HEALING HEALTHCARE: THE ROLE OF LINGUISTIC OPENNESS IN MEDIATING THE RELATIONSHIP BETWEEN STRATEGIC ALLIANCE AND INNOVATION PERFORMANCE

INTRODUCTION

“Good ideas may not want to be free, but they do want to connect, fuse, recombine. They want to reinvent themselves by crossing conceptual borders. They want to complete each other as much as they want to compete.” — Steven Johnson

Pharmaceutical companies have pushed the door open in space. Collaborating with NASA, they have grown crystals at the International Space Station and tried to create the next game-changing opportunity from orbit to the drug discovery and development for earthlings. Can you imagine a partnership between a drug manufacturer and a space agency? It is one small example, but also a giant leap about how innovating companies handle product research and development (R&D) with partnerships. Ever since Chesbrough brought it up in the early 2000s, open innovation has developed to be a rich concept. Scholars have studied it theoretically and empirically. Specifically, the early scholars emphasized its advantageous impacts, and recent research shows growing interest in finding the boundary of open innovation.

Major scholar attention about open innovation has centered on the “what” and “how” questions on a firm level. However, it is yet developed about the “when” inquiry. The “when” here explores how quickly does open innovation positively impact on a particular stage of innovation. Several scholars have emphasized the significance of filling up this gap. Exploring the time puzzle of open innovation can provide insights into the efficiency of the open innovation process (Brockman, Khurana, and Zhong 2018; West and Bogers 2014). It supports coordination of openness at a firm level (West et al. 2006) and therefore could potentially
maximize R&D efforts. Last but not least, it enhances the opportunity of “[winning] by making the best use of internal and external knowledge in a timely way” (Chesbrough 2003, p.52).

The purpose of this study is to answer the following question: when does open innovation improve innovation performance in large companies? To address this question, we focus on the pharmaceutical industry, where R&D suffers from sky-high expenses, declining productivity, and challenging New Drug Approval (NDA) process. Using the data from 24 large drug companies during 2000-2018, a generalized structural equation model together with some natural language processing techniques are also applied. Furthermore, we draw upon a theory from firm behavior, attention based view (Ocasio 1997, 2011), which suggests a linguistic perspective of open innovation. The linguistic openness measures the attention of the top management team and is captured through topic modeling. With this leadership approach, we seek to extend academic understanding about the internalization of open innovation.

**THEORY AND HYPOTHESES**

**Open innovation**

Open innovation, by definition, is an organizing principle that uses "a wide range of external actors and sources to help [organizations] achieve and sustain innovation" (Laursen and Salter 2006, p.131). It is the "openness" that makes open innovation stand out from other innovative concepts and enables organizations creating a distinctive competitive advantage (Henkel, Schöberl, and Alexy 2014). This competitive advantage is gained by breaking the organizational boundary which was used to be a key strategic building block (Teece 1986). From the perspective of open innovation, penetrating boundaries allows external knowledge flow in and out, and companies benefit from this open process (Simard and West 2006). Hence, opening up turns out to be an organizational capability that creates and captures values through external
accesses (Chesbrough 2003; West 2006). Then, how exactly do companies facilitate open innovation?

This question resonates with a classic inquiry in the field of innovation, how do companies innovate to capture value from it (Teece 2006). Innovations are used to happen when mobilized knowledge changes the existing knowledge landscape (Chesbrough 2003). Indeed, established companies have tried to achieve it internally and within its boundaries (O’Connor 2006). However, “[the] complexity of new technologies often goes beyond the capabilities of individual companies and forces innovating companies to cooperate with other firms and organizations to reduce the inherent uncertainties associated with novel product markets” (Vanhaverbeke and Cloodt 2006, p.258). And that is when the “openness” takes place, especially in large companies that aim at maintaining the vibes of innovation (Laursen and Salter 2006a) and then achieving breakthrough innovations (O’Connor 2006). Similar to the innovation framework, open innovation also builds on the goal of value capture (West et al. 2014). Furthermore, it is believed that by taking external ideas inside, additional value can be created (Chesbrough 2006b).

Scholars have explained several ways of facilitating open innovation. Based on knowledge management, for example, useful information and ideas “can come from inside or outside the company and can go to market from inside or outside the company as well” (Chesbrough 2006b, p.1). That is to say, open innovation mobilizes knowledge in both inbound (sourcing and acquiring) and outbound directions (revealing and selling) (Dahlander and Gann 2010). In particular, obtaining innovations from external sources has attracted the most scholarly attention, and the range of inbound extends to searching, enabling, incentivizing, and contracting (West and Bogers 2014). In this study, the “openness” is limited to the inbound condition.
Moreover, from a partnership perspective, innovating firms have used venture capitals to “catalyze their own innovation process” (Chesbrough 2003, p.55). Merger and acquisition is also a useful path to obtain new technologies or win market entrée (O’Connor 2006). Together with strategic alliances, these are the forms of open innovation in terms of knowledge insourcing and inter-organizational collaboration (Laursen and Salter 2014).

**Alliance.** This study focuses on the alliance form of open innovation. First, open innovation, by construction, is an external search strategy utilizing a wide range of inter-organizational collaborations (Laursen and Salter 2006b; Vanhaverbeke 2006). The strategic alliance is one form of formal ties. Teaming up with other organizations under contractual agreements and arranged channels has several benefits than informal ties. For example, an open innovation initiative can easily take strategic alliances; those formal partnerships bring knowledge spillover advantages and informal information flows, which in turn makes the process more open (Simard and West 2006).

