I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15077

Breakthroughs, Backlashes and Artifcial General Intelligence: An Extended Real Options Approach

Thomas Gries
Wim Naudé

FEBRUARY 2022

DISCUSSION PAPER SERIES

IZA DP No. 15077

Breakthroughs, Backlashes and Artifcial General Intelligence: An Extended Real Options Approach

Thomas Gries
Paderborn University

Wim Naudé
University College Cork, RWTH Aachen University and IZA

FEBRUARY 2022

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Breakthroughs, Backlashes and Artifcial General Intelligence: An Extended Real Options Approach

Breakthroughs and backlashes have marked progress in the development and diffusion of Artificial Intelligence (AI). These shocks make the investment in developing an Artificial General Intelligence (AGI) subject to considerable uncertainty. This paper applies a real options model, extended to account for stochastic jumps, to model the consequences of these breakthroughs and backlashes characterising on investment for an AGI. The model analytics indicate that the average magnitude and frequency of stochastic jumps will determine the optimum amount of time and money to invest in pursuing an AGI and that these may be too expensive and time-consuming for most private entrepreneurs.

JEL Classification: O31, O32, C61, C65

Keywords: artificial intelligence, real option models, radical innovation, risk

Corresponding author:
Wim Naudé
Technology and Innovation Management (TIM)
RWTH Aachen University
Kackertstraße 7
52072 Aachen
Germany
E-mail: wnaude@ucc.ie

1 Introduction

Modern Artificial Intelligence (AI) is the outcome of the availability of big data, accelerated increases in computing power, and advances in Machine Learning (ML)[1]. This type of AI is often labelled "narrow" to distinguish it from a (potential) super-intelligence, more formally labelled an Artificial General Intelligence (AGI) (Naudé, 2021). Although an AGI does not yet exist, some anticipate that it would be only a matter of time before breakthrough innovations would render it possible. Many AI experts expect an AGI to become a reality before 2060.[2]

An AGI would be a genuinely radical innovation, so radical that it could perhaps even be humans' last invention (Good, 1966). Whether it would be desirable, however, is hotly contested. Some hope that it will enable exponential innovation, culminating in a *Singularity* (Kurzweil, 2005), while others have warned that it could wipe humanity out or be misused by whoever invents it first, given its winner-takes-all consequences (Bostrom, 2014; Russel et al., 2015; Armstrong et al., 2016).

Whether an AGI will become a reality before 2060 depends on the extent and sustainability of investment into the broad R&D effort required. So far, developments in AI have traced a cyclical pattern of optimism followed by despair, breakthroughs followed by eventual backlashes. Historical accounts of the development of AI describe both these breakthroughs and backlashes, labelling the latter as "AI winters" and identifying such winters in the 1970s, the 1980s and 1990s (Floridi, 2020). These winters or backlashes were marked by a reduction in investment in R&D and a decline in venture capital (Liu et al., 2017). In 1973, for example, the Lighthill Report in the UK brought to an end a phase of investment in AI when

[1] Machine Learning (ML) refer to the use of algorithms and neural networks "in a narrow domain or application to make and improve predictions by 'learning' from more and more data about a specific domain" (Le Cun, 2015).

[2] See e.g., `https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/` and `https://aiimpacts.org/ai-timeline-surveys/`

it concluded that "in no part of the field have discoveries made so far produced the major impact that was then promised" (Hendler, 2008, p.2). Thus, the failure of scientists to come up with a viable technology resulted in a backlash.

Conversely, the latest AI winter gave way to a "spring" around 2006 following technological breakthroughs in deep learning by Hinton et al. (2006) and Hinton and Salakhutdinov (2006). Since 2012 there has been a boom in AI R&D and venture capital investments (Naudé, 2021). This has fuelled hopes that this investment in AI R&D will ultimately lead to the development of an AGI. For instance Joshi (2019) expressed the expectation that "the rapid rate at which AI is developing new capabilities means that we might be get close to the inflection point when the AI research community surprises us with the development of artificial general intelligence."

The sensitivity of investment in a revolutionary technology such as AI to either breakthroughs a or backlashes is characteristic of most investments subject to substantial uncertainty, especially when coupled with winner-take-all outcomes and possible adverse social consequences. So far, however, there has been little consideration of this in the literature dealing with the possible realisation of an AGI. Rather, the literature has dealt with the technical considerations, for instance debating whether the ML approach will be able to lead to an AGI or whether a completely different tack is needed; outlining how an AGI can be made safe (the control and alignments problems) or discussing the possibility and nature of an arms race for an AGI and how this could be influenced by AI governance (Armstrong et al., 2016; Bostrom, 2017; Naudé and Dimitri, 2020). The AGI literature has been more concerned about aligning the behavior of an AGI in its post-invention period rather than on what determines investment in R&D in the pre-invention period.

Our contribution is to use an extended real options model, as proposed by Bilkic et al. (2013) that incorporates a stochastic Compound Poisson Process to study the pre-invention period in AI R&D. From this we conclude that the investment effort for an AGI may ultimately be too time consuming and too expensive, to be feasible for most private firms. This has implications for the governance of AI and AI arms races, as it suggests that it will be largely government funded agencies or state-owned enterprises efforts (e.g. by the USA or Chinese governments) and/or a few large corporations (such as Google or Alibaba) that will invent an AGI, if ever.

The paper is structured as follows. Section 2 provides an overview of the relevant literature pertaining to investment in AI and real option modelling. Section 3 sets out our extended real options model. In section 4 the model is used to analyse the duration and extent of R&D investment towards an AGI. Section 5 concludes.

2 Relevant Literature

2.1 Breakthroughs, Backlashes and the Uncertainty of Investment in AI

We begin by stressing the difference between incremental and radical innovations. Incremental innovations are "relatively minor changes in technology and provide relatively low incremental customer benefits per dollar" while radical innovations "involve substantially new technology and provide substantially greater customer benefits" (Chandy and Tellis, 1998, p.476). Radical innovations tend to be disruptive (O'Connor and McDermott, 2004). Given that an AGI is a general-purpose technology that could be very disruptive by leading to either a Singularity and/or an Apocalypse, it clearly counts as a "radical" innovation.

Investing in R&D for a radical innovation such as an AGI involves a significant degree of uncertainty - much more so than the uncertainty characterising investment for incremental innovation. It requires that the firm or agency undertaking the R&D investments be prepared to make possibly huge investments in R&D over a sustained period, even if it does not deliver immediate results, with the hope that a breakthrough may eventually occur to make all of the investment worthwhile. As we have mentioned in the introduction, research into AI actively started in the 1950s - and although many advances have been booked, more than seventy years later the development of an AGI has not yet been achieved.

Indeed, during these seventy years there have been several major events, which can be described as "jumps" that have either been negative or positive for the development of an AGI. For example, the Lighthouse report of 1973 sent a major signal to investors that progress in AI was at an dead-end, after which investment in AI collapsed (Hendler, 2008; Liu et al., 2017). Between 1980s and 1987 there was again a boom in AI development, inspired by breakthroughs in the technology of expert systems (Durkin, 1996). By the end of the 1980s a backlash again set it, mainly because the application of expert systems turning out to be too expensive and not profitable. Hundreds of AI-based firms went bankrupt (Newquist, 1994). And since 2006, the boom in AI has been fuelled by technological breakthroughs in the technique of deep learning by Hinton et al. (2006) and Hinton and Salakhutdinov (2006). It is feared by some that this boom will come to an end soon, perhaps due to a backlash against "awful AI," [3] or due to a slowdown in advances in computing power (Moore's Law) (Floridi, 2020; Hao, 2019; Naudé and Dimitri, 2021).

