or services, more efficient production technology or better business models.

In our model, an entrepreneur needs to successfully innovate to be able to enter a product market, where she faces competition from an incumbent firm which uses ML on its' previously

[1] Using a survey, Bughin et al. (2017) estimate that businesses—mainly large companies—spent \$20–30 billion on AI development in 2016. Venture capital, private equity, and other external sources invested \$6–9 billion.

collected proprietary operational data. The entrepreneur chooses between different types of projects, where a project with a lower probability of success is associated with higher efficiency and profitability, given that it succeeds. Our analysis shows that more extensive use of ML on the incumbent's previously collected proprietary operational data implies that the incumbent becomes efficient and entrepreneurial barriers increase. However, our analysis also reveals that when the entrepreneur has limited access to the incumbent's data, more efficient ML on abundant incumbent data imply that the entrepreneur's incentives to embark on projects with higher risk but higher potential value become stronger. The mechanism is the following: With access to more data, the incumbent becomes more aggressive in the product market, and—for a given project—entry becomes less profitable, if it succeeds. This forces the entrepreneur to switch from a "safer" project with a high associated probability of success—but less value if it succeeds—to a project associated with more risk—but higher value, in the case of success. Greater importance of incumbents' previously collected proprietary data in the market process should therefore result in more risk-taking, as entrepreneurs steer their innovation efforts toward projects that have a low probability of success but a high payoff if they do succeed. Thus, while we should observe abundant entrepreneurial failures, when entrepreneurs do succeed with their projects, the associated inventions and business models should become more creative as ML becomes more prominent and incumbents' previously collected proprietary data become more important.

The appropriate regulatory response to this risk of market domination associated with the use of ML is not easy to determine. Antitrust enforcement officials have already recognized that challenges may arise when large incumbent firms control the vast majority of operational data. For example, FTC commissioner Terrell McSweeny has noted that "it may be that an incumbent has significant advantages over new entrants when a firm has a database that would be difficult, costly, or time consuming for a new firm to match or replicate." In its new strategy for the digital industry, the EU emphasizes the need to ensure that small and medium-sized businesses have adequate access to data and the competence required to implement ML. In the words of EU commissioner Margrethe Vestager, "*the real guarantee of an innovative future comes from keeping markets open so that anyone—big, small—can compete to produce the very best ideas" (Web Summit, 2019). In June 2022, the Bundeskartellamt initiated a proceeding against the technology company Apple to review its tracking rules and the App Tracking Transparency Framework under competition law. Andreas Mundt, President of the Bundeskartellamt, stated: "We welcome*

business models which use data carefully and give users choice as to how their data are used. A corporation like Apple which is in a position to unilaterally set rules for its ecosystem, in particular for its app store, should make pro-competitive rules. We have reason to doubt that this is the case when we see that Apple's rules apply to third parties, but not to Apple itself. This would allow Apple to give preference to its own offers or impede other companies. Our proceeding is largely based on the new competencies we received as part of the stricter abuse control rules regarding large digital companies which were introduced last year (Section 19a German Competition Act - GWB). On this basis, we are conducting or have already concluded proceedings against Google/Alphabet, Meta/Facebook and Amazon."[2] Himel and Seamans (2017) discuss how policy makers might address these issues and describe several policy solutions to consider, including provisions that would institute temporary data monopolies, data portability regimes, and the use of trusted third parties. A key feature of all these suggestions is that incumbents' monopoly access to their operational data should be limited in some way.

To capture this aspect in our model, we assume that the entrepreneurial firm can obtain access to a share of the incumbent's operational data to improve its products and increase customers' willingness to pay. Our analysis then shows that policymakers should consider how these operational data policies affect not only the amount but also the quality of entry. In particular, our analysis shows that policies that make operational data generally available stimulate the amount of entrepreneurship but that this growth could come from entrepreneurs who take on too little risk from the point of view of a society in which consumers benefit from creative inventions. These findings suggest that entrepreneurship policies that reduce the cost of becoming an entrepreneur with access to ML technology might substitute policies regarding access to incumbent operational data.

2. Relation to the literature

This paper contributes to the literature on how the use of ML on big data may affect barriers to entry and entrepreneurship and its implications for intellectual property (IP) and antitrust policy (Bessen (2018)). Farboodi, Mihet, Philippon, and Veldkamp (2019) propose a model where data accumulation increases the skewness of the firm size distribution, as large firms generate more

[2] The Bundeskartellamt. https://www.bundeskartellamt.de/SharedDocs /Meldung/EN/Pressemitteilungen/2022/14_06_2022_Apple.html

data, but where data-savvy small firms can overtake incumbents provided that they can finance their initial money-losing growth. Others contend that operational data alone are not likely to pose an entry barrier (Lambrecht and Tucker 2015, Sokol and Comerford 2016). Bajari et al. (2019) find that increasing the number of online products that Amazon tracks does not significantly improve ML prediction accuracy after a certain point, suggesting that data quantity may function as only a low barrier to entry. We add to this literature by examining the quality of the products or processes with which entry occurs, as well as the likelihood of entry. This qualitative aspect is of fundamental importance since the benefits of industrial restructuring depend not only on the pace at which firms are replaced but also on the nature of the novel products or processes. In particular, we show that the application of ML to incumbents' previously collected proprietary data increase the barrier to entrepreneurship because incumbents' competitive advantage increases due to oligopolistic strategic effects. However, we also show that increased use of ML by incumbents increases entrepreneurs' willingness to take on risk and lengthens the technological jumps that entrepreneurs provide to society.

This paper also contributes to the literature on the effects of privacy, data protection policy and competition (see for instance Acquisti et al (2016)). Jia et al. (2021) find that the EU General Data Protection Regime (GDPR) might constitute a barrier to entry for startups. Campbell (2015) propose a model of how regulatory attempts to protect consumers' data privacy affect the structure of competition and find that the consent-based approach may disproportionately benefit firms that offer a larger scope of services, thus making small firms and new firms the most adversely affected. This prediction is also supported empirically in recent work on the EU GDPR (Johnson and Shriver, 2022; and Batikas et al., 2020). What has not been examined, however, is how such regulations affect the quality of the products and services offered by entrant firms. We add to this stream by proposing a model where machine-learning-by-doing mechanisms with entrepreneurial innovations are central. This enables us to show that policies designed to reduce incumbents' advantages in using ML on previously collected proprietary data could stimulate the amount of entrepreneurship but that this entrepreneurship may takes on too little risk from the point of view of a society in which consumers benefit from breakthroughs that challenge incumbents.

Finally, our paper adds to the literature on firm asymmetries and risk behavior in the R&D process. Rosen (1991) and Cabral (2003) show that small firms may have an incentive to choose

a risky strategy due to strategic output effects in the product market; i.e., small firms do not take on low-risk–low-return projects since they cannot exploit large output improvements. Färnstrand Damsgaard et al. (2017) show that entrepreneurial firms may choose riskier strategies because, unlike incumbents, they will not have already sunk a large share of their entry (commercialization) costs before the outcome of an R&D process is determined. Haufler, Norbäck, and Persson (2014) study the effects of tax policies on entrepreneurs' choice of riskiness (or quality) of an innovation project and show that limited loss offset provisions in the tax system induce entrepreneurs innovating for entry to choose projects with inefficiently low risk but that the same distortion does not arise when entrepreneurs sell their innovation in a competitive bidding process. Henkel et al. (2015) show that independent entrepreneurs who innovate for sale choose riskier R&D projects than incumbents since incumbents have an incentive to opt for safer R&D projects to improve their bargaining power in subsequent acquisitions. We add to this literature by pointing out that the development of ML and the buildup of incumbent proprietary data induce entrepreneurs to take on more risk but that policies that make operational data generally available may be suboptimal. The reason is that this can reduce entrepreneurs' willingness to take on risk.

3. Model

To examine the effects of the application of ML on entrepreneurs' incentives to innovate and enter existing markets, we develop a framework in which firms can use ML to improve their production processes. The model combines active learning-by-doing mechanisms (see, Thompson (2010)) with entrepreneurial innovation mechanisms, as modeled in Färnstrand Damsgaard et al. (2017). In this framework, an incumbent firm obtains an advantage from being able to employ ML on previously collected proprietary data and incoming sales data to increase consumers' willingness to pay. We refer to such data as operational data. However, the incumbent firm also faces potential competition from an entrepreneurial firm that can invest in research to develop new products and also use ML on operational data to increase consumers' willingness to pay.

In stage 1, the entrepreneur can invest in an R&D project that—if successful—generates an invention. The invention can take several forms, all of which increase the profits of its owner. It can be a new product, a product of higher quality or a new or improved production process. For simplicity, we assume that the invention is a product innovation that increases consumers' willingness to pay. The entrepreneur chooses among an infinite number of independent R&D

projects. There is a cost of running a project, and to capture this, we assume that the entrepreneur can undertake only one project.[3] Along the technological frontier, the entrepreneur thus faces a choice between projects that have a high probability of success but deliver a small increase in willingness to pay in case of success and projects that are riskier but also feature a higher increase in willingness to pay if successful.

In stage 2, the outcome of the entrepreneur's R&D project is revealed, where the entrepreneur stays in the market if she is successful and exits otherwise. In the final stage, stage 3, product market interaction takes place, where for simplicity competition is modeled as Cournot competition (in differentiated goods or services). The product market profits depend on whether the entrepreneur succeeds with her R&D project and on the type of project undertaken. Key to the model is that both firms use ML in stage 3 to process sales information to increase consumers' willingness to pay but the incumbent has an advantage in the form of access to preexisting sales/customer data.

In what follows, we analyze the equilibrium of the proposed game, following the usual backward induction procedure.

3.1. Stage 3: Product market

3.1.1. Consumers

Consumers have quasilinear quadratic utility and solve the following utility maximization problem:

$$\underset{\{q_E,q_I,q_0\}}{Max} \quad : \quad U = u(q_E, q_I) + q_0 \tag{3.1}$$

$$s.t \quad : \quad u(q_E, q_I) = a_E \cdot q_E + a_I \cdot q_I - \frac{1}{2} \cdot \left[q_E^2 + q_I^2\right] - q_E \cdot q_I \tag{3.2}$$

$$s.t \quad : \quad P_E \cdot q_E + P_I \cdot q_I + q_0 = m, \tag{3.3}$$

where q_E represents the quantity consumed of the entrepreneur's good, q_I is the quantity consumed of the incumbent's good, and q_0 is the quantity consumed of a numeraire good—or outside good. The subutility function $u(q_E, q_I)$ in (3.2) over the entrepreneur's and the incumbent's goods is linear quadratic.[4] The consumer budget set is given in (3.3), where m is exogenous consumer

[3] See Gilbert (2006) for a motivation.

[4] For a review of quasilinear quadratic utility models, see Choné and Linnemer (2019).

income, P_E is the price of the entrepreneur's product, and P_I is the price of the incumbent's product. The price of the outside good is normalized to one.

Solving for the amount of the outside good q_0 from the budget constraint (3.3) and substituting the quadratic utility $u(q_E, q_I)$ in (3.2) into the direct utility in (3.1), we can rewrite the direct utility as follows:

$$U = [a_E - P_E] \cdot q_E + [a_I - P_I] \cdot q_I - \frac{1}{2} \cdot \left[q_E^2 + q_I^2\right] - q_E \cdot q_I + m. \tag{3.4}$$

Taking the first-order condition for consumer maximization, $\frac{\partial U}{\partial q_i} = 0$ for $i = E, I$, we obtain the residual demand facing each firm:

$$\frac{\partial u}{\partial q_i} = P_i = a_i - q_i - q_j, \text{ for } i, j = \{E, I\}, i \neq j. \tag{3.5}$$

From (3.5), consumers' willingness to pay for a firm's product $\frac{\partial u}{\partial q_i}$ is decreasing in the firm's own output q_i and in the rival's output q_j. We now assume that consumers' willingness to pay, as measured by the intercept a_i, can be affected by firms' use of ML.

3.1.2. Residual demand for the incumbent's product and ML

Firms use ML and big data to increase consumers' willingness to pay by affecting the demand intercept a_i in (3.5) in two distinct ways. For the incumbent firm, the demand intercept is given as

$$a_I = a + \underbrace{\alpha \cdot d_I}_{\text{ML on old data}} + \underbrace{\alpha \cdot q_I}_{\text{ML on new data}}, \tag{3.6}$$

where a is the part of consumers' willingness to pay that is unaffected by ML.

- The incumbent uses historical customer data d_I (available from previous sales) to increase consumers' willingness to pay. This is illustrated in the upper panel in Figure 3.1, starting with the case where the incumbent is a monopolist. Applying ML to preexisting data results in an upward shift of the demand intercept from a to $a + \alpha \cdot d_I$, where we can think of the parameter α as indicating how productive ML is (in using data to increase consumers' willingness to pay), which is a function of the (exogenous) state of computer technology.

- The incumbent can also use the information in contemporaneous sales, q_I, to increase willingness to pay, where we assume that consumers' willingness to pay also increases at

the rate α. This is also shown in the upper panel in Figure 3.1, where ML applied to data on contemporaneous sales makes demand more elastic, shifting the demand curve $a - q_I$ to $a - q_I + \alpha \cdot q_I$. An example would be the information gathered from the road mileage and driving patterns of buyers of self-driving cars, where performance and safety in new cars increase from advanced learning with the generation of new data. Here, we thus take the shortcut of assuming that the incumbent learns directly from its stage 3 sales, similar to the mechanism in the learning-by-doing literature. We also assume that each consumer does not internalize the information that she gives firms with her purchase, captured by the term $\alpha \cdot q_I$, in her consumption choice. This does not seem to be at odds with reality, as it seems notoriously hard for consumers to reap benefits from information sharing.

Putting these two channels of ML together, panel (i) of Figure 3.1 depicts the inverse demand for the incumbent's product when the incumbent faces no competition from the entrepreneur, $P_I^M = a + [\alpha \cdot d_I + \alpha \cdot q_I] - q_I$. In panel (ii) of Figure 3.1, we then depict the incumbent's residual demand—the demand facing the incumbent when q_E units are supplied by the entrepreneur, $P_I = P_I^M - q_E$, or

$$P_I = P_I^M - q_E = a + [\alpha \cdot d_I + \alpha \cdot q_I] - q_I - q_E. \tag{3.7}$$

3.1.3. Residual demand for the entrepreneur's product

The demand for the entrepreneur's product is more involved since consumers' willingness to pay for the entrepreneur's product depends on whether her innovation project succeeds.