Second, from resources based view, strategic alliances are built based on matching resources and capabilities in a relatively long period (Shaikh and Levina 2019). The partnership duration is long enough to devote major attention from adjusting changes into fulfilling shared goals and maximize complementary resources. Additionally, established formal ties attract new partnerships as well (Shaikh and Levina 2019).

Lastly, strategic alliances generally involve a wide range of organizations. This level of openness can enhance the innovative capacity of the focal company (Laursen and Salter 2006a). Research has also shown that multi-domain partnerships benefit the creation of innovations (Maula, Keil, and Salmenkaita 2006). However, except for the benefits, there are also costs and dangers associated with form ties, which is a research gap that several scholars have emphasized.
(i.e., Felin and Zenger 2014; Soh and Subramanian 2014; Teece 2006; West, Vanhaverbeke, and Chesbrough 2006). Hence, this research plans to explore the double-edged sword of strategic alliance in open innovation.

**The “when” question.** Is openness the panacea for innovation? Indeed, there are many benefits of open innovation. However, a paradox of openness exists. Openness brings valuable ideas to a general new combination of knowledge for innovations, but the following market development of them needs a shield from followers and competitors (Laursen and Salter 2014). Recently, more scholars start paying attention to the constraints of the open innovation process. For instance, one study investigated patent applications and found that those using knowledge of diverse domains turn out to be more likely to fail (Ferguson and Carnabuci 2017). Other possible consequences contain appropriation and opportunism (Brockman et al. 2018).

For the research of open innovation, the “what (the content of open innovation)” and “how (the process)” questions have been largely and broadly investigated by scholars (Huizingh 2011, p.2). Especially early studies are less good at answering the “when” and “how often” questions (West et al. 2014).

When does open innovation increase or impair the productivity of R&D? This “when” question has been one of the underdeveloped topics in a rich-development domain of open innovation (Huizingh 2011). One relevant research studied the efficiency of open innovation through the lens of societal trust and found out some hazards of external knowledge (Brockman et al. 2018).

However, it is an important inquiry and requires more scholarly attention. Because understanding the “when” relates to the detailed process after openness comes into a firm (West and Bogers 2014). Understanding the "when" helps facilitate the coordination of open innovation
by emphasizing the time variable, and then supporting to seek the answer for the following questions: "[if] projects move faster through the R&D system, does this result in more incremental innovation output? Or does a higher metabolic rate result in the faster incorporation of new knowledge, and in more (re)combinations of technologies in a given period" (West et al. 2006, p.291). Understanding the “when” provides the potentials of accelerating R&D cycle of radical innovation (O’Connor 2006), and “[winning] by making the best use of internal and external knowledge in a timely way” (Chesbrough 2003, p.52).

Previous research has shown that the relationship between partnership and R&D productivity is dynamic due to the moderation effects including the nature of knowledge as well as the age of firms (Soh and Subramanian 2014). Another similar research found different nature of partnership results in the different financial performance of R&D projects. (Du, Leten, and Vanhaverbeke 2014). In this study, we seek to understand the relationship between open innovation (measured by the number of the strategic alliance) and R&D productivity (measured by the number of products in discovery, incubation, and acceleration respectively). Accordingly, we assume that open innovation will impact the stages of R&D productivity differently. Moreover, we are interested in testing the following question, to what extent to which the stage of R&D productivity benefits from strategic alliances.

**Innovation performance and open innovation**

Innovation performance is captured by R&D productivity. Open innovation and R&D productivity are intertwined in terms of appropriating values. Open innovation can extend internal R&D knowledge with external connections to create and capture values (Chesbrough 2003). R&D, on the one hand, serves as an important asset to profit from investing key resources in the long-term research projects (Chesbrough 2003; West et al. 2014). On the other hand, R&D
builds entry barriers for competition with other companies (Chesbrough 2003). However, without devoting R&D investments into new products, companies hardly appropriate the returns or win the competition (Teece 1986). To develop new products, both research and development are necessary steps. In detail, research is “the exploration of new frontiers, punctuated by occasional flashes of insight that lead to exciting new discoveries,” while development “takes the output of research as an input into its own process… [Development] is fundamentally about making and hitting schedule targets and budgets, to convert discoveries into new products and services” (Chesbrough 2003, p.31-32). The “R” and “D” are continuous but different processes. It is natural to ask, then, does open innovation impact both processes in the same manner?

With the power of open innovation, large companies seek breakthrough innovations via R&D activities. Radical innovation, another name for non-incremental innovation, is perceived as a capacity with three competencies (O’Connor 2006). In detail, discovery is described as a conceptualization, with the “creation, recognition, elaboration, articulation of opportunities,” experimentation (or incubation) takes the opportunity to a step further as a business proposition, and commercialization (or acceleration) aims at “ramping up the business to stand on its own” (O’Connor 2006, p.69) In the pharmaceutical industry, for instance, new drug approval (NDA) is regarded as breakthrough innovation (Cohen and Caner 2016). Interestingly, the R&D procedure for NDA is comprised of at least three stages: drug discovery, clinical trial, and FDA review and market development, which corresponds well to the competencies that a radical innovation generally requires.

In fact, radical innovation sustains and renews companies' innovative position in the competition. Increasing radical innovations with a shortened life cycle and shrunk financial investment would be the ideal goal. It is also the goal of open innovation to accelerate the
metabolism of radical innovation by insourcing inbound and outbound flows of knowledge (O’Connor 2006). Hence, taking pharmaceuticals as an example again, when does open innovation benefit drug discovery and development, and does it impact on each stage of the R&D for NDA benefit the same way become important questions for open innovation.