Thus, both positive events such as breakthroughs (e.g. of deep learning) and negative major events such as backlashes due to fears about awful AI, and/or high costs of adoption, growing government regulation of data, or even unmet expectations, may cause the expected value of R&D going into AGI to exhibit major stochastic "jumps." These stochastic jumps takes place against the fact that any investment aiming at a radical innovation would already be characterised by high costs, as a result of the costs of creating complementary inputs or markets to exploit radical innovations, because such R&D investment projects are complex to

[3] See for instance the GitHub site devoted to Awful AI, at `https://github.com/daviddao/awful-ai`

manage, and requires patience and a long-term focus (Freeman, 1994; Marvel and Lumpkin, 2007; Golder et al., 2009, p.474). The high costs and uncertainty of generating a radical innovation explain why it is most often the outcome of large firms' efforts, often with the support of (mission-oriented) governments (Ahuja and Lampert, 2001; Azoulay et al., 2018; Hill and Rothaermel, 2003; Mazzucato, 2015, 2018).

The upshot is that a breakthrough that will lead to an AGI is not guaranteed. In particular since the current state of scientific knowledge about intelligence, the human brain, and related aspects such as consciousness may still be too rudimentary to allow a breakthrough AGI innovation - see for instance the arguments in Meese (2018), Allen and Greaves (2011) and Koch (2012).

2.2 Real Option Models

Real option models are well-established tools to inform inter-temporal optimization choices in the face of irreversible cost in investment, where deferring a decision has value (Dixit, 1989; Dixit and Pindyck, 1994). It therefore seems well suited to study the question of how much and for how long to invest in R&D for an AGI. [4]

Examples of studies using real option theory to study the optimal timing and investment in technology invention and adoption include Farzin et al. (1998), Whalley (2011), Sarkar (2000), Jorgensen et al. (2006) and Doraszelski (2001). For instance, Doraszelski (2001) found that uncertainty about when a new technology will arrive and how disruptive it may be, may lead firms to postpone innovations or technology adoptions. Sarkar (2000) considered circumstances when an increase in uncertainty might cause a firm to accelerate investments in R&D.

These studies, typical of the literature on real options models, model the stochastic process

[4] In this section we draw on Bilkic et al. (2013).

either by a Brownian motion or a simple Poisson process. This may not be appropriate when dealing with significant random jumps associated with breakthroughs and backlashes in AI as described in the previous sub-section. Particularly, one needs to allow for uncertainty about the time of occurrence of a jump (breakthrough or backlash) as well as about its magnitude. As pointed out by Bilkic et al. (2013) the innovation literature has neglected the decision on how long to sustain investment in R&D for radical innovation, particularly in the face of non-trivial stochastic jumps.

Such non-trivial stochastic jumps have been studied in the fields of financial economics, business cycle theory and natural hazards. Merton (1976) introduced jump processes into financial economics to deal with “rare” events that lead to sudden large changes in values of financial assets. Studies on natural hazards have modelled these as sudden (downward) jumps and jump diffusion processes in certain measurements (Cox and Pedersen, 2000; Cox et al., 2004; Yang and Zhang, 2005; Jang, 2007). A discussion of general “jump diffusion processes” and their characteristics are contained in Kou (2002); Kou and Wang (2003), Cai and Kou (2011) and Pham (1997).

In the next section we use the real options model extended by Bilkic et al. (2013)[5] to include stochastic jumps, which we then deploy in section 4 to analyse the magnitude and duration of investment in R&D for an AGI.

3 The Model

Before deriving our model mathematically, we provide an intuitive description.

[5]The working paper by Bilkic et al. (2013) is based on the mathematical modeling extensively described in Bilkic (2014).

3.1 Intuitive Description

When a firm or government scientific agency invest in R&D for a radical innovation such as an AGI it has to overcome two challenges. The first to achieve technical success by developing a workable AGI (as in this case). The second challenge is to commercialize the new technology. successfully. In the case of commercialization, it may not be automatic to assume that an AGI will be profitable: it may be nationalized or restricted through regulations and/or a competing AGI.

The period before invention and commercialization, which is wherein the world is at the moment of writing, is characterized by ongoing investments in technological innovation - for example, in expanding and improving on deep learning algorithms, gathering and improving the quality of data, and boosting computer power further. It is expensive and may take a very long time, as explained in section 2 - the AI R&D stage is already exceeding seventy years.

The firm or agency constantly evaluates whether, and if so when, the AGI is ready to be launched and commercialized. Therefore at any moment, this evaluation could lead the firm or agency to (i) enter the market with the new invention; or (ii) postpone market entry to the future and continue investment in R&D; or (iii) realise that eventual commercialization will be unprofitable - perhaps due to determining in its evaluation that the technology that they have been working on is a dead-end, and/or that a consumer and regulatory backlash is imminent.

In the case of (ii), the firm or agency postpones market entry to invest further in R&D. Further investments may improve the technology. Also, by in effect exercising the option of waiting with market entry, the firm may benefit from changes in the market, such as new complementary technologies and favorable changes in attitudes and regulations on AGI.

The question that our model helps to answer in this regard is, what criteria should the optimizing firm be using when evaluating whether and when to enter the market, or instead to exercise the option of waiting and investing more? or to stop pursuing investment towards radical innovation altogether?

In the rest of this section we describe how our model helps to answer these questions, by focusing on four key decisions: (i) how much accumulated investments to make towards an AGI (sub-section 3.2); (ii) how to evaluate the potential profit at market entry (sub-section 3.3); (iii) how to model the expected profit flow after commercialization (sub-section 3.4); and the (iv) whether to consider the option value of postponing commercialization and continue with R&D investment towards an AGI (sub-section 3.5).

3.2 How much to Invest in an AGI

The time at which a firm or agency commences its R&D investment project aimed at investing an AGI is $t = 0$. This could be at the beginning of a new AI boom phase. The current AI boom phase started roughly around 2007 given availability of data and ICT connectivity (see e.g. (Friedman, 2016)). How much, and for how long to maintain this R&D investment project is, as we discussed in section 2, of unknown duration.

The R&D cost is denoted by C. This cost, which we assume is constant, will differ amongst firms and agencies depending on firm-specific innovation capabilities. Total R&D investment is denoted by $I(T)$, consisting of the sum C incurred during the R&D period. Thus, at time T, when the AGI is launched, the total value investment in R&D is

$$I(T) = \int_0^T Ce^{r(T-t)}dt, \tag{1}$$

Here r denotes the risk-free interest rate.

3.3 Potential Profit at Market Entry

Whilst undertaking R&D, progress in the technological sophistication of the AGI will at each moment in time be associated with a particular level of profit (Y) that can be realised if it should be placed on the market then. This reflects the fact that an AGI invention will be at various states of sophistication - some AGIs can be better than others in terms of quality, safety, and other features (Bostrom, 2017; Armstrong et al., 2016). However, at any period, the profit associated with the state of the AGI, even if positive, may not be adequate. As we mentioned, the profitability of an eventual AGI for the firm or agency developing it will depend on factors such as whether it may be nationalized or restricted through regulations and/or face a competing AGI.

Hence the firm or agency may continue its R&D efforts, perhaps to comply with or by-pass regulations, improve the alignment of the AGI with human interest, or improve the quality and safety to the extent that it will outperform a potential rival's AGI. Each additional period that the firm abstains from market entry, it benefits from the possibility of further technical progress or favorable changes in market conditions. Hence, postponing market entry could potentially be profitable - however, it also does not need to be. For instance, while further R&D could produce a breakthrough that will deliver the AGI, it could also fail, running up against fundamental technological problems, or encountering stricter legislation and regulations. The breakthrough would represent an upward jump in the economic value that will be realised from R&D, and the backlash would represent a downward jump.