The entrepreneur succeeds with her innovation If the entrepreneur succeeds with her innovation, her demand intercept is given by

$$a_E|_{Succeed} = a \quad + \underbrace{\gamma \cdot \alpha \cdot d_I}_{\text{ML on incumbent's old data}} + \underbrace{\alpha \cdot q_E}_{\text{ML on new data}} + \underbrace{[b - \beta \cdot \rho_E]}_{\text{Innovation succeeds}} \tag{3.8}$$

The entrepreneur can also use information from consumers' purchases of her product and apply ML to make the product more attractive, increasing consumers' willingness to pay. This is shown in panel (i) of Figure 3.2, where ML applied to contemporaneous sales shifts the demand curve without ML, $a - q_E$, to the demand curve under ML, $a - q_E + \alpha q_E$. If the entrepreneur has access to the incumbent's historical data, d_I (whether through a general agreement or by law or

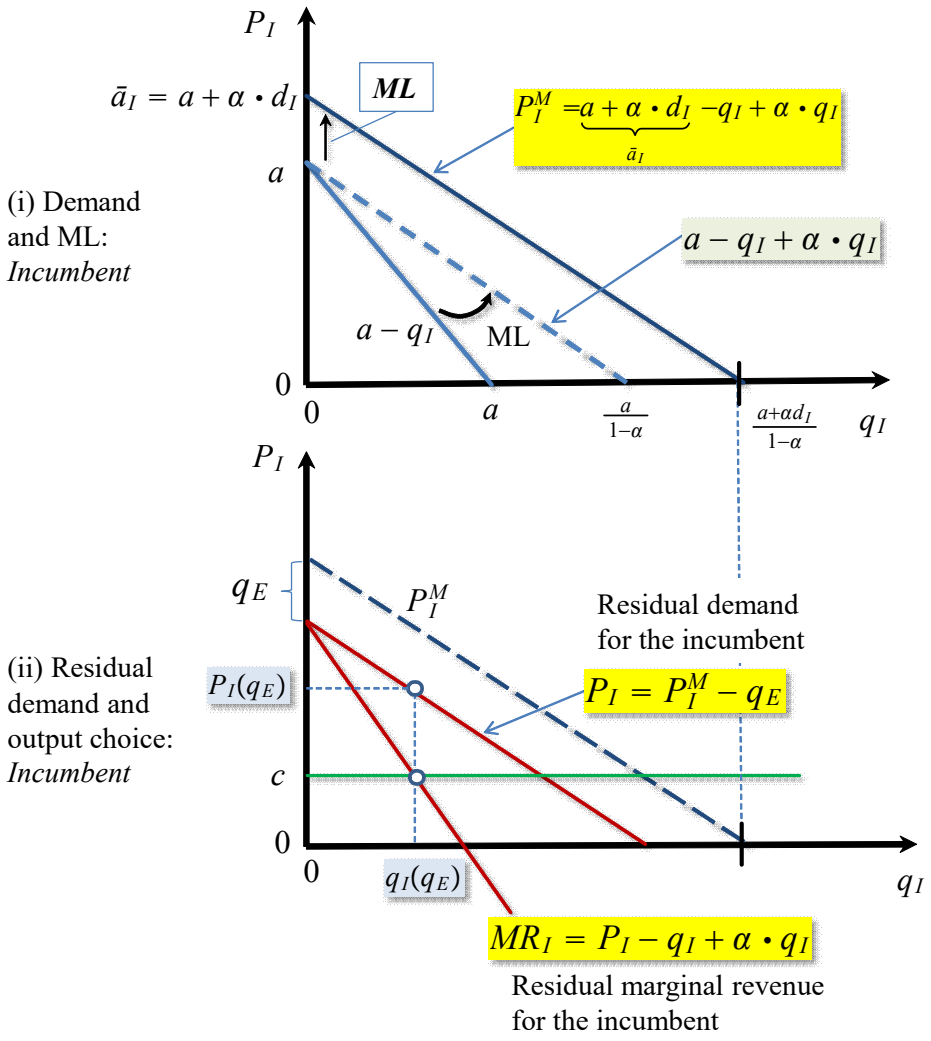


Figure 3.1: Panel (i) shows how ML affects the consumers willingness to pay for the incumbent's product and the incumbent's inverse demand in absence from competition from the entrepreneur. Panel (ii) then derives the incumbent's residual demand and its residual marginal revenue and illustrates its profit maximizing output choice.

regulation), the entrepreneur can also use this data source to increase consumer willingness to pay.

In what follows, we shall assume that the entrepreneur is disadvantaged by not having access to her own historical data, i.e. $d_E = 0$, and by having inferior access to the incumbents' data, where $\gamma \in [0, 1)$ captures the share of the incumbent's data that the entrepreneur has access to. We then define incumbency data advantage as follows:

Definition 1. *The incumbent has privileged data access:* $d_I > 0 = d_E$ *and* $\gamma \in [0, 1)$.

While the entrepreneur has an inherent disadvantage in having weaker access to historical data, she can compensate for this by succeeding with her innovation. Adding new features to her product increases consumers' willingness to pay by an amount $b - \beta\rho_E \geq 0$, where $\beta \in [0, b)$ and $\rho_E \in [0, 1]$ is the probability that the project succeeds. Note how consumers' willingness to pay for the entrepreneur's product is higher if she has taken a greater risk in her research project, i.e., if she has succeeded with a project with a lower probability of success ρ_E. That is,

$$\frac{\partial a_E}{\partial \rho_E} \cdot \underset{(-)}{d\rho_E} = -\beta \cdot \underset{(-)}{d\rho_E} > 0. \tag{3.9}$$

This reflects a natural trade-off where *riskier projects* have a *greater value for consumers if they succeed*. We turn to the project choice in more detail in the next section.

The upward shift of the demand intercept from a to $a + [b - \beta \cdot \rho_E] + \gamma \cdot \alpha \cdot d_I$ in the upper panel in Figure 3.2 illustrates how the entrepreneur can increase consumers' willingness to pay after succeeding with her innovation project by using ML with limited access to the incumbent's data d_I. When ML on contemporaneous sales data is accounted for, the inverse demand for the entrepreneur *without* competition from the incumbent $P_E^M = a + [b - \beta \cdot \rho_E] + \gamma \cdot \alpha \cdot d_I + \alpha \cdot q_E - q_E$ is drawn in panel (i) of Figure 3.2. In panel (ii) of Figure 3.2, we depict the entrepreneur's residual demand—the demand facing the incumbent when q_I units are supplied by the incumbent, that is,

$$P_E = P_E^M - q_I = a + [(b - \beta \cdot \rho_E) + \gamma \cdot \alpha \cdot d_I + \alpha \cdot q_E] - q_E - q_I. \tag{3.10}$$

The entrepreneur fails with her innovation If the entrepreneur fails with her innovation, she is left out of the increase in willingness to pay $b - \beta \cdot \rho_E$ in $a_E|_{Succeed}$ shown in (3.8). We shall assume that consumers' willingness to pay for the entrepreneur's product $a_E|_{Fail} = a_E|_{Succeed} - [b - \beta \cdot \rho_E]$ is too low to secure profitable entry into the product market if she fails.

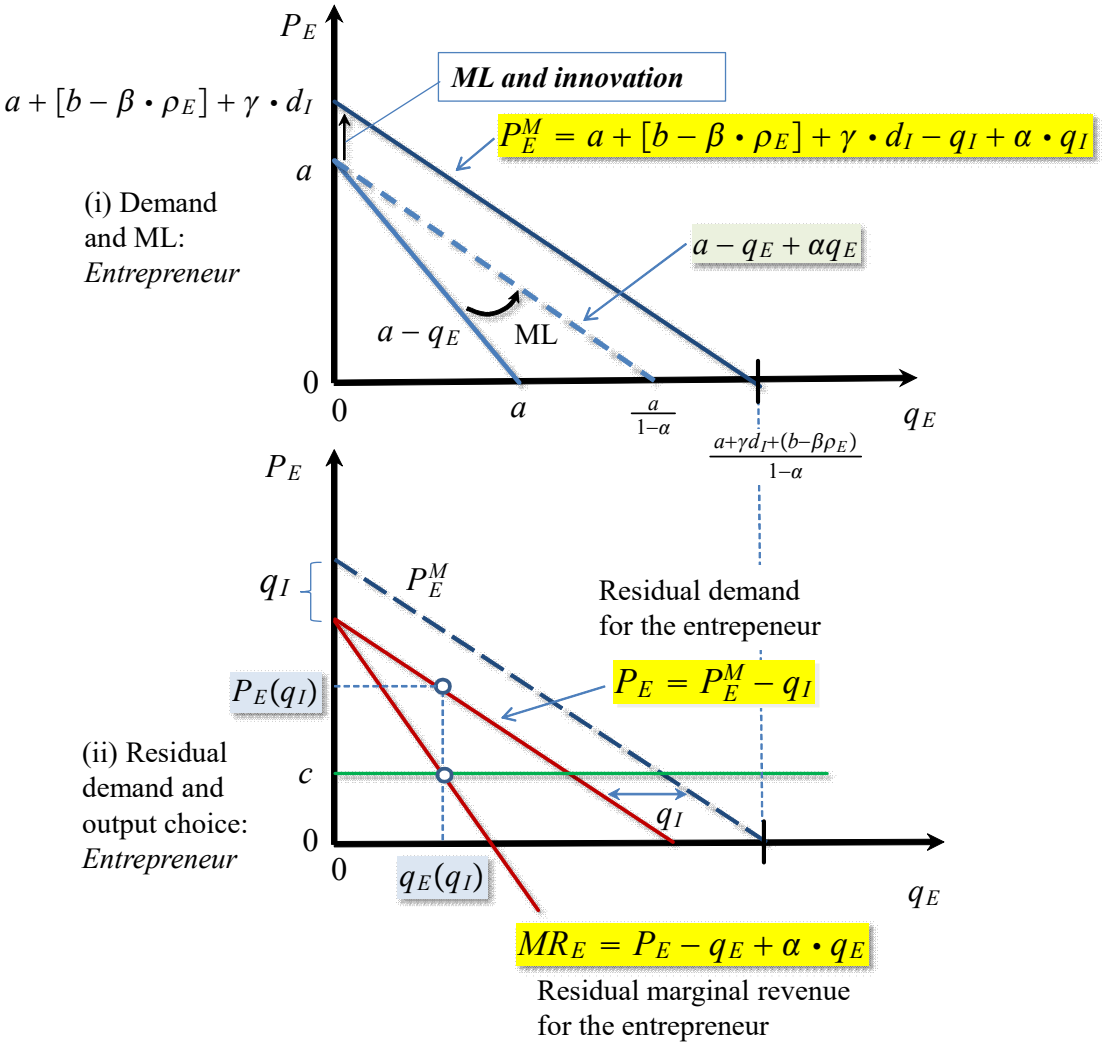


Figure 3.2: Panel (i) shows how ML affects the consumers willingness to pay for the entrepeneur's product given acccess to the incumbents data and depitcs its inverse demand in absence from competition from the incumbent. Panel (i) also illustrates that a succeful innovation project increases consumes willingness to pay—but more so if she has taken a greater risk in her research project. Panel (ii) derives the entrepeneur's residual demand and its residual marginal revenue and illustrates its profit maximizing output choice.

This feature can be formalized in several ways: One is to introduce a sufficiently high entry cost f for the entrepreneur before the quantity competition in stage 2 takes place. Including such an entry cost in the analysis will not affect our results qualitatively, but computations become more involved. Alternatively, we could assume that failing—in addition to the loss of the increase in willingness to pay $b-\beta\cdot\rho_E$ —would also be associated with a loss of confidence in the entrepreneur's product with the cost f effectively becoming a penalty for failing: $a_E|_{Fail} = a_E|_{Succeed} - [b-\beta\cdot\rho_E] - f < 0$. A third approach would be to focus the analyses on outcomes where the incumbent's data advantage is so significant that the entrepreneur would optimally choose a zero output level in the Cournot competition if she fails with her innovation.

For ease of expositional—but with no loss of generality—we shall assume that the entrepreneur will not enter the product market if she fails with the invention. Thus, when the entrepreneur fails with her innovation, the incumbent is a monopolist in the market with inverse demand $P_I^M = a + [\alpha\cdot d_I + \alpha\cdot q_I] - q_I$.

3.1.4. Optimal quantity

The profit maximization problem of firm i is

$$\max_{\{q_i\}} \pi_i = [P_i - c]\, q_i,\ i = \{I, E\}, \tag{3.11}$$

where each firm's price P_i is given from the residual demand functions (3.7) and (3.10) and where we assume that each firm faces a constant marginal cost c. The first-order conditions $\frac{\partial \pi_i}{\partial q_i} = 0$ imply that each firm chooses its output such that marginal revenue equals marginal cost or

$$\underbrace{P_i - (1-\alpha)\, q_i^*}_{MR_i} = \underbrace{c}_{MC_i},\ i = \{I, E\}, \tag{3.12}$$

where $dP_i/dq_i = -(1-\alpha)$ from (3.7) and (3.10) and where $\frac{da_i}{dq_i} = \alpha$ from (3.6) and (3.8), where the latter expressions capture how ML increases consumers' willingness to pay from information on contemporaneous sales.[5] These first-order conditions (giving each firm's best-response to its rival) are also illustrated in the lower panels of Figures 3.1 and 3.2.

[5] The second-order condition is fulfilled since $\frac{\partial^2 \pi_i}{\partial q_i^2} = 2(1-\alpha) < 0$.