To begin with, we should not assume that open innovation only brings positive effects to the R&D productivity. To put it differently, “not all innovation problems are suited equally well to this type of process” (Terwiesch and Xu 2008, p.1542). Unsuccessful applications of open innovation have not been reported much practically (West and Bogers 2014) or theorized much by scholars (Dahlander and Gann 2010). Or, a turning point exists where openess results in unproductivity of R&D (Laursen and Salter 2006a). If it is possible, where is this turning point?

**Discovery.** The discovery phase of R&D for radical innovation creates and refines a potential opportunity among many other candidates. As the first stage of an R&D process, it typically takes place inside an innovating firm (Terwiesch and Xu 2008). Open innovation, however, makes the discovery struggle between Intellectual Property (IP) protection and opening up internal knowledge (O’Connor 2006). Several scholars have shared this concern. For example, companies as both buyers and sellers of IP (Chesbrough 2003) need to facilitate open innovation wisely. Moreover, IP possibly weakens the degree of collaborating innovation (West 2006). And the balance between exposing resources to outsiders and protecting IP makes it challenging to milk from the discovery productivity (Dahlander and Gann 2010).

The fogged boundary of open innovation poses a threat to the discovery stage. Considering the greatly uncertain nature of discovery output (O’Connor 2006), it is challenging to handle too much knowledge as well as too many innovating opportunities (West and Bogers 2014). The inflow of external knowledge also stimulates an attitude of rejection and distrust (Chesbrough
as well as other “strong inertial forces” (Henkel et al. 2014, p.888). There is also the risk when contractual partners do not perform as the focal firm expects (Teece 1986). These arguments lead us to predict:

**H1 (a).** *Open innovation will negatively affect the outcome of the discovery stage of R&D productivity (Path C1 in Figure 1).*

**Experimentation.** The following stage of R&D evolves an innovative opportunity to market potentials. At this stage, the focus is to mature the outcome of discoveries by experiments after experiments. Interaction skills turn out to be the most important intangible resource (O’Connor 2006). Hence, collaborations and partnerships actually accelerate the process of experimentation (Maula et al. 2006). It is promising that open innovation benefits the incubation of innovative opportunities. Hence, we hypothesize:

**H1 (b).** *Open innovation will positively affect the outcome of the experimentation stage of R&D productivity (Path C2 in Figure 1).*

**Commercialization.** Radical innovation is defined as "the ability for an organization to commercialize products and technologies that have (a) high impact on the market in terms of offering wholly new benefits, and (b) high impact on the firm in terms of their ability to spawn whole new lines of businesses" (O’Connor 2006, p.64). By construction, the outcome of commercialization, radical innovations, contains more predictability than the discovery stage. At this last stage of R&D, it does more exploitation than exploration (O’Connor 2006). Indeed, previous research shows that building more partnerships contributes to the positive relationship between exploitation and radical innovations (Cohen and Caner 2016). Sometimes, serendipity explains the emergence of breakthrough innovation (Austin, Devin, and Sullivan 2012) and that also requires diverse domains of knowledge to act the "accidents." Moreover, "the type of
external knowledge sourced determines the likelihood of the creation of breakthrough innovation” (Phene, Fladmoe-Lindquist, and Marsh 2006, p.369). We thus predict:

\[ H1 \ (c). \text{Open innovation will positively affect the outcome of the commercialization stage of R&D productivity (Path C3 in Figure 1).} \]

**Large firms.** This study focuses on large companies. First, to create a sustainable competitive advantage, established companies can invest key resources in developing new products and processes for the long term (Chesbrough 2003). Second, large companies can leverage a rich pool of venture capital to absorb new knowledge (Chesbrough 2003). Third, open innovation at an organizational level keeps a firm contact with technological changes, especially in established companies (Christensen 2006). Fourth, compared to SMEs, large companies have developed a large pool of knowledge and that enables them commercializing discoveries from the R&D and scale innovation up (O’Connor 2006). Fifth, start-ups may devote most efforts in the discovery phase of the R&D process, large companies instead "focus on a broad range of activities in the value chain, seek downstream partnerships" (Soh and Subramanian 2014, p.811), and appropriating values from the entire R&D process. Lastly, equipped with key resources, established companies are good at exploiting the most benefits from their R&D outcomes (Teece 1986). Therefore, to examine the impact of open innovation on R&D productivity, large companies serve as the proper subjects for this study.

**Firm-centric and more.** Firm-level is a proper perspective for this study. One the one hand, open innovation, by definition, is a paradigm that describes a firm’s external search (Chesbrough 2006b); values are captured only through the open innovation business model of a particular company(West et al. 2006). On the other hand, as a coordination system (Srikanth and Puranam 2014), the firm is the subject to deliver the outcome of innovations(West et al. 2006).
Moreover, the majority of open innovation research is firm-centric (Randhawa, Wilden, and Hohberger 2016; West and Bogers 2014; West et al. 2006).

Though the major focus of open innovation research by far is firm-level (West et al. 2014), other levels of perspectives on open innovation are also recommended (Chesbrough 2003; West et al. 2014). Some recent studies have pursued various angels and generated interesting results. For example, one study explores the relationship between employee characteristics and openness of focal firms (Bogers, Foss, and Lyngsie 2018); one compares CEOs’ characteristics with the adoption of open innovation (Ahn, Minshall, and Mortara 2017); another individual-level of open innovation looks at the leadership style and its impact on the open innovation process (Edelbroek, Peters, and Blomme 2019). However, the "people side" of open innovation is underdeveloped and more research is required (West et al. 2014). Therefore, this study hopes to take the “human” level of open innovation into consideration and bridge the firm-level openness with linguistic open innovation.