Jump processes are needed to model these potential breakthroughs or backlashes / failures in R&D that will cause significant stochastic shocks to the profit (Y) from the AGI. Therefore, the development of potential profits ($\tilde{Y}$) from market entry during any R&D period can be specified an *Compound Poisson Process*:

$$d\tilde{Y} = \tilde{Y} \int\limits_{U_1} u N_1(t, du) \quad for \quad 0 < t < T. \tag{2}$$

Here the stochastic "jump" is the integral $\int\limits_{U_1} z N_1(t, dz)$ where $N_1(t, dz)$ a Poisson Process that has an intensity of λ_1 and U_1 is the step height at which non-trivial stochastic jumps accumulates. It is a Borel Set (Λ) in $\mathbb{R}$ such that $0 \notin \Lambda$.

The essence of equation (2) is to specify the accumulation of stochastic jumps with the implication that the occurrence of breakthroughs and backlashes in this model is unpredictable.

3.4 Profit Flow after Market Entry

When a firm or agency have decided to commercialize its current form of AGI, its profits depend on market conditions - which are highly uncertain as discussed in section 2. The conditions may either be good or bad for profits. For example, if the market contains some other technological invention or resource that are complementary to their AGI technology it will be good for profits. However, market conditions may be characterised by a consumer backlash against an AGI, with governments restricting its free use, which would be bad for profits. It could also be that potential profitability is negatively affected by a macro-economic recession, perhaps due to an unexpected crisis, such as posed by a global pandemic or global financial crisis.

Based on these considerations profit flow after market can be described as following a second Compound Poisson Process:

$$dY = Y \int\limits_{U_2} z N_2(t, dz) \quad \text{for} \quad T < t. \tag{3}$$

In equation (3) the accumulation of stochastic jumps associated with good or bad macroeconomic prospects for profits are denoted by the integral $\int\limits_{U_2} z N_2(t, dz)$. The term $N_2(t, dz)$ describes a Poisson Process with λ_2 intensity. As previously, the stochastic jumps have an uncertain step height of U_2, which is a Borel Set (Λ) in $\mathbb{R}$ such that $0 \notin \Lambda$. It indicates that stochastic jumps are random - and can be positive and negative.

Once the AGI is launched, the profit stream will evolve according to this random process. If we assume a risk neutral firm or agent, the market value of its R&D V^{gross} is given by the expected present value of the profit stream:

$$\begin{aligned} V^{gross} &= \frac{Y(T)}{\left(r - \int\limits_{f^{-1}(U_2)} z\upsilon_2(dz) - \int\limits_{U_2} \left[\ln(1+z) - z\right] \upsilon_2(dz) \right)}; \\ r &> \int\limits_{f^{-1}(U_2)} z\upsilon_2(dz) + \int\limits_{U_2} \left[\ln(1+z) - z\right] \upsilon_2(dz), \end{aligned} \tag{4}$$

with r the risk-free interest rate. f is the function $f(z) = ln(1+z)$ and υ_2 the Lévy Measure of the Poisson Process N_2, which describes the expected number of jumps. The firm or agent undertaking the investment is assumed to have an infinite life.

The expected net value from launching the AGI is calculated as the expected gross value (4) minus the costs of innovation $I(T)$:

$$V = V^{gross} - I(T).$$

3.5 Option Value of Innovation

Although market entry may be profitable at a particular point in time, it may still be better to defer it and continue investment in R&D - deferral may have an option value as we discussed, which we now denote by F. Taking this into account in the Compound Poisson Process (2), results in the Hamilton-Jacobi-Bellman equation:

$$rFdt = E(dF). \tag{5}$$

This indicates that the expected return on investment R&D is equal to the expected rate of capital appreciation over the period dt.

4 Model Analytics

The model described in section 3 can now be used to derive the duration and amount of R&D to be invested in an AGI. First, we derive the duration and implied amount of investment (section 4.1), whereafter we consider the determinants of the duration (section 4.2).

4.1 Expected Time of Market Entry

The firm or agent investing in R&D aiming at an AGI will constantly evaluate (V) against the option value (F) of continued R&D investment (Bilkic et al., 2009).

F is obtained by solving (see Bilkic et al. (2013)):

$$\max\left\{V^{gross}(T) - I(T), F(T)\right\}. \tag{6}$$

This shows that the expected time of market entry results from first determining the level of profits (the threshold Y^*) that will trigger such entry. When this threshold is reached the value of current R&D exceeds (F) (Bilkic et al., 2009). Consequently, market entry is indicated.

The expected time at which market entry will occur will secondly be determined by whether the pre-market entry value $\tilde{Y}$ is equal or greater to this threshold value or not.

This leads us to recall proposition 1 from Bilkic et al. (2013):

Proposition 1 *For the costs of innovation I, a pre-market entry value of current profits* (2)*, and future profit streams after market entry as per* (3)*, we can determine the threshold Y^* that would trigger the commercialization of the AGI invention*

$$
\begin{aligned}
Y^* &= \frac{\beta}{\beta - 1}\left[r - \int\limits_{f^{-1}(U_2)} z\upsilon_2(dz) - \int\limits_{U_2}\left[\ln(1+z) - z\right]\upsilon_2(dz)\right] I(T) \qquad (7)\\
\textit{with } r &> \int\limits_{f^{-1}(U_2)} z\upsilon_2(dz) + \int\limits_{U_2}\left[\ln(1+z) - z\right]\upsilon_2(dz)
\end{aligned}
$$

β is an implicit function resulting from the differential equation $rFdt = E(dF)$ with solution $F = B\tilde{Y}^{\beta}$.

Proof. see Appendix 1. ■

In determining the expected time at which the pre-market entry value ($\tilde{Y}$) will reach the threshold (7) and trigger market entry, we use the double exponential distribution of Kou and Wang (2003) as for this an analytical solution is possible in the event that overshooting occurs. It can be written as

$$h(z) = p\eta_1 e^{-\eta_1 z} 1_{\{z \geq 0\}} + q\eta_2 e^{\eta_2 z} 1_{\{z<0\}},$$

where p is the probability of *breakthrough* and q of a *backlash*. It is the case that $p+q=1$. $\frac{1}{\eta_1}$ and $\frac{1}{\eta_2}$ are the means of the two exponential distributions. These are distributions of deferment of entry, given breakthroughs or backlashes.

In order to analytically determine the expected time at which the pre-market entry value $(\tilde{Y})$) is expected to reach the threshold, we allow for a non-linear development of the threshold. As in Bilkic et al. (2009) and Bilkic et al. (2013) for the random process ($\tilde{Y}$ see (2)) we derive the expected time of first realizing or overshooting a certain market entry profit $\tilde{Y}_i$ (given the existing value $\tilde{Y}_0$). By calling on the Girsanov Theorem we can derive a probability density function for $\tilde{T}_i$. Given that we can determine the expected time of first realizing or overshooting of each value of the strictly monotonic increasing sequence $\tilde{Y}_i \in \left\{\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_n\right\}$ with $0 = \tilde{Y}_1 \leq \tilde{Y}_2 \leq \ldots \leq \tilde{Y}_n$ and $\tilde{Y}_1 > \tilde{Y}_0$ we can write the expected first realization or overshooting time as a function of $\tilde{Y}_i$ and of the overall drift of the jump process $\bar{u} = \lambda_1 \left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)$.