3.1.5. The Nash–Cournot equilibrium

To derive the Nash–Cournot equilibrium, it is useful to derive the firms' reaction functions. Using (3.6) and (3.8) and defining $\Lambda = a - c$, define consumers' net willingness to pay for each firm's product $\bar{\Lambda}_i$ as

$$\bar{\Lambda}_I(\Lambda, \alpha, d_I) = \Lambda + \alpha \cdot d_I, \tag{3.13}$$

$$\bar{\Lambda}_E(\Lambda, b, \beta, \rho_E, \alpha, d_I, \gamma) = \Lambda + [b - \beta \cdot \rho_E] + \gamma \cdot \alpha \cdot d_I. \tag{3.14}$$

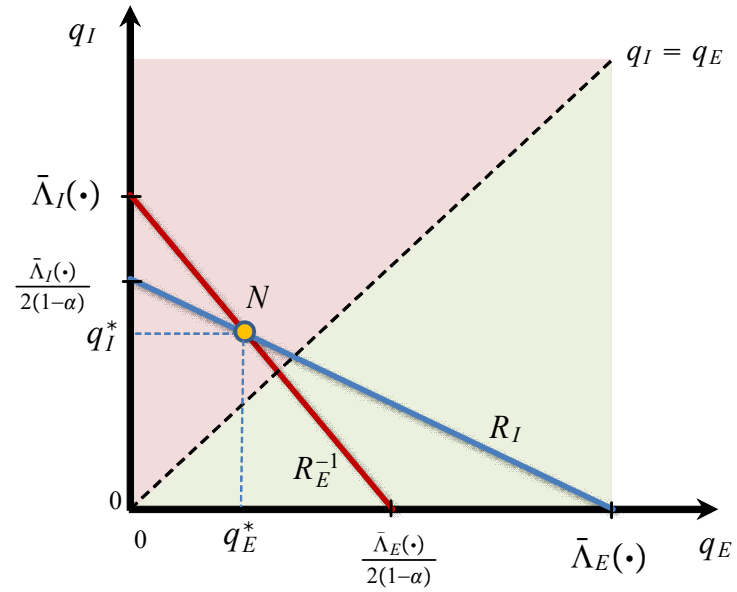
As shown, consumers' net willingness to pay increases when ML techniques become more efficient due to better computers, i.e., when α increases. For a given computer technology, applying ML to more data allows firms to better infer consumer preferences and further increase consumers' willingness to pay. However, since the entrepreneur does not have full access to the incumbent's data, $\gamma \in [0, 1)$, the increase in net willingness to pay is smaller for the entrant than for the incumbent: $\partial\bar{\Lambda}_I(\cdot)/\partial d_I = \alpha > \gamma\alpha = \partial\bar{\Lambda}_E(\cdot)/\partial d_I$. Again, this can be compensated for if the entrepreneur succeeds with her innovation, in which case the net willingness to pay for the entrepreneur's product $\bar{\Lambda}_E(\cdot)$ rises with the new product's features, $b - \beta \cdot \rho_E > 0$, with the increase in willingness to pay endogenously determined by the project choice ρ_E.

From (3.7)–(3.14), we can derive the firms' reaction functions:

$$R_i(q_j) = \frac{\bar{\Lambda}_i(\cdot) - q_j}{2(1-\alpha)},\ i, j = \{E, I\}, i \neq j. \tag{3.15}$$

The reaction function for firm i, $R_i(q_j)$ gives the optimal output choice q_i for given choice of output by firm j, q_j. The reaction function of the incumbent $R_I(q_E) = \frac{\bar{\Lambda}_I(\cdot) - q_E}{2(1-\alpha)}$ is depicted as the downward-sloping dark-blue curve in panel (i) in Figure 3.3. The downward slope captures the fact that the firms' quantities are strategic substitutes: If the incumbent believes that the entrepreneur will produce more, she will expect a lower price for her product, and—as shown in Eq. 3.12—lower marginal revenue, which induces her to produce less to bring marginal revenue equal to marginal cost. Since Figure 3.3 is drawn with the output of the incumbent on the vertical axis and the output of the entrepreneur on the x-axis, we use the inverse reaction function for the entrepreneur, $R_E^{-1}(q_E) = \bar{\Lambda}_E(\cdot) - 2(1-\alpha) q_E$. The reaction function of the entrepreneur is also downward sloping again, displaying the fact that quantities are strategic substitutes: If the entrepreneur expects the incumbent to produce high output, she will expect a low price for her product and lower marginal revenue, which will induce her to choose lower output.

(i) Reaction functions and the Nash-equilibrium in the product market



(ii) The change in the Nash-equilibrium when the incumbent has more historical data

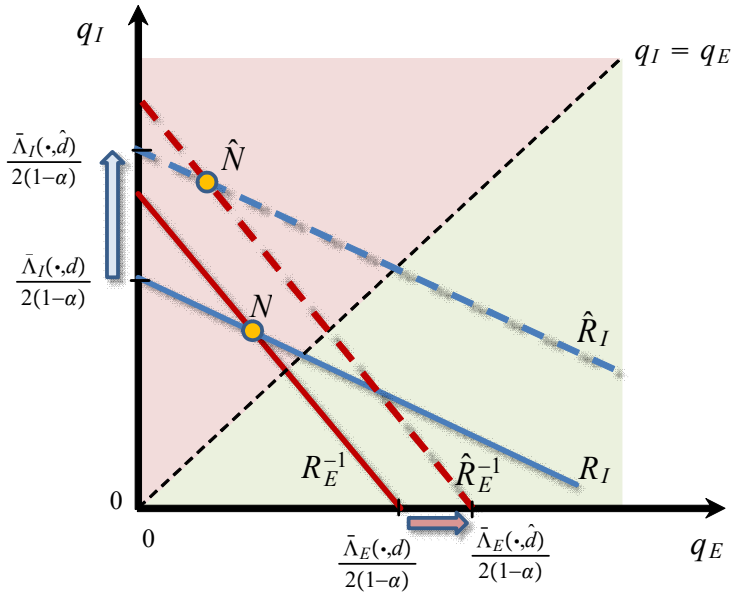


Figure 3.3: Illustrating the Nash-equilibrium in the product market for a given (succesful) innovation project chosen by the entrepeneur. Panel (i) illustrates the initial Nash equilbrium with the incumbent assumed to be the larger firm. Panel (ii) illustrates the shift in the Nash equilibrium toward reinforced incumbent domination (under weak access to the incumbent's historical data for the entrepeneur).

The Nash–Cournot equilibrium is reached when both firms choose output optimally and correctly infer the output choice of their rival; i.e., the Nash–Cournot equilibrium is given from the intersection of the reaction functions at point N in panel (i) in Figure 3.3. It is straightforward to verify that the Nash–Cournot equilibrium is

$$q_I^* = \frac{2(1-\alpha)\Lambda_I - \Lambda_E}{(1-2\alpha)(3-2\alpha)} = \frac{(1-2\alpha)\Lambda - (b-\beta\rho_E) + (2(1-\alpha)-\gamma)d_I\alpha}{(1-2\alpha)(3-2\alpha)} \tag{3.16}$$

$$q_E^* = \frac{2(1-\alpha)\Lambda_E - \Lambda_I}{(1-2\alpha)(3-2\alpha)} = \frac{(1-2\alpha)\Lambda + 2(b-\beta\rho_E)(1-\alpha) - (1-2(1-\alpha)\gamma)d_I\alpha}{(1-2\alpha)(3-2\alpha)}, \tag{3.17}$$

where $1-2\alpha > 0$ ensures the stability of the equilibrium.[6] Stability, i.e., $\alpha \in [0, 1/2)$, ensures that $3-2\alpha > 0$ and $2(1-\alpha)-\gamma > 0$.

3.1.6. Data availability and product market outcome

Our main interest lies in exploring how the equilibrium in the product market and the entrepreneur's incentives to innovate are affected by access to data and the use of ML. Let us first explore how the amount of historical data in the hand of the incumbent d_I affects the Nash equilibrium in (3.16) and (3.17) for a *given project choice* of the entrepreneur ρ_E.

To proceed, we make use of the following definition:

Definition 2. *The entrepreneur has (i) strong access to the incumbent's historical data d_I if and only if $1-2(1-\alpha)\gamma < 0$ and (ii) weak access to the incumbent's historical data d_I if and only if $1-2(1-\alpha)\gamma > 0$.*

We will use this definition to describe how the incumbent's amount of historical data affects the equilibrium behavior of the entrepreneur. We will explore how the entrepreneur's behavior is affected by her access to the incumbent's historical data, γ, in more detail in Section 3.3. The Appendix also illustrates the impact on the entrepreneur's behavior when the effectiveness of ML, α, varies.

We can derive the following results:

Proposition 1. *When the incumbent's amount of historical data d_I increases,*

(i) *The incumbent always expands its output, $\frac{\partial q_I^*}{\partial d_I} > 0$;*

[6] This ensures that the reaction function of the entrant is steeper than that of the incumbent.

(ii) *The entrant expands her output only when she has strong access to the incumbent's historical data* d_I: $\frac{\partial q_E^*}{\partial d_I} > 0$ iff $1 - 2(1-\alpha)\gamma < 0$ and $\frac{\partial q_E^*}{\partial d_I} < 0$ iff $1 - 2(1-\alpha)\gamma > 0$.

(iii) *The incumbent always expands its output more than the entrepreneur:* $\frac{\partial q_I^*}{\partial d_I} > \frac{\partial q_E^*}{\partial d_I} > 0$.

To prove parts (i) and (ii), partially differentiate (3.16) and (3.17) to obtain:

$$\frac{\partial q_I^*}{\partial d_I} = \frac{2(1-\alpha)-\gamma}{(1-2\alpha)(3-2\alpha)}\alpha > 0, \tag{3.18}$$

$$\frac{\partial q_E^*}{\partial d_I} = \begin{cases} -\frac{1-2(1-\alpha)\gamma}{(1-2\alpha)(3-2\alpha)}\alpha > 0, \text{ for } 1-2(1-\alpha)\gamma < 0, \\ -\frac{1-2(1-\alpha)\gamma}{(1-2\alpha)(3-2\alpha)}\alpha < 0, \text{ for } 1-2(1-\alpha)\gamma > 0. \end{cases} \tag{3.19}$$

where again stability, i.e., $\alpha \in [0, 1/2)$, ensures $2(1-\alpha)-\gamma > 0$.

As shown in (3.18), the incumbent strictly increases its output with access to more data, $\frac{\partial q_I^*}{\partial d_I} > 0$. As shown in the upper line in (3.19), this is also the case for the entrepreneur when she has strong access to the incumbent's historical data, i.e., $\frac{\partial q_E^*}{\partial d} > 0$ if $\gamma \in (\frac{1}{2(1-\alpha)}, 1]$.[7] However, as shown by the lower line, in (3.19), if the entrepreneur has weak access to the incumbent's historical data, $\gamma \in [0, \frac{1}{2(1-\alpha)}]$, the entrepreneur's output contracts when the incumbent has access to larger amounts of historical data, $\frac{\partial q_E^*}{\partial d_I} < 0$. Figure 3.3 (ii) provides an illustration of the interaction in the latter case: Increases in the amount of historical data held by the incumbent d_I and both firms' (differential) application of ML to these data—and to new data from contemporaneous sales—induce both firms to increase sales as consumers' willingness to pay increases. Hence, both firms' reaction functions shift outward. However, with access to the incumbent's historical data suppressed, the entrepreneur's reaction function shifts outward less than the incumbent's, and the incumbent reinforces her market dominance.

To prove part (iii), first note that when the entrepreneur has weak access to the incumbent's historical data, i.e., $1-2(1-\alpha)\gamma > 0$, it immediately follows that $\frac{\partial q_I^*}{\partial d_I} - \frac{\partial q_E^*}{\partial d_I} > 0$. When the entrepreneur has strong access to the incumbent's historical data, i.e., $1-2(1-\alpha)\gamma < 0$, (3.18) and (3.19) directly imply $\frac{\partial q_I^*}{\partial d_I} - \frac{\partial q_E^*}{\partial d_I} = \alpha\frac{1-\gamma}{1-2\alpha} > 0$.

However, the amount of historic data d_I held by the incumbent does not only affect the product market equilibrium—the amount of data and the access to it by the entrepreneur also affect the entrepreneur's innovation incentives through its effects on the entrepreneur's project choice, ρ_E. This innovation channel—which we have ignored so far—is the subject of the next section.

[7] Note that at the limit $\alpha = 1/2$, $\frac{1}{2(1-1/2)} = 1$, so that $\gamma \in [0, 1]$ is fulfilled.

3.2. Stage 2: R&D by the entrepreneur

In this stage, the entrepreneur decides on her optimal R&D project. Using the direct profit function (3.11), the residual demand (3.10), the net willingness to pay (3.14) and the Nash quantity in (3.17), we can write the reduced-form product market profit for the entrepreneur:

$$\pi_E(\rho_E) = \left(\underbrace{\bar{\Lambda}_E(\cdot, \rho_E) + \alpha \cdot q_E^*(\rho_E) - q_E^*(\rho_E) - q_I^*(\rho_E)}_{P_E - c} \right) \times q_E^*(\rho_E). \tag{3.20}$$

By assumption, the entrepreneur will only enter the market if the selected R&D project turns out to be successful in stage 1.[8] This outcome occurs with probability ρ_E and generates the net profit $\pi_E^*(\rho_E)$ for the entrepreneur. The entrepreneur's expected profit is therefore given as:

$$\underset{\{\rho_E\}}{Max} \quad : \quad \mathrm{E}[\Pi_E] = \rho_E \times \pi_E(\rho_E), \tag{3.21}$$

$$s.t \quad : \quad \rho_E \in [0,1], \tag{3.22}$$

$$s.t \quad : \quad \pi_E(\rho_E) > 0. \tag{3.23}$$

Let us first focus on an interior solution: a solution ρ_E^* that fulfills the constraints (3.22) and (3.23). The first-order condition for an interior solution, $\frac{dE[\Pi_E]}{d\rho_E} = 0$, is then

$$\underset{\text{Securing Success (SS):}}{\pi_E(\rho_E^*)} = \underbrace{-\rho_E^* \times \frac{d\pi_E(\rho_E^*)}{d\rho}}_{\text{Cost of Going Safer (CGS):}} > 0. \tag{3.24}$$

As shown in (3.24), we can understand this first-order condition from two distinct effects.

The securing-success effect The left-hand side of (3.24) gives the *increase in expected profit from choosing a marginally safer project* and is simply the reduced product market profit from succeeding, $\pi_E(\rho_E)$. We label this the *securing success (SS) effect*.