**Linguistic openness: the language of open innovation**

Open innovation is a concept with a wide spectrum. Scholars need to explore other facets of it (Randhawa et al. 2016). For example, applying new measures or combining the measurement as two or more levels enrich the understandings of open innovation (West et al. 2014). Therefore, this study applies both the firm-level and the linguistic measurements of open innovation. Since open innovation has impacted other strategic and managerial fields (West et al. 2014), we hope to bring the “openness” to the language studies as well.

**Attention based view.** Attention based view is a theory in firm strategy and behavior. According to Ocasio (1997), explains the process of organizational operation under the attention
of their decision-makers. The central argument is that the selective focus of decision-makers
decides and facilitate strategic actions (Ocasio 1997).

Attention based view provides a novel approach to understanding open innovation and its
impact on R&D productivity. First, combining resource-based view, attention based view offers
the possibility of better explaining open innovation strategy and value creation. Once strategic
alliances are built and the external knowledge penetrates the boundary of a firm, it is the
organizational attention that guides the action of knowledge flows. It is also a tunnel that shapes
the degree of openness within a firm boundary and decides the absorbing capacity of openness.
Attention based view is closely related to the research question of this study, that is, when does
open innovation enhance or discourage R&D productivity. In the situations that valuable
knowledge flows in while the selective focus is absence, R&D productivity may hardly benefit
from the open innovation strategy. In other cases, when decision-makers' attention centers at
openness's impact on R&D productivity, the open innovation actions may deliver a satisfactory
outcome. Second, attention based view can contribute to the human side of open innovation—a
research gap that several scholars have emphasized. Existing research involves with traits of
employees (Bogers et al. 2018) and CEOs (Ahn et al. 2017), and looks at their impact on
innovation outcome. The attention based view, however, “opens the black box of the firm to
highlight the importance of situated attentional processes in selectively focusing decision-
makers’ attention” (Ocasio 1997, p.202). Hence, it has great potential in providing another
interesting angel about human-related research of open innovation.

The measure of attention in open innovation: linguistic openness. Here, organizational
or decision-makers' attention is still at the firm-level. But it is approached from a different theory
which in turn inspires a new measure for open innovation, linguistic openness. There is no better
measurement than vocabularies to represent "the socially structure pattern of attention" (Ocasio 1997, p.188). Linguistic openness is defined as a leadership approach to understanding open innovation and it is captured by the vocabularies used by the top management team to describe open innovation. Linguistic openness aims at capturing the degree to which openness is part of a chairman or CEO's theory of value creation for the firm.

First, vocabularies serve as the common ground for organizational coordination. In detail, “groups, divisions, and organizations need a shared vocabulary to communicate effectively to coordinate their behavior and accomplish complex organizing practices” (Loewenstein and Ocasio 2005, p.31). As a systematic innovation, open innovation requires coordination with various parts of the business system (Maula et al. 2006). Shared vocabularies, especially those that represent decision-makers' attention, can be an effective indicator of the flows of strategic action and resources. Hence, capturing the linguistic traces of decision makers' attention can be used as a measure of openness to predict R&D productivity.

Second, some studies have applied the theory of firm behavior and the measure of language in the research for open innovation. For instance, Laursen and Salter (2006a) used firm behavior to measure openness and studied its relationship with firm performance. Another research studied the paradox of open innovation by measuring product innovation rumors as a form of open innovation (Hannigan, Seidel, and Yalis-Douglas 2018). The successful application of linguistics in previous research also motivates this study. Therefore, organizational attention on open innovation captured by decision-makers' use of language serves as a combined firm-centric and linguistic-level measure of openness strategy.

Third, there is a growing interest in research using vocabulary measures in the field of strategy and management. However, this phenomenon is not new. There has been a rich history
about vocabularies use in the organizational studies, "both theoretically and empirically" (Loewenstein and Ocasio 2005, p.50). Then, a disconnection followed. The recent turn of scholarly attention indicates that vocabularies are not only the vehicles of information but also the building blocks of meaning; moreover, they are closely linked to strategic practice (Loewenstein and Ocasio 2005). Here, vocabularies are defined as the "systems of words, and the meaning of these words, used by collectives at a different level of analysis – groups, organizations, communities of practice, institutional fields—in communication, thought, and action" (Loewenstein and Ocasio 2005, p.46). It is exciting to combine the linguistic explanatory power in exploring a rich concept as open innovation. Thus, it is interesting to test if different types of open innovation offer sources to the top management team about where to distribute their attention. It is possible that the number of strategic alliances impacts on decision-makers' vocabularies about openness.

H2. Companies with more strategic alliances are more likely to show openness linguistically than those that are not (Path A in Figure 1).

Lastly, linguistic openness in the empirical part of this study is measured by openness topics in the Letter to Shareholders from companies’ annual reports. This decision stands on the shoulder of previous work about open innovation. In fact, as early as the year of 2003, openness-related statements from a firm's annual report were quoted as evidence of firm openness in Chesbrough's earliest book about open innovation. Even the word combination of “open” and “innovation,” have taught practitioners “a new language to speak about the nature of R&D, helping to shift the dominant logic of R&D away from the internal discovery toward external engagement” (West et al. 2014, p.805).
When applying open innovation strategy, the “forward-looking attentional perspective of top management facilitates an organization’s ability to overcome structural inertia and core rigidities” (Ocasio 2011, p.1292). Moreover, “[successful] strategic performance thereby requires the sustained focusing of attention and effort associated with controlled attentional processing” (Ocasio 1997, p.203). To capture a firm's attention in open innovation, we measure to what extent does linguistic openness corresponds to the changes in R&D productivity? Furthermore, previous research that studied the relationship between openness and innovative performance found that those extend the depth and breadth of external search tend to be more innovative (Laursen and Salter 2006a). Therefore, we theorize:

**H3. Companies that exhibit openness linguistically are more likely to have better R&D productivity at the experimentation stage than those do not (Path B in Figure 1).**

Figure 1 portrays the relationship between hypotheses.