Proposition 2 *Using the Compound Poisson Process (2) we can derive the expected time of first realization or overshooting $E\tilde{T}$ of each market profit $\tilde{Y}_i \in \left\{\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_n\right\}$ with $0 = \tilde{Y}_1 \leq \tilde{Y}_2 \leq \ldots \leq \tilde{Y}_n$ and $\tilde{Y}_1 > \tilde{Y}_0$ as a function of $\tilde{Y}_i$, and hence determine the expected time until any market entry profit $\tilde{Y}$ is reached for the first time as:*

$$E(\tilde{T}) = \frac{1}{\bar{u}} \left[\tilde{Y}_i + \frac{\mu_2^* - \eta_1}{\eta_1 \mu_2^*}(1 - e^{-Y^* \mu_2^*})\right]. \tag{8}$$

$\bar{u}$ refers to the overall drift $\bar{u} = \lambda_1 \left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)$ of the jump process and μ_2^* is a constant derived in the technical note, for which $0 < \eta_1 < \mu_2^* < \infty$ holds.

Proof. see Appendix 1. ■

To determine the the expected time at which the pre-market entry value $(\tilde{Y})$) is expected to reach the threshold T^* we have to compare all points $(\tilde{Y}_i, E(\tilde{T}, \tilde{Y}_i))$ and (Y_i, T) from the sets of the threshold as well as from the ET_i set and choose the point that is included in both sets.

The resulting T^* determines the expected time at which the pre-market entry value $(\tilde{Y})$) is expected to attain the threshold. As the image set of ET_i is a sequence there may be no exact match with the threshold. In this case we choose the first point in time for which the threshold exceeds ET_i, that is $T(Y) > E(\tilde{T}, \tilde{Y}_i)$.The next proposition provides conditions for the existence of a solution to the expected time at which the pre-market entry value $(\tilde{Y})$) is expected to cross the threshold.

Proposition 3 *With the threshold $Y^*(T)$ (see (9)), the expected first-time realization of initial market profit $E\tilde{T}$ (see (8)), and condition (10) and (11) there exists an expected time to enter the market $T^* = E(T) > 0$. T^* is the first time point for which the following conditions hold:*

$$T^*(Y) \geq E(\tilde{T}, \tilde{Y}_i). \tag{9}$$

$$\frac{1}{r}\ln\frac{\tilde{Y}_1 + KC}{KC} < \frac{1}{\bar{u}}\left[\tilde{Y}_1 + \frac{\mu_2^* - \eta_1}{\eta_1\mu_2^*}(1 - e^{-\tilde{Y}_1\mu_2^*})\right], \tag{10}$$

$$\frac{1}{r}\ln\frac{\left(\tilde{Y}_t + KC\right)\left(\tilde{Y}_i + KC\right)}{\left(\tilde{Y}_s + KC\right)\left(\tilde{Y}_j + KC\right)} < \frac{1}{\bar{u}}\left[\begin{array}{c}\tilde{Y}_t + \tilde{Y}_i - \tilde{Y}_s - \tilde{Y}_j \\ +\frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(e^{-\tilde{Y}_s\mu_2^*} + e^{-\tilde{Y}_j\mu_2^*} \\ -e^{-\tilde{Y}_t\mu_2^*} - e^{-\tilde{Y}_i\mu_2^*})\end{array}\right] \tag{11}$$

with $K = \frac{\beta}{\beta-1}(r - \int\limits_{f^{-1}(U_2)} z\upsilon_2(dz) - \int\limits_{U_2}\left[\ln(1+z) - z\right]\upsilon_2(dz) - \alpha)$.

Proof. See Appendix 1. ■

The existence of T^* means that at any period a firm or agency could consider deferring entry, and continue to invest in R&D. Because the threshold is higher than the expected market profit it means that innovation costs during the R&D phase (preceding T^*) and the option value to enter the market is not yet compensated by the current value of R&D.

In addition, condition (10) illustrates the logic of the decision problem.[6] The decision in favor of initiating a research process for a certain time will only be positive if the minimum profit is sufficiently small compared to the R&D costs at the beginning, and if the time path of the research value can be expected to attain the profit threshold as the result of the research activity.

Further, T^* indicates the expected duration of research under present conditions. T^* is the answer to the question of how long a firm or agency will invest in R&D. Second, with T^* the firm or agency aiming for an AGI can also determine their expected total investment volume $I(T^*)$.

4.2 Determinants of Expected Time of Market Entry

In the previous sections we showed that high uncertainty or stochastic shocks, and irreversibility, are important ingredients of a radical innovation process such as that for an AGI. In this section we provide a closer examination of the effects of such stochastic shocks. In particular, we consider the frequency of jumps, and the effect of the jump-size during the innovation period, when we can already identify a market entry but when the firm has not yet decided to commercialize its available AGI.

Proposition 4 *An increase in the frequency of sudden breakthroughs or backlashes during the R&D period is generally ambiguous. However, an increase in λ_1 may lead to an earlier market entry ET^* if $\frac{p}{\eta_1} - \frac{q}{\eta_2} > 0, 1 < \frac{\mu_2^* - \eta_1}{\eta_1} e^{-Y^* \mu_2^*}$ and the sum of upward jumps is sufficiently*

[6] Both conditions are required for the existence of a solution to the problem.

large to outweigh of the sum of negative jumps so that $\frac{\partial\beta}{\partial\lambda_1} < 0$.

$$\frac{\partial E(T^*)}{\partial\lambda_1} = -\frac{\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}{\left[\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)\right]^2}\left[Y^* + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-Y^*\mu_2^*})\right] + \frac{1}{\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}\left[\begin{array}{c} -\frac{\frac{\partial\beta}{\partial\lambda_1}}{(\beta-1)^2}(r - \int\limits_{f^{-1}(U_2)} z\upsilon(dz) \\ -\int\limits_{U_2}[\ln(1+z)]\,\upsilon_2(dz)) \end{array}\right] \cdot(1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}) < 0.$$

Proof. see Appendix 2. ■

The conditions $\frac{p}{\eta_1} - \frac{q}{\eta_2} > 0$ and $1 < \frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}$ are connected to the probabilities of sudden breakthroughs or backlashes and the mean waiting times until an event occurs. The decision whether to enter the market if the frequency of jumps increases strongly depends on the direction of those jumps, and therefore whether a sudden breakthrough or a backlash happens. In general, an increase in λ_1 implies that more fundamental events are occurring. Hence, sudden positive discoveries become more frequent and the increase in the value of research results accelerates.

Proposition 5 *An increase in the magnitude of breakthroughs or backlashes during the R&D period is generally ambiguous. However, an increase in u leads to an earlier market entry if* $\frac{p}{\eta_1} - \frac{q}{\eta_2} > 0, 1 < \frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}$ *and the sum of upward jumps is sufficiently large to outweigh of the sum of downward jumps so that* $\frac{\partial\beta}{\partial u} < 0$.

$$\frac{\partial E(T^*)}{\partial u} = \frac{(1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*})}{\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}\left[\frac{-\frac{\partial\beta}{\partial u}}{(\beta-1)^2}\left(\begin{array}{c} r - \int\limits_{f^{-1}(U_2)} z\upsilon(dz) \\ -\int\limits_{U_2}[\ln(1+z)-z]\,\upsilon_2(dz) \end{array}\right)I(T)\right] < 0.$$

Proof. See Appendix 2. ■

An increase in u means that breakthroughs become more beneficial and threats less disastrous. Larger upward jumps suggest that research steps are larger up to a sudden breakthrough and hence a successful market entry can be expected earlier. Hence, discoveries become more important and have a more significant impact, with just one new research result potentially resulting in the final invention needed to trigger market entry.