The cost-of-going-safer effect The right-hand side of (3.24) represents *the reduction in expected profit from choosing a marginally safer project*, which we label the *cost-of-going-safer (CGS) effect*. The downside of choosing a safer project stems from a lower consumer willing-

[8] As explianed at the end of Section 3.1.3, it is straightforward to formalize that the entrepreneur stays out of the market if the innovation project fails.

ness to pay and more aggressive competition from the incumbent. To see this, use the envelope theorem in (3.20) to obtain[9]

$$\frac{d\pi_E(\rho_E)}{d\rho_E} = \left(\overbrace{\underbrace{\frac{\partial a_E}{\partial \rho_E}}_{(-)} + \underbrace{\frac{\partial P_E}{\partial q_I}}_{(-)} \times \underbrace{\frac{dq_I^*}{d\rho_E}}_{(+)}}^{(-)} \right. \\ \left. \begin{array}{cc} \\ \text{Direct demand effect} & \text{Strategic effect} \end{array} \right) \times q_E^*(\rho_E) < 0. \tag{3.25}$$

The first term shows that choosing a project with a marginally higher probability of success reduces consumers' willingness to pay (if the project is successful), $\frac{\partial a_E}{\partial \rho_E} = -\beta < 0$. This reduces the entrepreneur's product market price from (3.10). The second term captures that a lower willingness to pay for the entrepreneur's product also induces the incumbent rival to be more aggressive in the product market. This follows since $\frac{dq_I^*}{d\rho_E} = \frac{\beta}{(1-2\alpha)(3-2\alpha)} > 0$ from (3.16), which further reduces the entrepreneur's product market price since $\frac{\partial P_E}{\partial q_I} = -1 < 0$ from the residual demand in (3.10). Using the information in (3.25), we can then rewrite the *CGS effect* in (3.24) as

$$-\rho_E \frac{d\pi_E(\rho_E)}{d\rho_E} = \rho_E \left(1 + \frac{1}{(1-2\alpha)(3-2\alpha)}\right) \beta \times q_E^*(\rho_E) > 0. \tag{3.26}$$

3.2.1. The optimal project choice

We are now ready to determine the optimal project. First, note that since $P_E - c = (1-\alpha) q_E^*(\rho_E)$ holds from (3.12), the reduced product market profit in (3.20) is a quadratic function of the Nash output

$$\pi_E(\rho_E) = (1-\alpha)\left[q_E^*(\rho_E)\right]^2. \tag{3.27}$$

Inserting (3.26) and (3.27) into the first-order condition in (3.24), we then obtain

$$\left((1-\alpha) q_E^*(\rho_E) - \left(1 + \frac{1}{(1-2\alpha)(3-2\alpha)}\right) \beta \times \rho_E\right) \times q_E^*(\rho_E) = 0. \tag{3.28}$$

Note that this first-order condition (3.28) holds if the bracketed expression is zero, output is zero, or both of these conditions hold. Thus, we have two candidates for the optimal project, $\hat{\rho}_E$ and ρ_E^*:

$$q_E^*(\hat{\rho}_E) = 0, \tag{3.29}$$

$$(1-\alpha) q_E^*(\rho_E^*) - \left(1 + \frac{1}{(1-2\alpha)(3-2\alpha)}\right) \times \beta \times \rho_E^* = 0, \; q_E^*(\rho_E^*) > 0. \tag{3.30}$$

[9] Changes in the entrepreneur's own output $q_E^*(\rho_E)$ have only a second-order effect on the reduce-form profit since output is already optimally set from (3.12).

From (3.17), we know that choosing an easier project comes with less consumer appreciation in terms of lower net willingness to pay, which contracts output, $\frac{\partial q_E^*(\rho_E)}{\partial \rho_E} = -\frac{2\beta(1-\alpha)}{(1-2\alpha)(3-2\alpha)} < 0$. However, choosing ρ_E to achieve zero output cannot be optimal since this would imply that expected profit is zero, $\Pi_E(\hat{\rho}_E) = \hat{\rho}_E \pi_E(\hat{\rho}_E) = \hat{\rho}_E(1-\alpha)\left[q_E^*(\hat{\rho}_E)\right]^2 = 0$. Thus, only$\rho_E^*$ in (3.30) can be a maximum.

To derive ρ_E^*, it is useful to rearrange (3.30) to obtain

$$\underbrace{(1-\alpha)\, q_E^*(\rho_E^*)}_{\text{Securing Success (SS):}} = \underbrace{\left(1 + \frac{1}{(1-2\alpha)(3-2\alpha)}\right) \times \beta \times \rho_E^*}_{\text{Cost of Going Safer (CGS):}}, \tag{3.31}$$

where the left-hand side is the SS effect and the right-hand side represents the CGS effect, *rewritten* in linear form noting (3.27).

In Figure 3.4(i), we now illustrate how the SS effect and CGS effect shape the equilibrium. The downward-sloping curve labeled SS is the SS effect—it shows the benefit from succeeding with a marginally safer project in terms of per-unit profit. The SS curve is downward sloping since the value of securing success is lower the more likely the project is to succeed (since the quality of the project is inversely related to success probability—see Equation 3.9). The upward-sloping curve labeled CGS is the CGS effect and shows the reduction in per-unit profit from a safer project from lower consumer willingness to pay and intensified competition from the incumbent. The CGS curve is upward sloping since the cost of going safer is higher the more likely the project is to succeed.

The optimal project ρ_E^* is thus given from the intersection of the SS locus and the CGS locus and illustrated at point A in Figure 3.4(i). Combining (3.17) and (3.30), we obtain

$$\rho_E^* = \frac{1}{6\beta}\left(\left(\frac{1-2\alpha}{1-\alpha}\right)\Lambda + 2b - \left(\frac{1-2(1-\alpha)\gamma}{1-\alpha}\right)\alpha d_I\right). \tag{3.32}$$

In an appendix, available upon request from the authors, we (i) verify that ρ_E^* is the unique maximum, i.e., $\frac{\partial^2 \Pi_E(\rho_E^*)}{\partial \rho_E^2} < 0$, and (ii) derive the conditions under which the ρ_E^* satisfies $\rho \in [0,1]$ and $q_i^*(\rho_E^*) > 0$.

3.2.2. Comparative statics on the entrepreneur's project choice

Let us now explore comparative statics results on the entrepreneur's project choice.

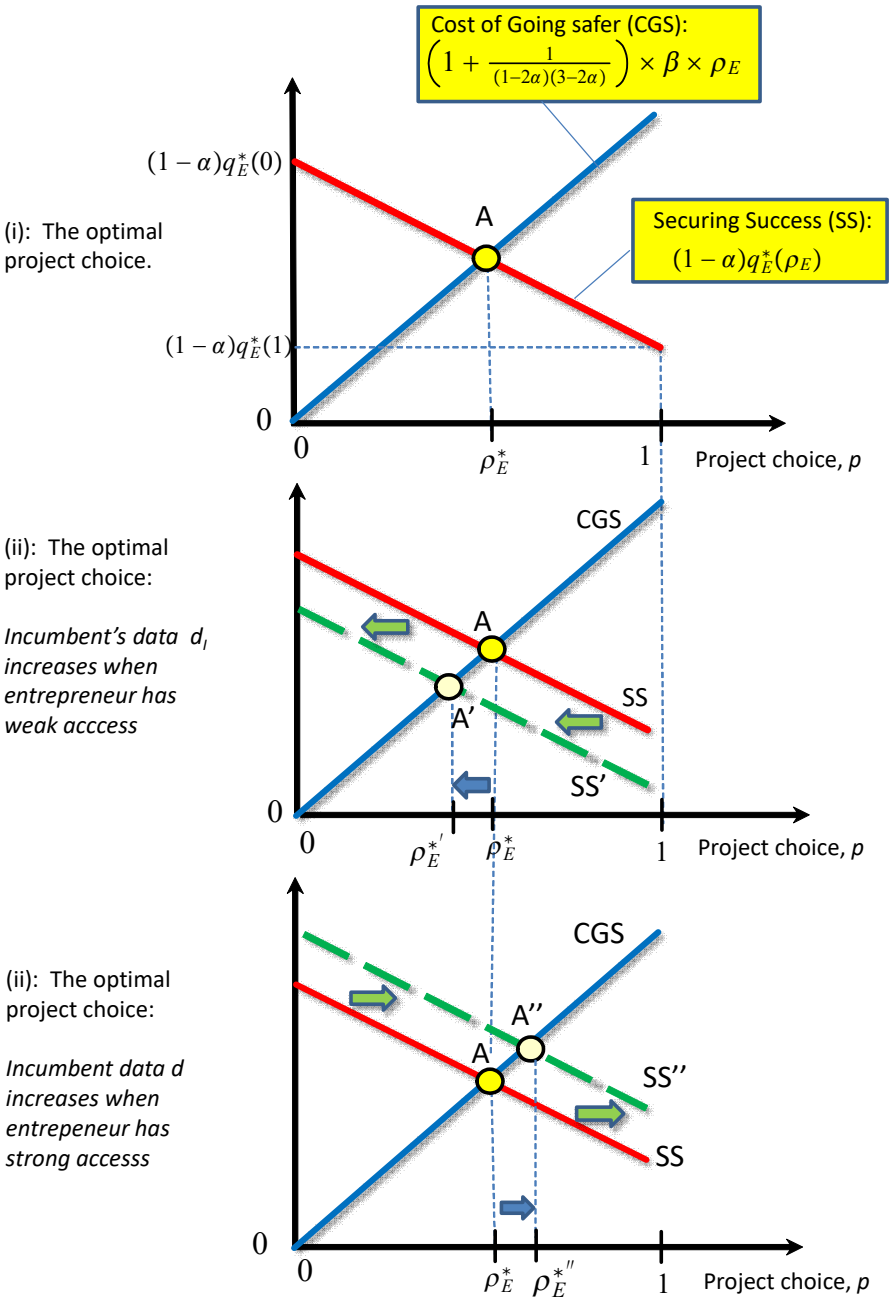


Figure 3.4: Panel (i) derives the optimal innovation project of the entrepeneur for a given amount of historical data held by the incumbent. Panel (ii) illustrates the change in project choice by the entrepreneur when she has weak access to the incumebent's historical data. Panel (iii) illustrates the change in project choice by the entrepreneur when she has strong access to the incumebent's historical data.

Amount of incumbent's historical data What is the effect on the entrepreneur's optimal project if the incumbent has access to more historical data?

We have the following proposition:

Proposition 2. *If the incumbent firms possesses more historical data d_I, then*

(i) *The entrepreneur chooses an R&D project with a lower probability of success when the entrepreneur has weak access to the incumbent's historical data, i.e., $\frac{d\rho_E^*}{dd_I} < 0$ if $1-2\gamma(1-\alpha) > 0$.*

(ii) *The entrepreneur chooses an R&D project with a higher probability of success when the entrepreneur has strong access to the incumbent's historical data, i.e., $\frac{d\rho_E^*}{dd_I} > 0$ if $1-2\gamma(1-\alpha) < 0$.*

By calculation from (3.32) and (3.8), we have

$$\frac{d\rho_E^*}{dd_I} = -\frac{\alpha}{\beta} \cdot \frac{(1-2\gamma(1-\alpha))}{(1-\alpha)} \tag{3.33}$$

Given that the entrepreneur has weak access to the incumbent's historical data, $1-2\gamma(1-\alpha) > 0$, more abundant historical data held by the incumbent decreases the entrepreneur's output, $\frac{\partial q_E^*}{\partial d_I} < 0$ (as shown in the lower line in Equation 3.19). This implies that the value of securing success decreases, as illustrated by a downward shift of the SS curve to the curve SS' in Figure 3.4 (ii). Since the CGS locus is unaffected from (3.31), we can infer that the incumbent's possession of more historical data induces the entrepreneur to choose a riskier project, moving from ρ_E^* to $\rho_E^{*\prime} < \rho_E^*$.

On the other hand, when the entrepreneur has strong access to incumbent's historical data, i.e., if $1-2\gamma(1-\alpha) < 0$ holds, the availability of more historical data increases the entrepreneur's output, $\frac{\partial q_E^*}{\partial d_I} > 0$ (as shown in the lower line in Equation 3.19). This implies that the value of securing success increases and is illustrated by an upward shift of the SS curve to SS" in Figure 3.4 (iii). The incumbent's possession of more historical data now induces the entrepreneur to go for a safer project, moving from ρ_E^* to $\rho_E^{*\prime\prime} > \rho_E^*$.

Entrepreneur's access to the incumbent's historical data (γ) What is the effect on the entrepreneur's optimal project if the entrepreneur's access to more historical operational data is improved? We have the following proposition:

Proposition 3. *If the entrepreneur obtains better access to the incumbent's historical data, then the entrepreneur chooses an R&D project with a higher probability of success, i.e.,* $\frac{d\rho^*}{d\gamma} > 0$.

By calculation from (3.32), we have

$$\frac{\partial \rho_E^*}{\partial \gamma} = \frac{1}{3}\frac{\alpha}{\beta} d_I > 0. \tag{3.34}$$

Thus, better access to the incumbent's historical data induces the entrepreneur to go for a safer project. A less risky project then adds less value consumers if it succeeds.

More efficient ML (α) We can also examine how the entrepreneur's project is affected if ML technology improves, which is captured by an increase in α. We can then state the following proposition:

Proposition 4. *If ML becomes more effective, the entrepreneur chooses an R&D project with a lower probability of success—and a higher consumer willingness to pay given success, i.e.,* $\frac{d\rho_E^*}{d\alpha} < 0$ *if* $\gamma < 1/2$.

By calculation from (3.32), we have

$$\frac{d\rho_E^*}{d\alpha} = \frac{1}{6}\frac{\left(2\gamma\left(\alpha^2 - 2\alpha + 1\right) - 1\right) d_I - \Lambda}{\beta\left(1-\alpha\right)^2} < 0 \text{ if } \gamma < 1/2. \tag{3.35}$$

When ML becomes more efficient, it becomes costlier for the entrepreneur to choose a safer project since the strategic effect of a more aggressive incumbent in the product market becomes stronger.