---

Insert figure 1 here

---

**RESEARCH SETTING**

We choose large companies from the pharmaceutical industry to test our hypotheses. A variety of previous research have also chosen this industry as empirical background for open innovation research (i.e., Caner, Cohen, and Pil 2017; Cohen and Caner 2016; Eisenman and Paruchuri 2019; Gupta et al. 2009; Karim and Kaul 2015; Keijl et al. 2016; Roy and Chaguturu 2014). Pharmaceutical companies serve as the proper subjects to explore the relationship between open innovation and R&D productivity in this study for several considerations.

First, innovating and collaborating blood runs in the pharmaceutical industry. Innovation is the secret behind the growth of the pharmaceutical industry and it has been proven to improve
the outcome of R&D (Schuhmacher et al. 2013). Also, to reach out for a wider range of patient base, to achieve a better reputation, and to broaden R&D networks than before, companies from this industry have continued collaborating with universities and innovative companies (Deloitte 2018). A potential positive relationship between alliances and desirable R&D outcome drives them to seek partnerships across diverse groups of purpose. Interestingly, when the concept of open innovation was brought up, companies from this industry was illustrated as the evidence (Chesbrough 2003). Research about open innovation has mainly focused on large high-tech firms (West and Bogers 2014). Thus, the pharmaceutical industry fits the boundary condition of open innovation. Furthermore, the ongoing collaborations and partnerships reinforce the idea that opening up contributes to the improvement of R&D productivity. Except for the firm level, openness results in innovative solutions in healthcare on a team level as well. (West et al. 2014)

Second, innovation has always been open in drug discovery and development (Dahlander and Gann 2010). Back in the 1970s, ideas from outside of the firm have been recognized as the origin of innovations (West et al. 2014). Particularly, according to Cragg and Newman (2013), medical materials firstly were found in nature and developed from herbs and plants; then in the 1970s, microbial substances were added to drug making techniques; in the late 1980s, chemistry became a new source for medicines which shifted drug-making from nature search to lab experiments; recent frontier techniques based on genetic differences of patients open up a new way of drug innovations.

Third, there is a crucial need for open innovation research in the pharmaceutical industry. The reality of the R&D productivity goes far apart from the desired trend, which is to create cost-efficient NDAs with fruitful quantities and shortened life cycles (Paul et al. 2010). Specifically, the decreasing R&D efficiency (Scannell et al. 2012), sky-high costs in R&D investment which
has climbed up 32% during 2014-2019 (IQVIA 2019), expensive average expenses of the complete development of a new drug ($2.87 billion) (DiMasi, Grabowski, and Hansen 2016), and no corresponding increase of output of NDAs (Pammolli, Magazzini, and Riccaboni 2011), are desperate for more insights of how to improve R&D efficiency and effectiveness. Since external search has been a useful source for the process of drug discovery and development (Bignami and Mattsson 2019), open innovation research in this industry is potentially advantageous practically.

Fourth, under so many urgent challenges, decision-making is indeed the “core function of any drug development firm” (Jekunen 2014, p.2009). Then, the focus of decision-making turns out to be a necessary measure to check the level of openness from the top management team. Additionally, decision-making is by no means perceived as a stable process (Jekunen 2014). Therefore, applying the linguistic openness measure under the theory of attention based view can track down every pharmaceutical firm's annual attention on open innovation.

Sixthly, this industry naturally provides testing filed for the research design of this study. The nature of the pharmaceutical business is time-sensitive. The timing to fulfill financial and social obligations as well as to deliver promising outcomes through R&D (Jekunen 2014) makes it a proper empirical background to explore the ‘when” question in this study. Moreover, drug discovery and development by regulation and design are comprised of several stages with clear boundaries. The discovery-experimentation-commercialization setting fits in very well. Except for time, geographic location does not separate the US pharmaceutical companies from the Europeans in terms of R&D productivity (Pammolli et al. 2011). Therefore, this study includes the top players of this industry regardless of their original nations. After all, one similarity among them exist. That is, they are all global-based entities.
METHODS

Sample and data

This study involves a sample of 456 firm-year observations from 24 largest pharmaceutical companies over 2000-2018. The level of analysis is firm. We describe the dependent and independent variables below and explain how their relationship will be tested using structural equation modeling. Linguistic openness serves as a mediator variable, and its measurement will be described below as well.

Dependent variables

Due to the nature of innovation performance, there are three dependent variables in this study: discovery, experimentation, and commercialization of R&D productivity.

Discovery. The data of the discovery variable is collected from Capital IQ. It is calculated as the sum of the following items: (1) number of products in discovery research (2) number of products in research and development (3) number of products in pre-clinical trials. It is lagged one year.

Experimentation. The data of the experimentation variable is collected from Capital IQ. It is calculated as the sum of the following items: (1) number of products in clinical trials (2) number of products in phase I (3) number of products in phase II (4) number of products in phase III (5) number of products in pre-registration. It is lagged one year.

Commercialization. The data of the commercialization variable is collected from Capital IQ. It is calculated as the sum of the following items: (1) the number of products approved during the period (2) the number of products launched during the period. It is lagged one year.