While the these two effects result from large uncertain events during the R&D phase, we next discuss two effects that result from large uncertain events after market entry and affects the investment decision.

Proposition 6 *An increase in the frequency of sudden breakthroughs or backlashes after market entry is generally ambiguous. However, an increase in* λ_2 *may lead to an earlier market entry* ET^* *if* $\frac{p}{\eta_1} - \frac{q}{\eta_2} > 0$, $\frac{\mu_2^* - \eta_1}{\eta_1} e^{-Y^* \mu_2^*} > 1$ *and the sum of upward jumps is sufficiently large to outweigh the sum of negative jumps.*

$$\frac{\partial E(T^*)}{\partial \lambda_2} = \frac{1}{\lambda_1 \left(\frac{p}{\eta_1} - \frac{q}{\eta_2} \right)} \left[\frac{\beta}{\beta - 1} \left(\begin{array}{c} - \int\limits_{f^{-1}(U_2)} z h(dz) \\ - \int\limits_{U_2} \left[\ln(1+z) - z \right] h(dz) \end{array} \right) I(T) \right] \left(1 - \frac{\mu_2^* - \eta_1}{\eta_1} e^{-Y^* \mu_2^*} \right) > 0.$$

Proof. See Appendix 2. ■

As we point out in the Appendix, the sign of the derivative according to λ_2 is ambiguous. However, we assume that breakthroughs are frequent and large enough to outweigh the sum of backlashes such that market entry will be preferred. More positive jumps indicate that more breakthroughs than threats and losses can be expected after market entry. Having invented an AGI and launched it on the market, new information may reveal additional and unexpected applications for the AGI. Upon market entry, information about the new product will spread, enabling the firm to expand the uptake of the AGI. This will make market entry more attractive, so the firm or agency may then want to try and enter the market with its first viable AGI prototype.

Proposition 7 *An increase in the magnitude of breakthroughs and backlashes after market entry is ambiguous. However, an increase in* z *leads to an earlier market entry if* $\frac{p}{\eta_1} - \frac{q}{\eta_2} > 0$, $\frac{\mu_2^* - \eta_1}{\eta_1} e^{-Y^* \mu_2^*} > 1$ *and the sum of upward jumps is sufficiently large to outweigh the sum of downward jumps.*

$$\frac{\partial E(T^*)}{\partial z} = \frac{1}{\lambda\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)} \left[\begin{array}{c} -\underbrace{\frac{\beta}{\beta - 1}}_{>0}\Big(\int\limits_{f^{-1}(U_2)} 1 \upsilon_2(dz) \\ + \int\limits_{U_2} \left[\frac{1}{(1+z)} - 1\right] \upsilon_2(dz)\Big) \underbrace{I(T)}_{>0} \end{array} \right] \left(1 - \frac{\mu_2^* - \eta_1}{\eta_1} e^{-Y^* \mu_2^*}\right) < 0$$

Proof. See Appendix 2. ■

An increase in z mean that opportunities become more beneficial and threats less disastrous. Larger upward jumps mean that the benefits from opportunities provided by market entry increase. At the same time, smaller downward jumps reduce the loss in profits generated by threats. Hence, opportunities are more beneficial than threats so that market entry is expected to have a greater payoff. We observe that further R&D becomes less attractive because the existing AGI prototype, even if not perfect, and perhaps subject to safety risks, promises such commercialization success that an earlier market entry is preferred.

From the model analytics described here we can conclude that stochastic shocks have a determining impact on the decision to invest in R&D for an AGI. The extent of R&D may be prolonged or terminated, potentially affecting the safety of an AGI, depending on such stochastic events.

5 Concluding Remarks

The firms or agents that invest in R&D in search of an AGI, which can potentially yield winner-take-all-benefits, need to take strategic decisions with respect to the possibility of the occurrence (or not) of stochastic events such as breakthroughs and backlashes. The key strategic decisions for a firm or agent that wishes to be a successful developer and provider of an AGI are a) how much to invest in R&D; and b) when to bring the AGI prototype innovation to market.

In this paper, we used an extended real options model to derive the optimal entry timing and show that stochastic jumps such as breakthroughs and backlashes affect both the investment amount and market entry decisions of firms and agents that aim to develop an AGI. Our contribution has been twofold: first, our paper considered the economic dimensions of developing an AGI - in contrast to most of the literature where considerations of technical and ethical feasibility dominate; and second by using a novel extended real options model that includes non-marginal stochastic jump processes to discuss the effects of such large uncertain events on the AGI investment decision.

From our model analytics we determined that the average magnitude of sudden events and the direction (up or down) and frequency of the associated jumps are the most important parameters in this highly uncertain decision process. These stochastic shocks imply that investment in an AGI will be too time consuming and expensive for most private entrepreneurs. Hence, if ever an AGI will be invented, it will most likely be either through the R&D investment efforts of a government, such as the USA or Chinese governments, or the efforts of a very large private firm with deep pockets and other resources for innovation in the AI field, such as *Google* or *Alibaba*.

The downside of this is that whether or not an AGI will ever be invented will depend on only a few efforts and may suggest that a *Singularity* will not be reached any time soon. The upside however is that an arms race for an AGI, which may result in an inferior and unsafe AGI being pushed into the market, may perhaps be easier to avoid that previously thought, as the coordination and collaboration possibilities are more feasible with fewer players in the markets. Governments may also threaten the post-commercialization profits of such firms and agencies thus to put pressure on them to postpone launching the AGI until such a date as a satisfactory quality can be ensured. Finally, it will be easier to govern and check the process of development of the AGI if the appropriate regulatory and transparency measures can be agreed on at a global level.

Appendices

Appendix 1: Expected Time of Market Entry

Proof of Proposition 1. Apply the boundary conditions

$$\begin{aligned} F(0) &= 0 \\ F(Y^*) &= V^{gross}(Y^*) - I \qquad \textit{value matching condition,} \\ \frac{dF(Y^*)}{dY} &= \frac{d(V^{gross}(Y^*) - I)}{dY} \qquad \textit{smooth pasting condition.} \end{aligned}$$

and solve the equation system for Y^*. ■

Proof of Proposition 2. The problem of determining the the expected time at which the pre-market entry value $(\tilde{Y})$) is expected to reach the threshold T^* can be solved analytically if we assume an explicit distribution of the jump sizes. Following Kou and Wang (2003) we assume the double exponential distribution

$$h(z) = p\eta_1 e^{-\eta_1 z} 1_{\{z \geq 0\}} + q\eta_2 e^{\eta_2 z} 1_{\{z<0\}},$$

where p is the probability of a positive jump (a breakthrough) and q for a negative jump (backlash), respectively. $\frac{1}{\eta_1}$ and $\frac{1}{\eta_2}$ are the means of the two exponential distributions. The moment generating function for $\tilde{Y}(t)$ with $\theta \in (-\eta_2, \eta_1)$ is

$$\phi(\theta, t) := E(e^{\theta \tilde{Y}(t)}) = \exp(G(\theta)t)$$

where the function G is defined as

$$G(x) := \lambda_1 \left(\frac{p\eta_1}{\eta_1 - x} + \frac{p\eta_2}{\eta_2 + x} - 1 \right).$$