In terms of Figure 3.4(i), this would cause the CGS locus to twist counterclockwise (not shown). Indeed, partially differentiating the right-hand side of (3.31), we see that the expected cost of going safer increases when α increases

$$\frac{\partial}{\partial \alpha}\left(\left(1 + \frac{1}{\left(1-2\alpha\right)\left(3-2\alpha\right)}\right) \times \beta \times \rho_E\right) = 8\frac{1-\alpha}{\left(3 + 4\alpha^2 - 8\alpha\right)^2} \times \beta \times \rho_E > 0. \tag{3.36}$$

The effect of more effective ML on the SS locus is more involved. The entrepreneur chooses output such that profit per unit equals the net reduction in revenues per unit from a unit expansion in sales, $P_E - c = (1-\alpha)q_E^*(\rho_E)$. More efficient use of data, increasing α, then makes expansion less costly and hence allows the entrepreneur to operate with a lower per-unit profit at an unchanged output level. This would shift SS downward in Figure 3.4(i) (again

not shown). Since more efficient ML increases consumers' willingness to pay, this also gives the entrepreneur an incentive to increase her output, which makes the SS effect stronger and shifts the SS' condition further upward. However, as derived above, if the entrepreneur has sufficiently low access to the incumbent's data, the entrepreneur always responds to more efficient ML by choosing a riskier project.

3.2.3. Why ML and big data may lead to more creation — but less destruction

Let us now put our results together and explore the main question of interest in this paper: *What is the impact of more protected big data and ML on the creative destruction process*?

From (3.8) and as illustrated in panel (i) in Figure 3.2, we know that when succeeding with the invention consumers willingness for the entrepreneur' product will increase with the amount

$$\Delta\ a_E|_{Succeed} = b - \beta \cdot \rho_E^*(d_I). \tag{3.37}$$

From (3.37) a *riskier project* (lower ρ_E^*) then has have a *greater value for consumers if it succeeds.* To this end, we shall then define creative entrepreneurship as follows:

Definition 3. ***Creative entrepreneurship:*** *Entrepreneurship is (more) creative when the entrepreneur takes on more risk and aims for a more innovative invention in her innovation decision.*

If the entrepreneur succeeds and enters the product market, this will have a business-stealing effect which will be destructive for the incumbent. We shall then define destructive entrepreneurship as follows:

Definition 4. ***Destructive entrepreneurship:*** *Entrepreneurship is destructive when the entrepreneur, through a successful innovation, can enter the market and overtake the incumbent's position as market leader.*

To capture destructive entrepreneurship, we could calculate the market share for each firm. However, without loss of generality, it turns to be easier to capture destructive entrepreneurship by comparing profits using reduced-form profits as function of the incumbent firm's historical data d_I. In the next section, we will show that results hold also with more traditional market shares.

To this end, let $\pi_i(\rho_E^*(d_I), d_I) \equiv \pi_i(q_i^*(\rho_E^*(d_I), d_I), q_j^*(\rho_E^*(d_I), d_I), d_I)$ for $i, j = \{E, I\}$ and $i \neq j$. Then, let the relative reduced-form profit of the entrant be $\varphi_E(d_I)$, or

$$\varphi_E(d_I) = \frac{\pi_E(\rho_E^*(d_I), d_I)}{\pi_I(\rho_E^*(d_I, d_I)} = \frac{(1-\alpha)\left[q_E^*(\rho_E^*(d_I), d_I)\right]^2}{(1-\alpha)\left[q_I^*(\rho_E^*(d_I), d_I)\right]^2} = \frac{q_E^*(\rho_E^*(d_I), d_I)}{q_I^*(\rho_E^*(d_I), d_I).} \tag{3.38}$$

From (3.38), it directly follows that destructive entrepreneurship can be captured by simply comparing firms' outputs

$$\varphi_E(d_I) > 1 \Leftrightarrow q_E^*(\rho_E^*(d_I), d_I) > q_I^*(\rho_E^*(d_I), d_I). \tag{3.39}$$

However, does an equilibrium exist where the entrepreneur overtakes the incumbent, i.e., a post-entry market equilibrium where $\varphi_E(d_I) > 1$? Does this occur when the incumbent has less or more historical data d_I? How does the entrepreneur's access to the incumbent's historical proprietary data γ affect the likelihood of this equilibrium? How does higher efficiency of ML α affect the likelihood of this equilibrium?

To begin our analysis, it useful to return to Figures 3.4 (i) and (ii). Recall that we started in a situation where the data advantage of the incumbent makes the incumbent the market leader: this Nash equilibrium is now reproduced at point N in Figure 3.5, where $\varphi_E^N(d_I) < 1$, as point N is above the 45-degree line where $\varphi_E = 1$. Consider what happens if the incumbent has access to more historical data. In Proposition 1, we showed that when the amount of historical data held by the incumbent increases, the incumbent expands more than the entrant given that we hold the project choice of the entrepreneur—and hence the quality of the innovation—constant. That is:

$$0 < \frac{\partial q_I^*(\rho_E^*(d_I), d_I)}{\partial d_I} > \frac{\partial q_E^*(\rho_E^*(d_I), d_I)}{\partial d_I} = \begin{cases} < 0; & 1 - 2\gamma(1-\alpha) > 0 \\ > 0; & 1 - 2\gamma(1-\alpha) < 0. \end{cases} \tag{3.40}$$

Suppose that the entrepreneur has weak access to the incumbents data in (3.40), $1-2\gamma(1-\alpha) < 0$. The movement from of the Nash-equilbrium from N to $\hat{N}$ in Figure 3.5 involves only the direct change in output from an increase in historical data, i.e., the change in output holding the entrepreneur's innovation project choice constant. However, Proposition 2(i) then showed that the entrepreneur chooses a riskier project when the incumbent has access to more historical data, given weak access of the entrepreneur to these data, i.e., $\frac{d\rho_E^*}{dd_I} < 0$. If such a project succeeds, it delivers a higher consumer willingness to pay, and the incumbent would therefore face more aggressive competition from the entrepreneur. This is shown by the movement from $\hat{N}$ to N',

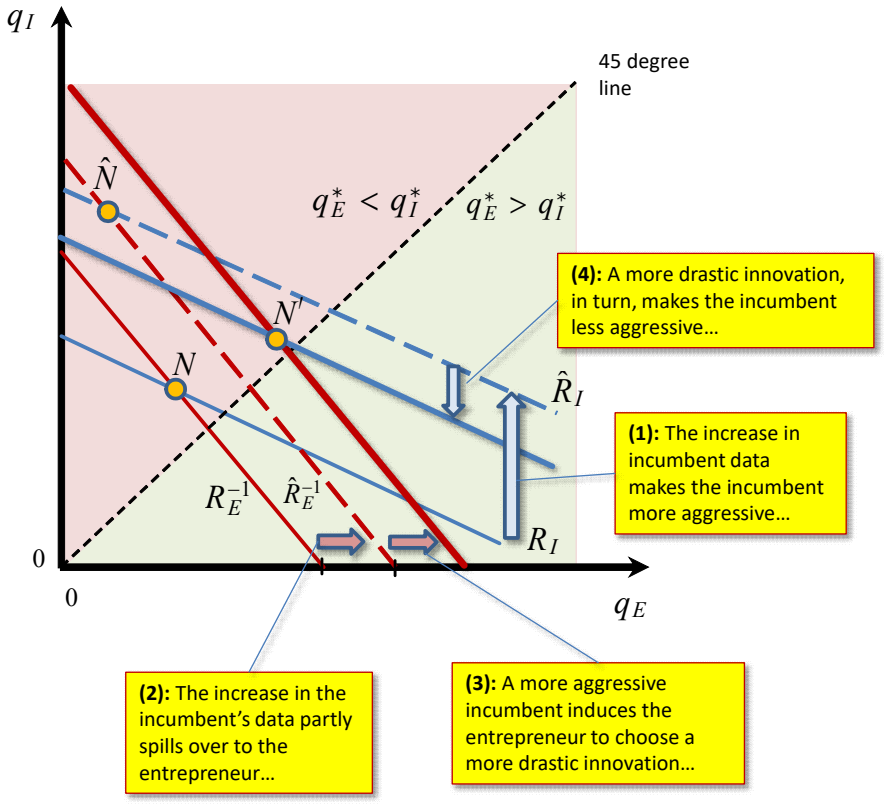


Figure 3.5: Illustrating how the Cournot-Nash-equilibrium in the product market is affected, when the incumbent has access to more historical data and the entrepreneur has weak access to the incumbent's data.

where the reaction function of the entrepreneur shifts further out in Figure 3.5 while the reaction function of the incumbent firm shifts in. In Figure 3.5 the incumbent remains the market leader in the new Nash-equilibrium N'. Can this new Nash equilibrium even move from above the 45-degree line (where $\varphi_E < 1$) to a point below the 45-degree line (where $\varphi_E > 1$) as a consequence of the incumbent having access to more historical data?

Simplify (3.39) further by using (3.16) and (3.17), we can obtain an intuitive condition for when entrepreneurship is destructive:

$$\varphi_E(d_I) > 1 \Leftrightarrow \underbrace{b - \beta\rho_E^*(d_I)}_{\text{Creation effect}} > \underbrace{(1-\gamma)\alpha d_I}_{\text{Big data incumbent advantage effect}}. \tag{3.41}$$

The left-hand side captures how creative the invention is given that it succeeds, while the right-hand side captures how strong the big data advantage is for the incumbent when it is able to

use ML on all its historical data. Note that if the creation effect of the innovation dominates the incumbent advantage effect, the invention is destructive, i.e. $\varphi_E(d_I) > 1$. Figure 3.6 illustrates how the amount of historical data possessed by the incumbent d_I affects the equilibrium market share of the entrepreneur, $\varphi_E(d_I)$. The left-hand side of the figure describes the case of weak access to the incumbent's historical data on the part of the entrepreneur. The right-hand side of the figure the case of strong access to the incumbent's historical data on the part of the entrepreneur.

The left column depicts the case of weak access of the entrepreneur to the incumbent's historical data, $1 - 2\gamma(1-\alpha) > 0$. The right column depicts the case of strong access of the entrepreneur to the incumbent's historical data, $1-2\gamma(1-\alpha) < 0$. Panels in (i) show the optimal project choice in terms of the probability of success as a function of the amount of historical data held by the incumbent, $\rho_E^*(d_I)$. Panels in (ii) show the extra willingness to pay (WTP) that a successful project brings consumers of the entrepreneur's product, i.e. our measure of creative entrepreneurship $b - \beta\rho_E^*(d_I)$, as well as the advantage the incumbent has in better access of data (IA), $(1-\gamma)\alpha d_I$. Panel in (iii) depict our measure of destructive entrepreneurship, i.e. the relative profit of the entrepreneur if she succeeds with her innovation, $\varphi_E(d_I)$.

Strong access to the incumbent's historical data From Proposition 2, we know that if the entrepreneur has strong access to the incumbent's historical data, she chooses a less risky project (a higher success probabilityρ_E^*) in response to the incumbent having more data. This is illustrated by the upward-sloping green curve in right diagram in panel (i) in Figure 3.6. The intuition is that when the entrepreneur has strong access to the incumbent's historical data, an increase in incumbent data wills will strengthen the called the Securing-Success effect (SS effect), that is, increase the cost of failure for the entrepreneur. This induces the entrepreneur to take on less risk (as illustrated in panel (iii) in Figure 3.4).

In the right diagram of panel (ii) in Figure 3.6, we then depict the creation effect, $b-\beta\rho_E^*(d_I)$, and the big data incumbent advantage effect, $(1-\gamma^S)\alpha d_I$, where γ^S indicate strong access, $1-2\gamma^S(1-\alpha) < 0$. The blue curve is the big data incumbent advantage effect, which naturally increases with the amount of historical data d_I for the incumbent. The green curve is the creation effect, which is decreasing in the amount of historical data d_I for the incumbent since—as shown in panel (i)—strong access to the incumbents data induces the entrepreneur to take on less risk, which (when the innovation succeeds), creates a less valuable innovation, leading to a lower