Independent variable
Strategic alliance. The data of this variable is collected from Capital IQ. It measures the number of strategic alliances per year per firm. Though there are several databases for collecting alliance data, our data from Capital IQ is valid. According to one previous research in understanding different data sources of strategic alliances, its findings indicate that "even though the databases only capture a sample of alliance activity, they may yield reliable results for many—if not all—research purposes" (Schilling 2009, p.233). Except for this independent variable, every other variable are lagged one year.

Mediator

Linguistic openness. The data for the mediator is collected from Letter to Shareholders in the annual report per year per firm. The data is collected from the Mergent Online database. In total, 233 letters have been collected. This mediator is a measure of open innovation. It is calculated by the topic modeling through Python. It is binary with 1 indicating that one letter explicitly contains open innovation topics, while 0 represents few or none coverage of openness.

Controls

R&D intensity. It is calculated by using R&D expenditure divided by sales per year per firm * 100 (in percentage %). The data is collected from Bloomberg. Research has shown that a firm’s R&D intensity is an important measure for open innovation research (Laursen and Salter 2006a). The data is in percentages.

Number of employees. The data is collected from Bloomberg per year per firm (thousand). The change of staff has shown an impact on a company's innovation process (Chesbrough 2006a). Hence, it should be controlled for this study.

Sales. The data of this variable is also collected from Bloomberg. It indicates the revenue per year per firm and the data has been logged.
Firm age. Bloomberg also provides the data for this variable. It is defined as the number of years (plus one) elapsed since the year of the company’s foundation.

Innovation preference. This dummy variable describes a company’s introverted or extroverted preference in innovation management. Extroverted firms are assigned to 1 while introverted one is 0. The list of companies' innovation preferences is based on previous research by Schuhmacher et al (2013).

Level of externally acquired R&D projects. This dummy variable describes a company’s proportion of externally acquired R&D projects. Firms with a high level are assigned to 1 while the others are 0. The list of companies' proportion of externally acquired R&D projects is based on previous research by Schuhmacher et al (2013).

M&A. The outcome of R&D can be highly influenced after M&A. Mergers and acquisitions (M&A) variable is to control for it, with 1 indicating the involvement of M&A in that year and after, and 0 without any M&A in that year.

Linguistic controls

The source of the following additional controls is the Letter to Shareholders in the annual report per year per firm.

Type-token ratio. This variable describes the total number of types divided by the total number of tokens per year per firm. It measures to what extent do CEOs or Chairmen use the variety of words to write the letters. It is achieved by the Natural Language Processing Toolkit (NLPT) through Python. This data is normalized before analysis. The normalization process transforms data into a standard distribution with a mean of 0 and a variance of 1. Through normalization, this set of data is adjusted to a common scale.
**Average word length.** This variable describes the average number of characters per word in a letter (per year per firm). It measures to what extent do CEOs or Chairmen use the length of words to write the letters. It is achieved by NLPT through Python. This data is normalized before analysis.

**Total number of words.** This variable describes the total number of words CEOs or Chairmen in a letter (per year per firm). It is achieved by NLPT through Python. This data is normalized before analysis.

**Concreteness.** The level of concreteness is a language measure that captures the persuasiveness in communication. It has shown its effectiveness in predicting strategic outcomes (Pan et al. 2018). Using the LIWC software, this variable is calculated by using the sum of three normalized lexical category scores (including verbs, numbers, nonspecific quantifiers, and past-focused) minuses the sum of another three normalized lexical category scores (including adjectives, nonspecific quantifiers, and future-focused).

**Open.** It is a binary variable with 1 indicating that the word “open” is used in a letter, while 0 means the absence of this word. It is achieved through Python.

**Knowledge.** It is a binary variable with 1 indicating that the word “knowledge” is used in a letter, while 0 means the absence of this word. It is achieved through Python.

**Innovation.** It is a binary variable with 1 indicating that the word “innova*” is used in a letter, while 0 means the absence of this word. It is achieved through Python.

**Alliance.** It is a binary variable with 1 indicating that the word “alliance” is used in a letter, while 0 means the absence of this word. It is achieved through Python.

**Partner.** It is a binary variable with 1 indicating that the word “partner” is used in a letter, while 0 means the absence of this word. It is achieved through Python.
Collaboration. It is a binary variable with 1 indicating that the word “collabora*” is used in a letter, while 0 means the absence of this word. It is achieved through Python.

Analytical approach: generalized structural equation modeling to test for mediation

We use structural equation modeling (SEM) to examine the mediating effect of linguistic openness on the relations between strategic alliance and R&D productivity. This study contains three dependent variables, and several paths to test (Path A, B, C1, C2, and C3). The advantage of SEM overrunning several separate regressions is that "the simultaneous equations control for measurement errors that might lead to under- or overestimation of mediation effects" (Kaplan and Vakili 2015, p.1448). The nature of our dependent variables are counts, and the mediator variable is binary, which leads us to use the generalized SEM (GSEM), model. GSEM in particular deals with counts and binary data (Stata 2013). We use a negative binomial function for regressions with count outcomes and a probit function for regressions with binary outcomes.

---

Insert table 1 here

---

RESULTS

Table 1 represents descriptive statistics for the whole sample. Table 2 reports the results of the structural equation models. Specifically, column 1 identifies the total effect of the independent variable and mediator on the discovery variable, column 2 and 3 identify the total effect on the experimentation and commercialization variables.