To determine the the expected time at which the pre-market entry value $(\tilde{Y})$) is expected to reach the threshold T^* times one has to consider the exact hit of a constant boundary, as well as overshooting. Accordingly, two cases can be distinguished. The Laplace Transformation of the first hitting time, which is when $\tilde{Y}(t)$ hits the boundary Y^* exactly is :

$$E(e^{-\varepsilon \tilde{T}_i} 1_{\{\tilde{Y}(\tilde{T}_i) = Y^*\}}) = \frac{\eta_1 - \beta_{1,\varepsilon}}{\beta_{2,\varepsilon} - \beta_{1,\varepsilon}} e^{-Y^* \beta_{1,\varepsilon}} + \frac{\beta_{2,\varepsilon} - \eta_1}{\beta_{2,\varepsilon} - \beta_{1,\varepsilon}} e^{-Y^* \beta_{2,\varepsilon}}$$

with $\mu_{1,\varepsilon}$ and $\mu_{2,\varepsilon}$ being the only positive roots of $G(\beta) = \varepsilon$ and $0 < \beta_{1,\varepsilon} < \eta_1 < \beta_{2,\varepsilon} < \infty$. For every overshoot $\tilde{Y}(\tilde{T}) - Y^*$ the Laplace Transformation is

$$E(e^{-\varepsilon\tilde{T}}1_{\left\{\tilde{Y}(\tilde{T}_i)-Y^*>y\right\}}) = e^{-\eta_1 y}\frac{(\eta_1 - \mu_{1,\varepsilon})(\mu_{2,\varepsilon} - \eta_1)}{\eta_1(\mu_{2,\varepsilon} - \mu_{1,\varepsilon})}\left(e^{-Y^*\mu_{1,\varepsilon}} - e^{-Y^*\mu_{2,\varepsilon}}\right) \quad \text{for all } y \geq 0.$$

The expectation of the expected time at which the pre-market entry value ($\tilde{Y}$)) is expected to reach the threshold T^* is finite, i.e. $E(T^*) < \infty$, if and only if the overall drift of the jump process is positive. Hence,

$$E(T^*) < \infty \Leftrightarrow \bar{u} = \lambda_1\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right) > 0.$$

Now for $\bar{u} > 0$ we determine the expected time at which the pre-market entry value ($\tilde{Y}$)) is expected to reach the threshold T^* as

$$E(T^*) = \frac{1}{\bar{u}}\left[Y^* + \frac{\mu_2^* - \eta_1}{\eta_1\mu_2^*}(1 - e^{-Y^*\mu_2^*})\right] \tag{12}$$

where μ_2^* is defined as the unique root of $G(\mu_2^*) = 0$ with $0 < \eta_1 < \mu_2^* < \infty$. ■

Proof of Proposition 3. For each $\tilde{Y}_i$ we can determine the corresponding expected time $E(\tilde{T}_i)$ when this market entry profit $\tilde{Y}_i$ is reached for the first time. In order to find the expected time of market entry we have to consider all combinations of profit levels $\tilde{Y}_i$ (and, Y_i respectively) and the required time to reach this level $E(\tilde{T}_i)$ (T_i, respectively) from the expected time at which the pre-market entry value ($\tilde{Y}$)) is expected to reach the threshold T^* and the threshold function. That is, we compare the image sets of those functions and choose the point in time which refers to the exact hit or overshoot of the threshold as the expected time of market entry. The proof for the existence of this point is as follows.

Assume that $\tilde{Y}_i$ is an element of the strictly monotonic increasing sequence $\left\{\tilde{Y}_0, \tilde{Y}_1, ..., \tilde{Y}_n\right\}$ with $0 = \tilde{Y}_0 \leq \tilde{Y}_1 \leq \ldots \leq \tilde{Y}_n$, so that $E(\tilde{T})$ can be written as a function of any $\tilde{Y}_i$. In this case $E(\tilde{T}, \tilde{Y}_i)$ is a strictly monotonic increasing sequence as well, and all pairs of values $(\tilde{Y}_i, E(\tilde{T}, \tilde{Y}_i))$ form its image set. In order to derive the expected time before market entry, which is the time at which $\tilde{Y}$ reaches the time-dependent threshold Y^* for the first time, we have to prove that there exists a point $(\tilde{Y}_i, E(\tilde{T}, \tilde{Y}_i))$ which is also in the image set of the function $T(Y)$. $T(Y)$ is determined by the threshold curve

$$\begin{aligned}
Y^*(T) &= \underbrace{\frac{\beta}{\beta-1}(r - \int\limits_{f^{-1}(U_2)} z\upsilon_2(dz) - \int\limits_{U_2} [\ln(1+z) - z]\,\upsilon_2(dz) - \alpha)}_{=:K} C\left(e^{rT}-1\right) \\
&= KC\left(e^{rT}-1\right) = KCe^{rT} - KC
\end{aligned}$$

$$\Rightarrow T^*(Y) = \frac{1}{r}\ln\frac{Y+KC}{KC}.$$

The proof is as follows. We analyze the functions $E(\tilde{T}, \tilde{Y}_i)$ and $T(Y)$ near the origin and show that one function lies above the other. Next we can show that although both curves increase, the increasing rate of one function decreases faster, leading to a image point which may be in both image sets. Hence, we consider the two functions in 0.

$$T(0) = \frac{1}{r}\ln\frac{0+KC}{KC} = 0$$

and

$$E(\tilde{T}, 0) = \frac{1}{\bar{u}}\left[0 + \frac{\mu_2^* - \eta_1}{\eta_1\mu_2^*}(1 - e^{-0\mu_2^*})\right] = 0.$$

As both functions start in 0 we consider their increment of growth between 0 and $\tilde{Y}_1$.

$$\begin{aligned}
&\frac{\frac{1}{r}\ln\frac{\tilde{Y}_1+KC}{KC} - \frac{1}{r}\ln\frac{0+KC}{KC}}{\tilde{Y}_1} \\
=\ &\frac{\frac{1}{r}\ln\frac{\tilde{Y}_1+KC}{KC}}{\tilde{Y}_1}
\end{aligned}$$

$$\begin{aligned}
&\frac{\frac{1}{\bar{u}}\left[\tilde{Y}_1 + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_1\mu_2^*})\right] - \frac{1}{\bar{u}}\left[\tilde{Y}_1 + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_1\mu_2^*})\right]}{\tilde{Y}_1} \\
=\ &\frac{\frac{1}{\bar{u}}\left[\tilde{Y}_1 + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_1\mu_2^*})\right]}{\tilde{Y}_1}.
\end{aligned}$$

The increment of growth of $T(Y)$ is smaller than the increment of growth of $E(\tilde{T}, Y)$ for

$$\frac{1}{r}\ln\frac{\tilde{Y}_1+KC}{KC} < \frac{1}{\bar{u}}\left[\tilde{Y}_1 + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_1\mu_2^*})\right].$$

Furthermore, we analyze the change in the increments of growth with the second difference quotient. For any $\tilde{Y}_i, \tilde{Y}_j, \tilde{Y}_s, \tilde{Y}_t$ with $\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_s < \tilde{Y}_t$ and $\tilde{Y}_i \neq 0$ the second difference quotient is