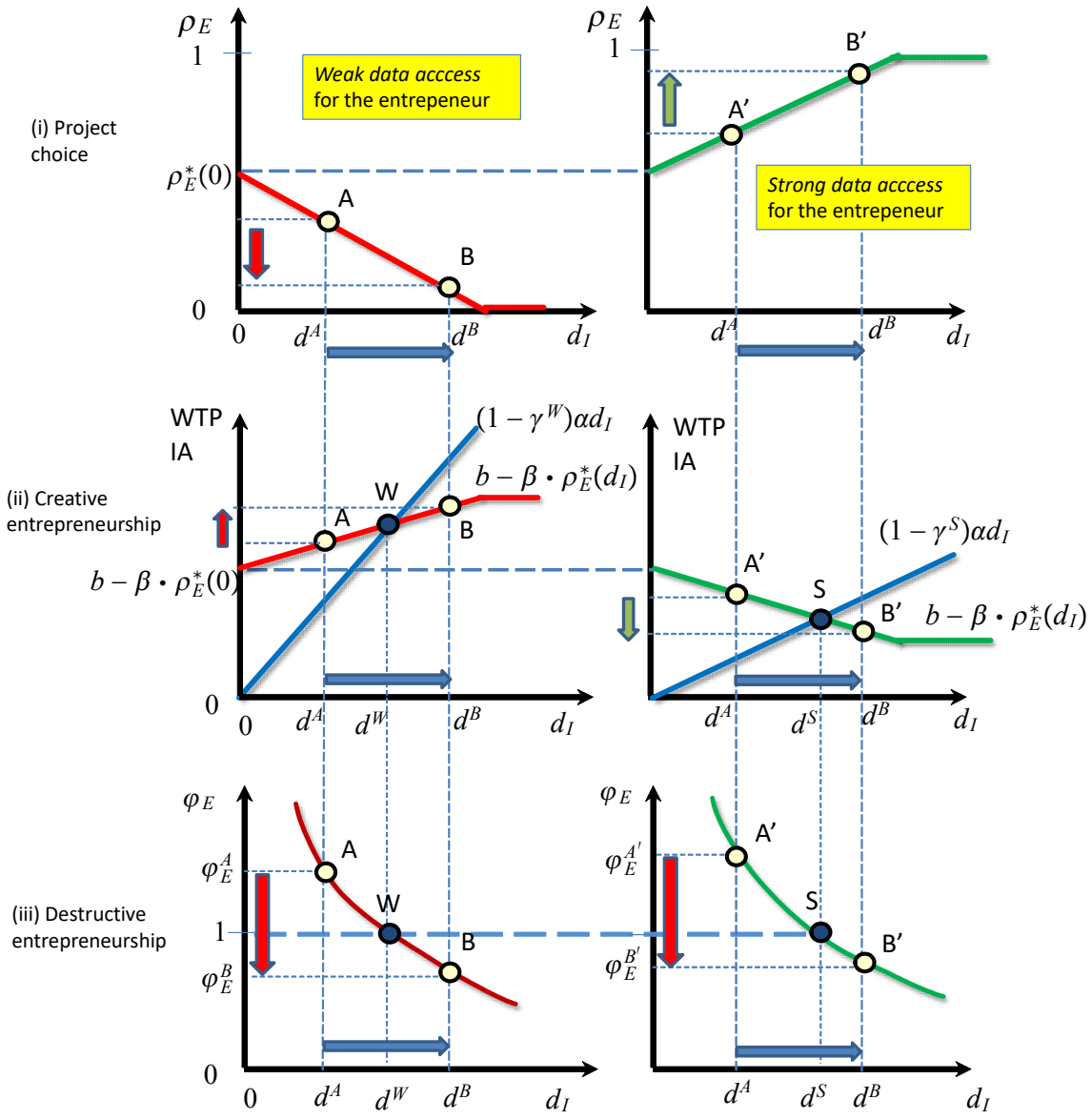


Figure 3.6: The left column depicts the case of weak access of the entrepeneur to the incumbent's historicsl data, $1-2\gamma\left(1-\alpha\right)>0$. The right column depicts the case of strong access of the entrepeneur to the incumbent's historicsl data, $1-2\gamma\left(1-\alpha\right)<0$. Panels in (i) show the optimal project choice in terms of the probability of success as a function of the amound of historicl data held by the incumbent, $\rho_E^*(d_I)$. Panels in (ii) show the extra willingness to pay (WTP) that a succesful project brings consumers of the entrepeneur's product, i.e. our measure of creative entrepeneurship $b-\beta\rho_E^*(d_I)$, as well as the advantage the incumbent has in better access of data (IA), $(1-\gamma)\alpha d_I$. Panel in (iii) depict our measure of destructive entrepreneurship, i.e. the relative profit of the entrepreneur if she succeeds with her innovation, $\varphi_E(d_I)$.

increase in consumer willingness to pay (WTP).

Let d^S be the level of incumbent data which equalizes the incumbent advantage effect and the entrepreneur's creation effect. For low levels of historical data, $d_I \in [0, d^S)$, the creation effect dominates the big data incumbent advantage effect, and the entrepreneur steals business from the incumbent in the product market and becomes the market leader. This is is shown in panel (iii) in the right-hand diagram where $\varphi_E(d_I) > 1$ for $d_I \in [0, d^S)$. At larger amounts of historical incumbent data, $d_I > d^S$, the incumbent advantage effect dominates the entrepreneur's creation effect, and the incumbent remains the market leader. This is as shown in panel (iii) in the right-hand diagram, where $\varphi_E(d_I) < 1$ for $d_I > d^S$.

Weak access to the incumbent's historical data What if the entrepreneur has weak access to the incumbent's historical data? This case is shown on the left side of Figure 3.6. She will then respond to an increase in the incumbent's historical data with a riskier project (a decrease in ρ_E), as shown by the upward-sloping red curve in the left diagram of panel (i) in Figure 3.6. The intuition is now that weak access to incumbents increasing data worsens the entrepreneur's position in the product market which, in turn, softens the securing-success effect (SS) as failure is less costly.

Turning to panel (ii) in Figure 3.6, the blue upward-sloping curve in the left diagram depicts the big data incumbent advantage effect, $(1 - \gamma^S)\alpha d_I$, where γ^W indicate weak access, $1 - 2\gamma^W(1 - \alpha) < 0$. The red curve is the creation effect, $b - \beta\rho_E^*(d_I)$, which is now increasing in the amount of historical data d_I for the incumbent since, as panel (i) shows, weak access to the incumbents data induces the entrepreneur to take on more risk, which (if the innovation succeeds), creates a more valuable innovation leading to a stronger increase in consumer willingness to pay (WTP).

Let d^W is the level of incumbent data which equalizes the incumbent advantage effect and the entrepreneur's creation effect. For low levels of incumbent data, $d_I \in [0, d^W)$, the creation effect again dominates the incumbent advantage effect, and the entrepreneur and becomes the market leader. This is shown in panel (iii) in the left-hand diagram, where $\varphi_E(d_I) > 1$ for $d_I \in [0, d^W)$. At higher amounts of historical incumbent data, $d_I > d^S$, the incumbent advantage effect again dominates the entrepreneur's creation effect, and the incumbent remains the market leader. This is as shown in panel (iii) in the left-hand diagram, where $\varphi_E(d_I) > 1$ for $d_I > d^W$.

Our analyses shows that regardless of the which type of access the entrepreneur has to the

incumbents data, the entrepreneur cannot become the market leader post-entry when the incumbent has access to a sufficient amount of historical data: put differently, when the incumbent has gathered enough data entry is less likely to be destructive. However, the analysis also shows that the entrepreneur responds differently in her choice of innovation project, which leads her to do go for more creative inventions under weak access to the incumbents data and less creative inventions when she has stronger access.

We can summarize with the following proposition:

Proposition 5. *Let d^S and d^W be defined from $b - \beta \cdot \rho_E^*(d^l) = (1-\alpha)\, d^l$ for $l = \{S, W\}$. Suppose that $d^A < \max(d^S, d^W) < d^B$. Then, if the incumbent's data d_I increases from $d_I = d_A$ to $d_I = d_B$,the following holds:*

(i) *Under weak access, $1 - 2\gamma\,(1-\alpha) > 0$,entrepreneurship becomes **more creative**, i.e. $b - \beta \cdot \rho_E^*(d_B) > b - \beta \cdot \rho_E^*(d_A)$ and **less destructive**, i.e. $\varphi_E(d_B) < 1 < \varphi_E(d_A)$*

(ii) *Under strong access, $1 - 2\gamma\,(1-\alpha) < 0$,entrepreneurship becomes **less creative**, i.e. $b - \beta \cdot \rho_E^*(d_B) < b - \beta \cdot \rho_E^*(d_A)$ and **less destructive**, i.e. $\varphi_E(d_B) < 1 < \varphi_E(d_A)$*

In the Appendix we provide numerical illustrations of Proposition 5. We then show how the amount of historical data that the incumbent possesses and the efficiency of ML affects (i) the R&D project choice of the entrepreneur, ρ_E^*, (ii) our measure *creative entrepreneurship*, i.e. the increase in willingness to pay for the entrepreneur's product if the entrepreneur succeeds with her R&D, $\Delta\, a_E|_{Succeed} = b - \beta \cdot \rho_E^*$, and (iii) our measure of *destructive entrepreneurship*, φ_E.

3.3. Stage 1: Becoming an entrepreneur

Let us now close the model and examine the incentive to become an entrepreneur and the way in which this incentive depends on the increasing importance of historical data possessed by the incumbent. To this end, we assume that the entrepreneur faces a fixed research and development cost or investment cost, F, of becoming an entrepreneur. This cost can consist of the cost of evaluating different types of possible business opportunities, the cost of setting up the basics of the business, the opportunity cost of becoming an entrepreneur in the form of forgone wage earnings, etc. In our setting, we can also think of F as a fixed cost for entrepreneurs' to get access to and knowledge of ML technology.

Since the fixed cost F is incurred before the entrepreneur takes her R&D decision, the expected profit for an entrepreneurial venture $\mathrm{E}[\Pi_E]$ becomes

$$\mathrm{E}[\Pi_E] = \rho_E^* \times \pi_E(\rho_E^*) - F. \tag{3.42}$$

We can then examine how the expected profit of the entrepreneur depends on the amount of historical data possessed by the incumbent by differentiating $\mathrm{E}[\Pi_E]$ w.r.t. to d_I:

$$\frac{d\mathrm{E}[\Pi_E]}{dd_I} = \underbrace{\left[\pi_E(\rho_E^*) + \rho_E^* \frac{\partial \pi_E(\rho_E^*)}{\partial \rho_E}\right]}_{=0} \frac{d\rho_E^*}{dd_I} + \rho_E^* \times \frac{\partial \pi_E(\rho_E^*)}{\partial d_I} = \rho_E^* \times \frac{\partial \pi_E(\rho_E^*)}{\partial d_I}, \tag{3.43}$$

where we use the fact that the f.o.c. for the project choice in period 1 implies that $\pi_E(\rho_E^*) + \rho_E^* \frac{\partial \pi_E(\rho_E^*)}{\partial \rho_E} = 0$. Using (3.27) and (3.19), we can rewrite (3.43):

$$\frac{d\mathrm{E}[\Pi_E]}{dd_I} = -2\rho_E^* \times (1-\alpha)\, q_E^*(\rho_E^*) \times \frac{(1-2(1-\alpha)\gamma)\alpha}{(1-2\alpha)(3-2\alpha)} = \begin{cases} < 0; & 1-2(1-\alpha)\gamma > 0: \text{ weak access} \\ > 0; & 1-2(1-\alpha)\gamma < 0: \text{ strong access.} \end{cases} \tag{3.44}$$

We can thus state the following proposition:

Proposition 6. *When the amount of historical data possessed by the incumbent increases, the incentive to become an entrepreneur increases when the entrepreneur has strong access to the incumbent's historical data, i.e., when* $1-2(1-\alpha)\gamma < 0$*, and decreases when the entrepreneur has weak access to the incumbent's historical data, i.e., when* $1-2(1-\alpha)\gamma > 0$*.*

Propositions 5 and 6 points to a policy dilemma. It is likely that the entrepreneur will have limited access to the incumbent's data if data access is unregulated. Proposition 5 then suggest that the trend towards greater big data availability and greater use of ML will lead towards less destructive entrepreneurship—with sustained incumbent market power—but also to less entrepreneurship as entering markets in which incumbents have big data advantages will be less profitable. This suggests a policy which levels the playing field between entrepreneurs and incumbents by forcing incumbents to give entrepreneurs access to their data is motivated. Indeed, such a policy would not only encourage more entrepreneurship—it would also gives rise to more destructive entrepreneurship.

To see this, first differentiate (3.42) in γ to obtain

$$\frac{d\mathrm{E}\left[\Pi_E\right]}{d\gamma} = 2\rho_E^* \times (1-\alpha)\, q_E^*(\rho_E^*) \times \tfrac{2(1-\alpha)\alpha d_I}{(1-2\alpha)(3-2\alpha)} > 0. \tag{3.45}$$

That is, better access for the entrepreneur to the incumbents data will increase the entrepreneur's expected profit from becoming an entrepreneur, which will make it more likely that she invest the fixed cost F in order to take the chance to become a successful entrepreneur.

Better data access will also lead to more destructive entrepreneurship. To see this, recall from (3.41) that entrepreneurship is destructive, $\varphi_E(d_I) > 1$, when the entrepreneur's creation effect, $b - \beta \cdot \rho_E^*$, is greater the incumbent big data advantage effect, $(1-\gamma)\alpha d_I$. From (3.34), it follows that that better data access will weaken the creation effect since the entrepreneur will choose a safer project. However, the decline in creative entrepreneurship is being dominated by a lower incumbent data advantage, that is

$$\underbrace{\left|\frac{\partial\left[b-\beta\cdot\rho_E^*\right]}{\partial\gamma}\right| = \frac{1}{3}\alpha d_I}_{\text{Decline in creative entrepreneurship}} \quad < \quad \underbrace{\alpha d_I = \left|\frac{\partial\left[(1-\gamma)\alpha d_I\right]}{\partial\gamma}\right|}_{\text{Decline in incumbent advantage}}. \tag{3.46}$$

The left-hand side in inequality in (3.46) reveals the drawback in trying to level the playing field between entrepreneurs and incumbents by giving the entrepreneur better data access—this policy weakens the incentives for creative entrepreneurship.

These findings suggest that policies supporting early entrepreneurial ventures might instead be warranted. Subsidizing the fixed cost to become an entrepreneur F would increase $\mathrm{E}\left[\Pi_E\right]$ in (6) and make entry by the entrepreneur more likely without reducing the entrepreneur's incentive to be creative (i.e. to take on more risk). Summing up, we can state the following Proposition:

Proposition 7. *A policy which makes operational data generally available (increasing γ), may be suboptimal: while it may make entrepreneurial entry more likely and increase destructive entrepreneurship (make the entrepreneur the new market leader), it may reduce creative entrepreneurship (create less value for consumers). An alternative or complementary policy, might be to subsidize the fixed R&D or investment cost F (i.e. reduce the cost of becoming an entrepreneur with access to the ML technology). This policy will promote entrepreneurial entry without reducing creative entrepreneurship.*

4. Concluding remarks

In this paper, we have investigated how ML applications and increased amounts of incumbent operational data affect entrepreneurship incentives. In a model, where (i) all firms have access to the same ML application but (ii) incumbents have access to historically generated operational data, we show how increased use of ML on operational data affects entrepreneurial entry and the type of entrepreneurship. In particular, we show that under weak access to incumbent's operational data for entrepreneurs, the development of big data and ML challenges the entrepreneurial process, raising entrepreneurial barriers. However, it is also shown that this process induces entrepreneurs to take on more risk, implying that entrepreneurial activity might become more conducive to high quality entrepreneurship. Thus, we show that machine learning and big data may make creative destruction more creative but less destructive.

Policy implications. This paper has important implications for both entrepreneurs and incumbents. First, entrepreneurs should take into account that challenging incumbents in the era of ML will be more difficult since incumbents' use of ML on previously collected proprietary data makes them more formidable competitors. This implies that entrepreneurs need to become riskier and more creative in the future to find a competitive edge. They may therefore seek support from angels or venture capital firms and use their financing and experience to become more novel in their ventures. Incumbents, conversely, have an incentive to aggressively employ ML application on their operational data so that they become so efficient that they force entrepreneurs to take on so much risk that entrepreneurship will seldom pose a (destructive) threat.