---

Insert table 2 here

---

Discovery
We do not see a negative effect of strategic alliance on the discovery stage of innovation performance (Hypothesis 1a with mediation effect). However, neither does the positive relationship between these two from the structural equation model show statistically significance. We find that the number of employees and sales are both positively related to the productivity of discovery with a p-value of less than 0.01. Besides, average word length in this stage shows a negative association with the mediator, while the total number of words and texts containing the word "open" both indicate positive associations with linguistic openness; all three are statistically significant at p<0.10 level. It makes sense methodologically: the larger the total number of words that a letter has, the higher the possibility of it showing openness linguistically; the more often a word like “open” shows in a letter, the higher the linguistic openness will be.

**Experimentation**

Turning to Model 2, which tests the total effect on experimentation, we find a strong positive relationship between strategic alliance and the dependent variable. In other words, holding all other variables at their means, as the number of strategic alliances increases by 1, the number of products in the experimentation phase of innovation increases by 0.024, which is statistically significant at p<0.01 level (Hypothesis 1b with mediation effect). Interestingly, firm age shows statistically negative association; that is to say, the older a company gets, the less experimentation productivity it will achieve. Moreover, another two control variables show statistical significance at p<0.001 level and they identify that a company with more introverted innovation preference or more internal R&D projects will generate a better outcome in the experimentation stage. Linguistic openness, however, presents a non-significant negative association with experimentation; hence, we do not find support for hypothesis 3.

**Commercialization**
Both innovation preference and externally acquired R&D variables indicate statistically negative associate with the dependent variable in Model 3 (last column). Other than that, hypothesis 1c is not supported but it does show a positive relationship.

**DISCUSSION AND CONCLUSION**

This study assumes that opening-up wider (more strategic alliances) gives more bets for companies that they can benefit from external resources and enhance their innovation performance. The research question, when does open innovation improve innovation performance, is conceptually explored and partially answered econometrically through this study. There is little doubt that we should consider more about the model itself and data handling to provide a better fitting empirical analysis. Additionally, strategic alliances not showing significant results in the pharmaceutical dataset reveal another possibility; that is to say, the industrial side has not figured out well with open innovation. The internalization of open innovation raises a concern and hopefully can be captured in future studies.

Open innovation is mainly captured through the number of strategic alliances in this study; in other studies, venture capitals have been applied to represent open innovation. However, open innovation is more than crowdsourcing. It is an organizing principle; hence the adoption of open innovation should associate with culture changes, and linguistic openness is a new way to understand it as a mode that works. It is new because previous studies look into open innovation through the lens of capabilities. A leadership lens of open innovation, instead, specifically understands whether openness is viewed as an enabler of the theory of value creation made by the top management team. Particularly, CEOs and chairmen, are the holders of companies’ identity, value, and culture. And this study emphasizes the top leader's role when deciding the entire set of strategic alliances (Koka and Prescott 2002).
We have shown evidence that open innovation exhibits a positive impact on the experimentation stage of innovation. But more work has to be done. Methodologically, for example, texts including quarterly earnings conference call transcripts may provide richer linguistic sources than the Letter to Shareholders from annual reports. Topic modeling can also be improved by adding precision: by showing weights per topic in each firm-year text, we can get a continuous linguistic openness variable than a binary one. And a larger range of texts enhances the analytical power of topic modeling as well. As to robustness check, we can look into a specific therapeutic area, for instance, oncology, for a better demonstration about the degree of the contribution that opening-up to positive innovation performance.
REFERENCE


Soh, Pek Hooi and Annapoornima M. Subramanian. 2014. “When Do Firms Benefit from University-Industry R&D Collaborations? The Implications of Firm R&D Focus on


Figure 1. The role of linguistic openness in mediating the relationship between strategic alliance and R&D productivity
Table 1 Descriptive statistics