$$
\begin{aligned}
& \frac{\frac{1}{r}\ln\frac{\tilde{Y}_t+KC}{KC} - \frac{1}{r}\ln\frac{\tilde{Y}_s+KC}{KC} - \frac{1}{r}\ln\frac{\tilde{Y}_j+KC}{KC} + \frac{1}{r}\ln\frac{\tilde{Y}_i+KC}{KC}}{\tilde{Y}_t - \tilde{Y}_i} \\
= & \frac{\frac{1}{r}\ln\frac{\tilde{Y}_t+KC}{\tilde{Y}_s+KC} - \frac{1}{r}\ln\frac{\tilde{Y}_j+KC}{\tilde{Y}_i+KC}}{\tilde{Y}_t - \tilde{Y}_i} \\
= & \frac{\frac{1}{r}\ln\frac{\left(\tilde{Y}_t+KC\right)\left(\tilde{Y}_i+KC\right)}{\left(\tilde{Y}_s+KC\right)\left(\tilde{Y}_j+KC\right)}}{\tilde{Y}_t - \tilde{Y}_i}
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{\begin{array}{c}\frac{1}{\bar{u}}\left[\tilde{Y}_t + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_t\mu_2^*})\right] - \frac{1}{\bar{u}}\left[\tilde{Y}_s + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_s\mu_2^*})\right] \\ -\frac{1}{\bar{u}}\left[\tilde{Y}_j + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_j\mu_2^*})\right] - \frac{1}{\bar{u}}\left[\tilde{Y}_i + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-\tilde{Y}_i\mu_2^*})\right]\end{array}}{\tilde{Y}_t - \tilde{Y}_i} \\
= & \frac{\frac{1}{\bar{u}}\left[\tilde{Y}_t - \tilde{Y}_s + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(e^{-\tilde{Y}_s\mu_2^*} - e^{-\tilde{Y}_t\mu_2^*})\right] - \frac{1}{\bar{u}}\left[\tilde{Y}_j - \tilde{Y}_i + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(e^{-\tilde{Y}_i\mu_2^*} - e^{-\tilde{Y}_j\mu_2^*})\right]}{\tilde{Y}_t - \tilde{Y}_i} \\
= & \frac{\frac{1}{\bar{u}}\left[\tilde{Y}_t + \tilde{Y}_i - \tilde{Y}_s - \tilde{Y}_j + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(e^{-\tilde{Y}_s\mu_2^*} + e^{-\tilde{Y}_j\mu_2^*} - e^{-\tilde{Y}_t\mu_2^*} - e^{-\tilde{Y}_i\mu_2^*})\right]}{\tilde{Y}_t - \tilde{Y}_i}.
\end{aligned}
$$

The function T increases more slowly than the sequence of E(T) for

$$
\frac{1}{r}\ln\frac{\left(\tilde{Y}_t+KC\right)\left(\tilde{Y}_i+KC\right)}{\left(\tilde{Y}_s+KC\right)\left(\tilde{Y}_j+KC\right)} < \frac{1}{\bar{u}}\left[\tilde{Y}_t + \tilde{Y}_i - \tilde{Y}_s - \tilde{Y}_j + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(e^{-\tilde{Y}_s\mu_2^*} + e^{-\tilde{Y}_j\mu_2^*} - e^{-\tilde{Y}_t\mu_2^*} - e^{-\tilde{Y}_i\mu_2^*})\right].
$$

■

Appendix 2: Derivatives of the Expected First Realization of Market Entry Profit Level:

Proof of Proposition 4.

$$\frac{\partial E(T^*)}{\partial \lambda_1} = -\underbrace{\frac{\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}{\left[\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)\right]^2}}_{(1)}\underbrace{\left[Y^* + \frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}(1-e^{-Y^*\mu_2^*})\right]}_{(2)}$$

$$+\underbrace{\frac{1}{\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}}_{(3)}\underbrace{\left[\begin{array}{c} -\frac{\frac{\partial\beta}{\partial\lambda_1}}{(\beta-1)^2}(r - \int\limits_{f^{-1}(U_2)} z\upsilon(dz) \\ -\int\limits_{U_2}\left[\ln(1+z)-z\right]\upsilon_2(dz))I(T)\end{array}\right]}_{(4)}\underbrace{(1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*})}_{(5)}$$

For the first term (1) we obtain

$$\frac{\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}{\left[\delta+\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)\right]^2} > 0 \Leftrightarrow \frac{p}{\eta_1}-\frac{q}{\eta_2} > 0 \quad \text{with } q = 1-p.$$

With the same condition, we obtain a positive sign also for term (3)

$$\frac{1}{\delta+\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)} > 0.$$

For the second term (2) it holds that

$$\underbrace{Y^*}_{>0} + \underbrace{\frac{\mu_2^*-\eta_1}{\eta_1\mu_2^*}}_{>0}\underbrace{(1-e^{-Y^*\mu_2^*})}_{\geq 0} > 0.$$

The sign of the fourth term (4) depends on whether $\frac{\partial\beta}{\partial\lambda_1}$ is positive or negative. Assuming $\frac{\partial\beta}{\partial\lambda_1} < 0$, then term (4) becomes

$$-\frac{\frac{\partial\beta}{\partial\lambda_1}}{(\beta-1)^2}\underbrace{(r - \int\limits_{f^{-1}(U_2)} z\upsilon(dz) - \int\limits_{U_2}\left[\ln(1+z)-z\right]\upsilon_2(dz))}_{>0}\underbrace{I(T)}_{>0} > 0$$

The last term (5)

$$1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}$$

is negative if

$$1<\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}.$$

Summarizing all conditions leads to $\frac{\partial E(T^*)}{\partial \lambda_1}<0$. ■

Proof of Proposition 5.

$$\frac{\partial E(T^*)}{\partial u}=\underbrace{\frac{1}{\lambda\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}}_{(1)}\underbrace{\left[\begin{array}{c}-\frac{\frac{\partial\beta}{\partial u}}{(\beta-1)^2}(r-\int\limits_{f^{-1}(U_2)}zv(dz)\\ -\int\limits_{U_2}\left[\ln(1+z)-z\right]\upsilon_2(dz))I(T)\end{array}\right]}_{(2)}\underbrace{(1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*})}_{(3)}.$$

As before, the term (1) is positive. Accordingly, the last component (3) is negative for $1<\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}$. The sign of (2) depends on whether $\frac{\partial\beta}{\partial u}\gtrless 0$. Assuming that $\frac{\partial\beta}{\partial u}<0$, it follows that $\frac{\partial E(\tilde{T})}{\partial u}<0$. ■

Proof of Proposition 6.

$$\frac{\partial E(T^*)}{\partial \lambda_2}=\underbrace{\frac{1}{\lambda_1\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}}_{(1)}\underbrace{\left[\underbrace{\frac{\beta}{\beta-1}}_{>0}\left(\begin{array}{c}-\int\limits_{f^{-1}(U_2)}zh(dz)\\ -\int\limits_{U_2}\left[\ln(1+z)-z\right]h(dz)\end{array}\right)\underbrace{I(T)}_{>0}\right]}_{(2)}\underbrace{\left(1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*}\right)}_{(3)}.$$

From the conditions above (1) is positive and (3) is negative. Hence, the sign of $\frac{\partial E(T^*)}{\partial \lambda_2}$ depends on the second term and especially on the sign of $-\int\limits_{f^{-1}(U_2)}zh(dz)-\int\limits_{U_2}\left[\ln(1+z)-z\right]h(dz)$. Assuming more negative than positive jumps lead to a positive sign. ■

Proof of Proposition 7.

$$\frac{\partial E(T^*)}{\partial z}=\underbrace{\frac{1}{\lambda\left(\frac{p}{\eta_1}-\frac{q}{\eta_2}\right)}}_{(>0)}\underbrace{\left[\begin{array}{c}-\underbrace{\frac{\beta}{\beta-1}}_{>0}(\int\limits_{f^{-1}(U_2)}1\upsilon_2(dz)\\ +\int\limits_{U_2}\left[\frac{1}{(1+z)}-1\right]\upsilon_2(dz))\underbrace{I(T)}_{>0}\end{array}\right]}_{(2)}\underbrace{(1-\frac{\mu_2^*-\eta_1}{\eta_1}e^{-Y^*\mu_2^*})}_{(<0)}$$

According to the above assumptions (1) is positive and (3) is negative. The second term again depends on the jump part. However, even if there are more negative than positive jumps the effect on the jump part is not so large as to result in a negative sign. Therefore the sign of the overall derivative is negative. ■

Acknowledgements

The core of the real options model developed in this paper is based on the earlier working paper by Bilkic et al. (2013). We thank Natasa Bilkic for her contributions to this earlier working paper and the ideas contained in the present paper. The usual disclaimer applies.