Data protection and data privacy issues have become a flashpoint in the media due in part to high-profile data breaches such as that at Equifax in 2017 and in part to high-profile exposure of Facebook users' personal data to Cambridge Analytica in 2016 and 2017. Partially in response to these events, government regulators have instituted tighter rules on data protection. This has most notably manifested in Europe in the form of the GDPR. The results derived in the paper suggest that policies that making operational data generally available may be suboptimal. The reason is that this can reduce entrepreneurs' willingness to take on risk. An alternative or complementary policy might be to support entrepreneurs' access to and knowledge of ML technology since it stimulates creative entrepreneurship. Subsidizing R&D by small entrepreneurial firms will increase effort but not reduce risk taking.

Limitations. The model has several limitations. The result that entrepreneurs will choose more creative (riskier) projects when ML becomes more efficient and incumbents' proprietary data becomes more important is dependent on the assumption that these strategies do not require excessive financial resources. If they do, our results would be less relevant since if ML becomes sufficiently efficient, it will simply block entrepreneurship entirely. However, the growing venture capital and angel market might relax such restrictions.

A second limitation is that we assume that the incumbent cannot acquire the entrepreneurial firm. A special case would be if a potential acquisition by the incumbent serves the sole purpose of shutting down the invention. While this feature is relevant in some cases, we believe that most sold patents are used efficiently. Indeed, there is ample evidence that many (leading) firms such as Microsoft, Google and Ericsson acquire startups to incorporate them highly efficiently into their businesses. However, to become sufficiently interesting to become a target, entrepreneurs still need to employ a sufficiently creative (risky) strategy.

Future Research. We have treated the human capital of the entrepreneur as a constant in our analysis. As Varian (2018) points out, in the traditional form of learning-by-doing, the learning is passive, but in practice, learning requires active investment in ML machinery and human capital. Thus, quality of human capital and financial capital likely affects ML application and the R&D process. Thus, endogenizing the human capital level in the analysis seems to be a fruitful avenue for future research.

References

[1] Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The Economics of Privacy." Journal of Economic Literature, 54(2): 442-492.

[2] Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki. 2019. "The Impact of Big Data on Firm Performance: An Empirical Investigation." AEA Papers and Proceedings, 109: 33-37.

[3] Batikas, Michail, Stefan Bechtold, Tobias Kretschmer, and Christian Peukert. 2020. "European Privacy Law and Global Markets for Data." London, Center for Economic Policy Research Discussion Paper, 14475. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14475

[4] Bessen, James. 2018. "The Policy Challenge of Artificial Intelligence" (July 25, 2018). CPI Antitrust Chronicle, June 2018, Boston Univ. School of Law, Law and Economics Research Paper No. 18-16. Available at SSRN: https://ssrn.com/abstract=3219887 or http://dx.doi.org/10.2139/ssrn.3219887

[5] Bughin, Jacques, Eric Hazan, Sree Ramaswamy, Michael Chui, Tera Allas, Peter Dahlström, Nicolaus Henke, and Monica Trench. 2017. "Artificial Intelligence the Next Digital Frontier?" McKinsey Global Institute Discussion Paper, June 2017.

[6] Cabral, Luís. 2003. "R&D Competition when firms Choose Variance." Journal of Economics & Management Strategy, 12(1): 139¬¬ – 150.

[7] Campbell, James, Avi Goldfarb and Catherine Tucker. 2015. "Privacy Regulation and Market Structure." Journal of Economics & Management Strategy, 24(1): 47-73.

[8] Choné, Philippe and Laurent Linnemer. 2019. "The quasilinear quadratic utility model: An overview." HAL Working Papers, hal-02318633.

[9] Cohen, Wesley M.. 2010. Chapter 4 - Fifty Years of Empirical Studies of Innovative Activity and Performance. In Hall, Bronwyn H. and Nathan Rosenberg (eds.). Handbook of the Economics of Innovation. North-Holland, Volume 1, Pages 129-213.

[10] Dutton, Tim. 2018. An Overview of National AI Strategies. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

[11] Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. "Big Data and Firm Dynamics." AEA Papers and Proceedings, 109: 38-42.

[12] Färnstrand Damsgaard, Erika, Per Hjertstrand, Pehr-Johan Norbäck, Lars Persson, and Helder Vasconcelos. 2017. "Why Entrepreneurs Choose Risky R&D Projects – But Still Not Risky Enough." The Economic Journal, 127(605): F164-F199.

[13] Gilbert, Richard. 2006. "Looking for Mr. Schumpeter: Where Are We in the Competition–Innovation Debate?" Innovation Policy and the Economy, 6: 159-215.

[14] Haufler, Andreas, Pehr-Johan Norbäck, and Lars Persson. 2014. "Entrepreneurial Innovations and Taxation." Journal of Public Economics, 113: 13-31.

[15] Henkel, Joachim, Thomas Rønde, and Marcus Wagner. 2015. "And the winner is–Acquired. Entrepreneurship as a contest yielding radical innovations." Research Policy, 44(2): 295-310.

[16] Himel, Samuel and Robert Seamans. 2017. "Artificial Intelligence, Incentives to Innovate, and Competition Policy." Antitrust Chronicle, Fall 2017 Vol 1(3).

[17] Jia, Jian, Ginger Zhe Jin, and Liad Wagman. 2021. "The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment." Marketing Science, 40(4): 661-684.

[18] Johnson, Garrett, Scott Shriver, and Samuel Goldberg. 2022. "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR" (January 31, 2022). Available at SSRN: https://ssrn.com/abstract=3477686 or http://dx.doi.org/10.2139/ssrn.3477686

[19] Lambrecht, Anja and Catherine E. Tucker. 2017. "Can Big Data Protect a Firm from Competition?" CPI Chronicle, January 2017.

[20] Martens, Bertin. 2018. "The Importance of Data Access Regimes for Artificial Intelligence and Machine Learning" (December 2018). JRC Digital Economy Working Paper 2018-09. Available at SSRN: https://ssrn.com/abstract=3357652 or http://dx.doi.org/10.2139/ssrn.3357652

[21] Sokol, D. Daniel and Roisin E. Comerford. 2016. "Does Antitrust Have a Role to Play in Regulating Big Data?" (January 27, 2016). Cambridge Handbook of Antitrust, Intellectual Property and High Tech, Roger D. Blair & D. Daniel Sokol editors, Cambridge University Press, Available at SSRN: https://ssrn.com/abstract=2723693

[22] Thompson, Peter. 2010. Chapter 10 - Learning by Doing. In Hall, Bronwyn H. and Nathan Rosenberg (eds.). Handbook of the Economics of Innovation. North-Holland, Volume 1, Pages 429-476.

[23] Rosen, Richard. 1991. "Research and Development with Asymmetric Firm Sizes." The RAND Journal of Economics, 22(3): 411-429.

[24] Varian, Hal. 2018. "Artificial Intelligence, Economics, and Industrial Organization. " National Bureau of Economic Research (NBER) Working Paper, No. 24839.

[25] Web Summit. 2019. In conversation with Margrethe Vestager. https://www.facebook.com/WebSummitHQ/videos/2453514394865115/

A. Appendix

A.1. Creative- and destructive entrepreneurship and big data: An illustration

We here provide numerical illustrations of Proposition 5. We start with the R&D project choice of the entrepreneur, ρ_E^*. We then turn to our measure *creative entrepreneurship*, i.e. the increase in willingness to pay for the entrepreneur's product if the entrepreneur succeeds with her R&D, $\Delta\ a_E|_{Succeed} = b - \beta \cdot \rho_E^*$. Finally, we look at our measure of *destructive entrepreneurship*, where we illustrate that results are qualitatively the same when we use our relative profit as measure of destructive entrepreneurship, φ_E, and when we use a more traditional market share measure. We study how each outcome varies with the amount of historical incumbent data d_I and the state of ML captured by the effectiveness parameter α. As prompted by Proposition 5, for each measure, we compare the two cases of weak- and strong access for the entrepreneur to the incumbent's data.

A.1.1. Risk-taking behavior and big data

We begin by examining how the risk behavior of entrepreneurs in the innovation process depends on the amount of historical data that the incumbent possesses and on the efficiency of ML. Start with the upper panel in Figure A.1 with weak access. Note that risk-taking increases as the entrepreneur is choosing a lower success probabilityρ_E^* when we move in north-east direction. From Proposition 5(i), she takes on more risk when incumbent data increases as successful entry is associated with less value from a weaker SS-effect: From Proposition 4, she also take on more risk when ML becomes more efficient in using these data. When ML becomes more efficient under weak access, the CGS effect is strengthened—it becomes more costly for the entrepreneur to choose a safer project due to the strategic effect of a more aggressive incumbent in the product market.

Then, turn to the lower panel with strong access. From Proposition 5(ii), we know that more incumbent data induces the entrepreneur to reduce her risk-taking as the value of entry increases which strengthens the SS-effect. As shown in the lower panel in Figure A.1, risk-taking now increases in the north-west direction. As indicated by (3.35), the entrepreneur will respond to more efficient ML by choosing more risky R&D projects if the amount of historical data is not too large, again due to a stronger CGS-effect. However, from (3.35) we also note that when the incumbent has abundant historical data we there may be a non-linear effect on risk-taking by the

entrepreneur. Indeed, we can see that when d_I is sufficiently large, an increase in α first induces the entrepreneur to go for safer projects which is then reversed when α becomes sufficiently large.

A.1.2. Creative entrepreneurship and big data

Having illustrated how the amount of historical incumbent data, d_I, and the effectiveness with which this data can be used by ML, α, affect the entrepreneur's risk-taking through her R&D project choice, ρ_E^*, we now turn to have these choices affect creative entrepreneurship. Recall that Definition 3 defines creative entrepreneurship in terms of how much consumer willingness to pay (WTP) increases when the entrepreneur succeeds with her R&D project, $\Delta\ a_E|_{Succeed} = b - \beta \cdot \rho_E^*$. Since the increase in WTP for a successful invention is larger if the entrepreneur succeeds with a more risky project, i.e. with a project with a lower success probability ρ_E^*, the size of the creation effect will be a direct mapping of the entrepreneur's level of risk-taking.

Indeed, comparing Figure A.1 and Figure A.2, we observe that creativity in entrepreneurship maps the R&D risk behavior of the entrepreneur: Under weak access in the top panel in Figure A.2, entrepreneurship is more creative in the northeast direction. Under strong access in the bottom panel in Figure A.2, in contrast, entrepreneurship is more creative in the north-west direction.

A.1.3. Destructive entrepreneurship and big data

Recall that Definition 4 defines destructive entrepreneurship as entrepreneurial entry with successful R&D where the entrepreneur becomes the market leader. Proposition 5 showed that entrepreneurship will be less destructive when the incumbent gets access to more historical data—regardless if the entrepreneur has weak or strong access to these data. This is illustrated in Figure A.3 which uses our relative profitability measure φ_E in (3.38).The top panel shows the case of weak access and the bottom panel shows the case of strong access. More abundant incumbent historical data eventually leads to less destructive entrepreneurship. The impact of more efficient use of these data through more efficient ML again depends on amount of data that the incumbent has access to. In both panels we see a nonlinear pattern where increasing the ML parameter α leading destructive entrepreneurship when the amount of historical data at the incumbent d_I is low whereas the opposite is true when the amount of data possessed by the incumbent is high. In Figure A.4 we the same pattern when we measure destructive

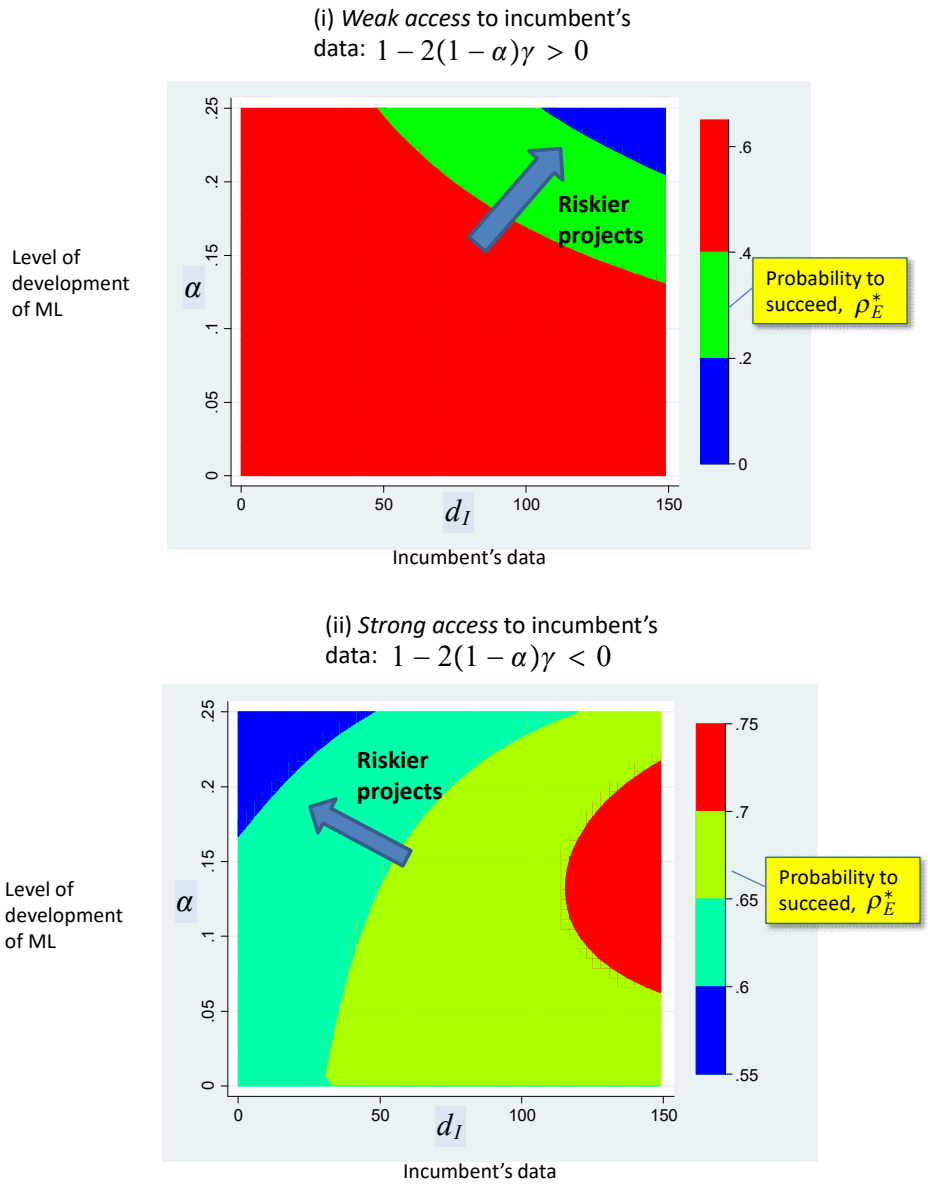


Figure A.1: Illustrating risk-taking by the entrepreneur through her choice of success probability ρ_E^*. Panel (i) shows contours of ρ_E^* as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has weak access to the incumbent's data. Panel (ii) shows contours of ρ_E^* as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has strong access to the incumbent's data. Parameter values set at $\Lambda = 15, b = 12, \beta = 10$ combined with $\gamma = 0.25$ in panel (i) and $\gamma = 0.75$ in panel (ii).

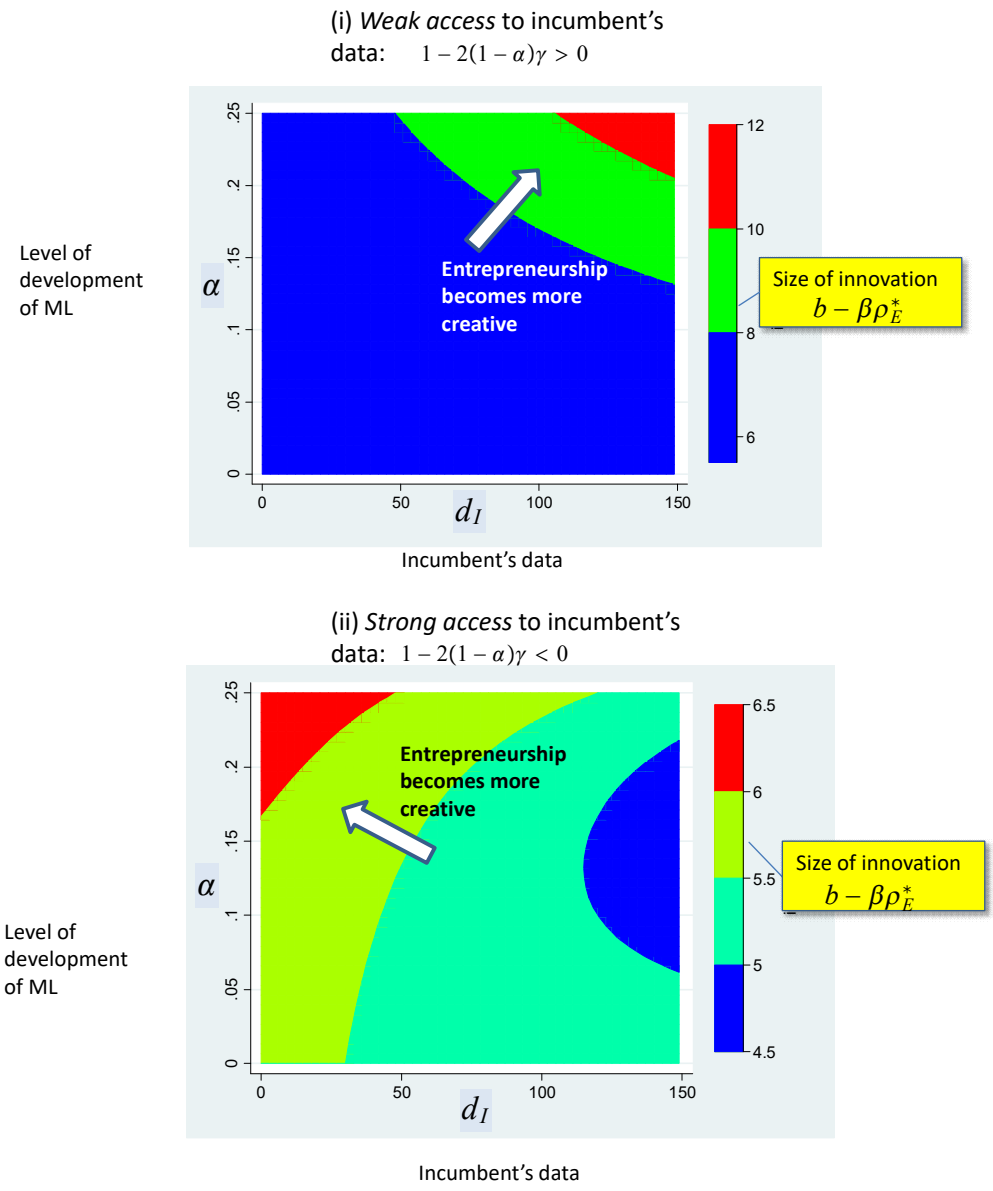


Figure A.2: Illustrating creative entrepeneurship as measured the size of a succesful innovation $b-\beta\cdot\rho_E^*$ in terms if the increase in consumers willingness to pay. Panel (i) shows contours of $b-\beta\cdot\rho_E^*$ as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has weak access to the incumbent's data. Panel (ii) shows contours of $b-\beta\cdot\rho_E^*$ as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine learning (ML), α, when the entrepreneur has strong access to the incumbent's data. Parameter values set at $\Lambda = 15, b = 12, \beta = 10$ combined with $\gamma = 0.25$ in panel (i) and $\gamma = 0.75$ in panel (ii).

entrepreneurship using the more conventional market share measure.

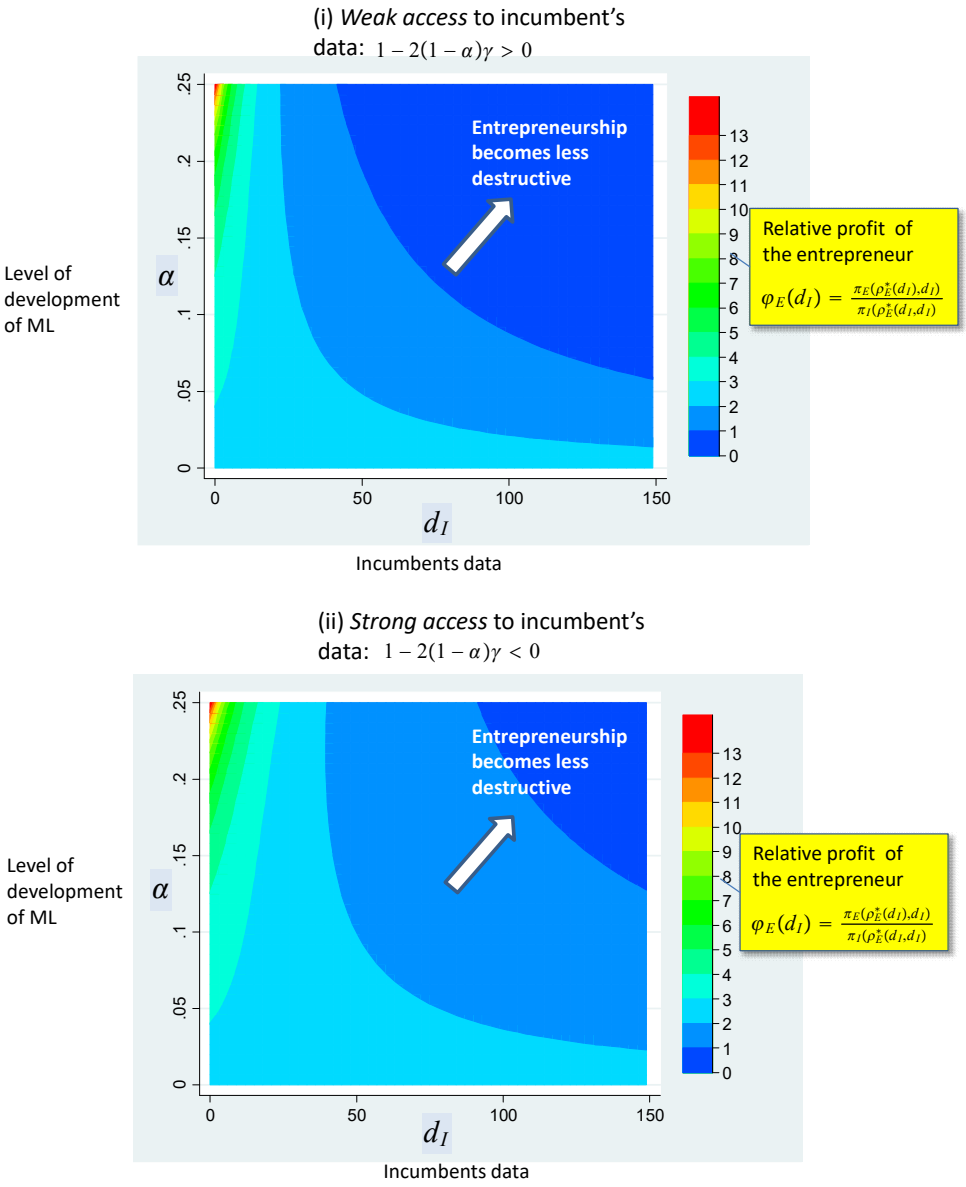


Figure A.3: Illustrating destructive entrepeneurship as measured the relative size the entrepeneur's profit, φ_E. Panel (i) shows contours of φ_E as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has weak access to the incumbent's data. Panel (ii) shows contours of φ_E as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has strong access to the incumbent's data. Parameter values set at $\Lambda = 15, b = 12, \beta = 10$ combined with $\gamma = 0.25$ in panel (i) and $\gamma = 0.75$ in panel (ii).

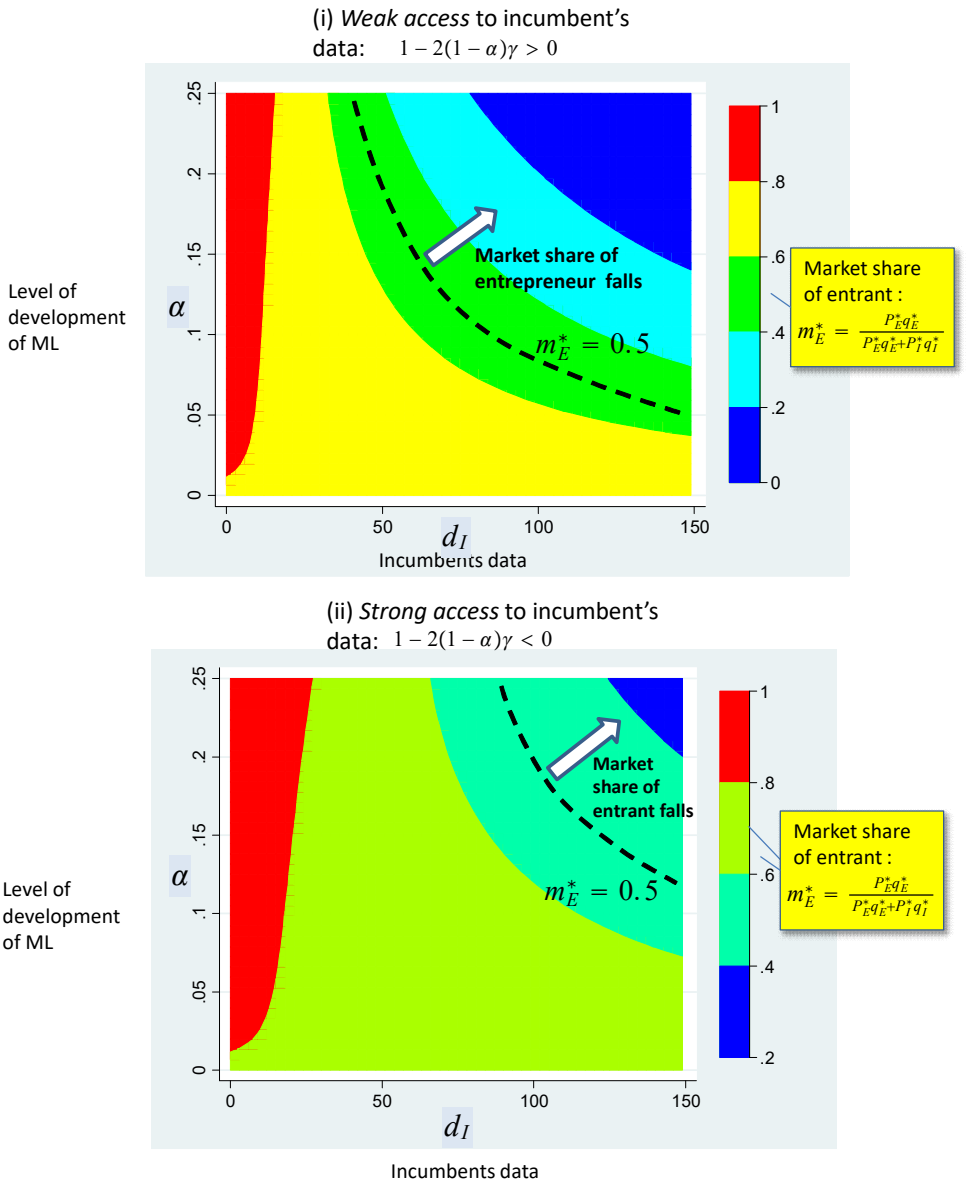


Figure A.4: Illustrating destructive entrepeneurship as measured the market share of the entrepeneur's m_E. Panel (i) shows contours of m_E as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has weak access to the incumbent's data. Panel (ii) shows contours of m_E as a function of the amount of historical data held by the incumbent, d_I, and the effectiveness of machine Learning (ML), α, when the entrepreneur has strong access to the incumbent's data. Parameter values set at $\Lambda = 15, b = 12, \beta = 10$ combined with $\gamma = 0.25$ in panel (i) and $\gamma = 0.75$ in panel (ii).