References

Ahuja, G. and Lampert, C. (2001). Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions. *Strategic Management Journal*, 22:521–543.

Allen, P. and Greaves, M. (2011). The Singularity isn't Near. *MIT Technology Review*, October 12.

Armstrong, S., Bostrom, N., and Schulman, C. (2016). Racing to the Precipice: A Model of Artificial Intelligence Development. *AI & Society*, 31:201–206.

Azoulay, P., Jones, B., Kim, J., and Miranda, J. (2018). Age and High-Growth Entrepreneurship. *NBER Working Paper Series no. 24489. National Bureau for Economic Research.*

Bilkic, N. (2014). Essays on Investment Decisions under Uncertainty. *Ph.D Thesis, Department of Economics, Paderborn University, Germany.*

Bilkic, N., Gries, T., and Naudé, W. (2013). The Radical Innovation Investment Decision Refined. *IZA DP No. 7338, Bonn: IZA Institute of Labor Economics.*

Bilkic, N., Gries, T., and Pilichowski, M. (2009). Stay in School or Start Working? The Human Capital Investment Decision Under Uncertainty and Irreversibility. *CESifo Working Paper, No. 2825. Center for Economic Studies and ifo Institute (CESifo), Munich.*

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. *Oxford: Oxford University Press.*

Bostrom, N. (2017). Strategic Implications of Openness in AI Development. *Global Policy*, pages 1–14.

Cai, N. and Kou, S. (2011). Option Pricing Under a Mixed-Exponential Jump Diffusion Model. *Management Science*, 57(11):2067–2081.

Chandy, R. and Tellis, G. (1998). Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize. *Journal of Marketing Research*, 35:474–487.

Cox, H., Fairchild, J., and Pedersen, H. (2004). Valuation of Structured Risk Management Products. *Insurance, Mathematics and Economics*, 34:259–272.

Cox, S. and Pedersen, H. (2000). Catastrophe Risk Bonds. *North American Actuarial Journal*, 4(4):56–82.

Dixit, A. (1989). Entry and Exit Decisions under Uncertainty. *Journal of Political Economy*, 97(3):620–638.

Dixit, A. and Pindyck, R. (1994). Investment under Uncertainty. *Princeton NJ: Princeton University Press.*

Doraszelski, U. (2001). The Net Present Value Method versus the Option Value of Waiting: A Note on Farzin, Huisman and Kort (1998). *Journal of Economic Dynamics & Control*, 25:1109–1115.

Durkin, J. (1996). Expert Systems: A View of the Field. *IEEE Intelligent Systems*, 11:56–63.

Farzin, Y., Huismann, K., and Kort, P. (1998). Optimal Timing of Technology Adaption. *Journal of Dynamics & Control*, 22:779–799.

Floridi, L. (2020). AI and Its New Winter: From Myths to Realities. *Philosophy & Technology*, 33:1–3.

Freeman, C. (1994). The Economics of Technical Change. *Cambridge Journal of Economics*, 18.:463–514.

Friedman, T. (2016). Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations. *New York: Picador.*

Golder, P., Shacham, R., and Mitra, D. (2009). When, By Whom, and How Are Radical Innovations Developed? *Marketing Science*, 28(1):166–179.

Good, I. (1966). Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers*, 6:31–88.

Hao, K. (2019). We Analyzed 16,625 Papers to Figure Out Where AI is Headed Next. *MIT Technology Review, 25 January.*

Hendler, J. (2008). Avoiding Another AI Winter. *IEEE Intelligent Systems*, 2:2–4.

Hill, C. and Rothaermel, F. (2003). The Performance of Incumbent Firms in the Face of Radical Technological Innovation. *Academy of Management Review*, 28(2):257–274.

Hinton, G. and Salakhutdinov, R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313:504–507.

Hinton, G., S.Osindero, and The, T.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18:1527–1554.

Jang, J. (2007). Jump Diffusion Processes and their Applications in Insurance and Finance. *Insurance, Mathematics and Economics*, 41:62–70.

Jorgensen, S., Kort, P., and Dockner, E. (2006). Venture Capital Financed Investments in Intellectual Capital. *Journal of Dynamics & Control*, 30:2339–2361.

Joshi, N. (2019). How Far Are We From Achieving Artificial General Intelligence? *Forbes*, 10 June.

Koch, C. (2012). Modular Biological Complexity. *Science*, 337(6094):531–532.

Kou, S. (2002). A Jump-Diffusion Model for Option Pricing. *Management Science*, 48(8):1086–1101.

Kou, S. and Wang, H. (2003). First Passage Time of a Jump Diffusion Process. *Advances in Applied Probability*, 35(2):504–531.

Kurzweil, R. (2005). The Singularity is Near: When Humans Transcend Biology. *New York: Viking Press.*

Liu, F., Shi, Y., and Li, P. (2017). Analysis of the Relation Between Articial Intelligence and the Internet from the Perspective of Brain Science. *Procedia Computer Science*, pages 377–383.

Marvel, M. and Lumpkin, G. (2007). Technology Entrepreneurs' Human Capital and Its Effects on Innovation Radicalness. *Entrepreneurship Theory and Practice*, November:807–828.

Mazzucato, M. (2015). The Entrepreneurial State: Debunking Public vs. Private Sector Myths. *London: Penguin Books.*

Mazzucato, M. (2018). Mission-Oriented Innovation Policies: Challenges and Opportunities. *Industrial and Corporate Change*, 27(5):803–815.

Meese, T. (2018). The How and Why of Consciousness? *Frontiers in Psychology*, Volume 9, article 2173.

Merton, R. (1976). Option Pricing when the Stock Returns are Discontinuous. *Journal of Financial Economics*, 3:125–144.

Naudé, W. (2021). Artificial Intelligence: Neither Utopian nor Apocalyptic Impacts Soon. *Economics of Innovation and New Technology*, 30(1):1–24.

Naudé, W. and Dimitri, N. (2020). The Race for an Artificial General Intelligence: Implications for Public Policy. *AI & Society*, 35(2):367–379.

Naudé, W. and Dimitri, N. (2021). Public Procurement and Innovation for Human-Centered Artificial Intelligence. *IZA Discussion Paper*.

Newquist, H. (1994). The Brain Makers: Genius, Ego, and Greed in the Quest For Machines That Think. *New York: Macmillan*.

O'Connor, G. and McDermott, C. (2004). The Human Side of Radical Innovation. *Journal of Engineering and Technology Management*, 21:11–30.

Pham, H. (1997). Optimal Stopping, Free Boundary, and American Option in a Jump-Diffusion Model. *Applied Mathematics and Optimization*, 35:145–164.

Russel, S., Dewey, D., and M.Tegmark (2015). Research Priorities for Robust and Beneficial Artificial Intelligence. *AI Magazine, Association for the Advancement of Artificial Intelligence, Winter, pp. 105-114*.

Sarkar, S. (2000). On the Investment-Uncertainty Relationship in a Real Options Model. *Journal of Economic Dynamics and Control*, 24:219–225.

Whalley, A. (2011). Optimal R&D Investment for a Risk-Averse Entrepreneur. *Journal of Economic Dynamics & Control*, 35:413–429.

Yang, H. and Zhang, L. (2005). Optimal Investment for Insurers with Jump-Diffusion Risk Process. *Insurance: Mathematics and Economics*, 37:615–634.