| Variable                              | Mean | Std. Dev. | Min  | Max   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| Discovery (DV)                        | 113.20 | 115.122 | 1.000 | 635.000 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | |
| Experimentation (DV)                  | 70.332 | 74.991 | 1.000 | 307.000 | -0.136 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | |
| Commercialization (DV)                | 7.253 | 9.823 | 1.000 | 80.000 | -0.047 | 0.479 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | |
| Linguistic Openness (M)               | 0.167 | 0.374 | 0.000 | 1.000 | -0.176 | -0.185 | -0.091 | 1.000 | | | | | | | | | | | | | | | | | | | | | | |
| Type-Token Ratio                      | 13.447 | 12.706 | 1.000 | 65.000 | 0.349 | 0.572 | 0.360 | -0.343 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| R&D Intensity                         | 10.033 | 11.306 | 5.148 | 136.605 | -0.561 | -0.059 | 0.035 | -0.043 | -0.170 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| Number of Employees                   | 46.228 | 41.054 | 0.090 | 135.696 | 0.114 | 0.140 | 0.345 | 0.376 | 0.673 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| Firm Age                              | 77.205 | 59.862 | 1.000 | 238.000 | 0.050 | 0.164 | -0.280 | 0.158 | 0.343 | 0.673 | 0.717 | 1.000 | | | | | | | | | | | | | | | | | | | | |
| Innovation Preference                 | 0.417 | 0.494 | 0.000 | 1.000 | 0.350 | 0.570 | -0.463 | -0.130 | -0.232 | 0.099 | 0.082 | -0.105 | 0.494 | 1.000 | | | | | | | | | | | | | | | | | | | |
| Externally Acquired R&D               | 13.060 | 2.501 | 0.000 | 1.000 | -0.446 | 0.097 | -0.423 | -0.287 | -0.181 | 0.256 | -0.551 | -0.712 | -0.947 | -0.179 | 1.000 | | | | | | | | | | | | | | | | | | | |
| M&A                                   | 0.762 | 0.427 | 0.000 | 1.000 | 0.165 | 0.073 | -0.280 | 0.046 | 0.282 | 0.012 | -0.063 | 0.019 | 0.008 | 0.357 | 0.162 | 1.000 | | | | | | | | | | | | | | | | | | | |
| Concentremess                         | 0.000 | 2.744 | 7.157 | 1613.048 | 0.000 | -0.349 | 0.167 | 0.324 | -0.483 | -0.406 | -0.417 | -0.080 | 0.441 | 0.096 | 1.000 | | | | | | | | | | | | | | | | | | | |
| Type-Token Ratio                      | 0.000 | 1.000 | 2.0.19 | 7.75 | -0.299 | -0.232 | -0.215 | -0.190 | 0.132 | -0.246 | -0.224 | -0.258 | -0.173 | 0.187 | -0.124 | 0.190 | 1.000 | | | | | | | | | | | | | | | | | | | |
| Average Word Length                   | 0.000 | 1.000 | -4.268 | 3.418 | -0.503 | -0.120 | -0.135 | -0.043 | -0.100 | 0.451 | -0.504 | -0.184 | -0.186 | -0.142 | 0.148 | 0.054 | 0.563 | 0.013 | 1.000 | | | | | | | | | | | | | |
| Total Number of Words                 | 0.000 | 1.000 | -1.767 | 3.420 | 0.464 | 0.122 | -0.004 | 0.262 | 0.109 | 0.139 | 0.265 | 0.239 | 0.265 | 0.119 | 0.139 | 0.076 | -0.820 | -0.180 | 1.000 | | | | | | | | | | | | | |
| Knowledge Test                        | 0.000 | 0.501 | 0.000 | 1.000 | 0.320 | 0.039 | -0.188 | -0.112 | -0.068 | -0.170 | 0.189 | 0.039 | 0.025 | -0.382 | 0.162 | -0.135 | -0.332 | 0.162 | 1.000 | | | | | | | | | | | | | |
| Open Test                             | 0.378 | 0.486 | 0.000 | 1.000 | 0.139 | 0.247 | -0.134 | -0.139 | 0.047 | 0.001 | -0.060 | -0.138 | 0.009 | 0.289 | 0.095 | 0.200 | -0.014 | 0.131 | 0.013 | 0.000 | 0.090 | 1.000 | | | | | | | | | | | | | |
| Innovation Test                       | 0.863 | 0.345 | 0.000 | 1.000 | 0.002 | 0.168 | 0.053 | 0.058 | 0.260 | 0.151 | 0.071 | -0.036 | -0.202 | -0.233 | 0.201 | -0.083 | 0.151 | -0.211 | -0.387 | 0.381 | 0.201 | 0.028 | 1.000 | | | | | | | | | | | | | |
| Alliance Test                         | 0.232 | 0.423 | 0.000 | 1.000 | 0.264 | 0.068 | 0.208 | -0.149 | 0.357 | 0.219 | 0.414 | 0.401 | 0.384 | 0.062 | 0.376 | -0.049 | -0.117 | -0.379 | -0.229 | 0.437 | 0.148 | 1.000 | | | | | | | | | | | | | |
| Partnership Test                      | 0.618 | 0.487 | 0.000 | 1.000 | 0.292 | 0.068 | -0.095 | 0.169 | 0.386 | -0.169 | 0.389 | 0.308 | 0.055 | 0.243 | -0.248 | 0.016 | 0.020 | -0.414 | -0.232 | 0.560 | 0.031 | -0.049 | 0.342 | 0.522 | 1.000 | | | | | | | | | | | | | |
| Collaboration Test                    | 0.455 | 0.499 | 0.000 | 1.000 | 0.176 | 0.150 | -0.253 | -0.139 | -0.130 | -0.168 | 0.152 | 0.231 | 0.298 | 0.289 | -0.193 | -0.067 | 0.011 | -0.425 | -0.109 | 0.541 | 0.090 | 0.200 | 0.249 | 0.279 | 0.598 | 1 |
Table 2 Tests of mediation (dv=innovation performance including discovery, experimentation, and commercialization)

<table>
<thead>
<tr>
<th></th>
<th>Discovery (A*B)+C1</th>
<th>Experimentation (A*B)+C2</th>
<th>Commercialization (A*B)+C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Openness</td>
<td>1.004</td>
<td>-0.147</td>
<td>-0.276</td>
</tr>
<tr>
<td>(M)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategic Alliance (IV)</td>
<td>0.003</td>
<td>0.024**</td>
<td>0.004</td>
</tr>
<tr>
<td>R&amp;D Intensity</td>
<td>-0.072</td>
<td>0.002</td>
<td>-0.007</td>
</tr>
<tr>
<td>Number of Employees</td>
<td>0.029**</td>
<td>0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td>Sales</td>
<td>2.592**</td>
<td>0.236</td>
<td>-0.045</td>
</tr>
<tr>
<td>Firm Age</td>
<td>-0.019</td>
<td>-0.004**</td>
<td>-0.002</td>
</tr>
<tr>
<td>Innovation Preference</td>
<td>0.858</td>
<td>-0.542**</td>
<td>-0.901**</td>
</tr>
<tr>
<td>Externally Acquired</td>
<td>-1.328</td>
<td>-0.945**</td>
<td>-1.795**</td>
</tr>
<tr>
<td>R&amp;D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M&amp;A</td>
<td>-0.352</td>
<td>0.374</td>
<td>0.073</td>
</tr>
<tr>
<td>Concreteness</td>
<td>0.056</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>Type-Token Ratio</td>
<td>0.244</td>
<td>0.244</td>
<td>0.244</td>
</tr>
<tr>
<td>Average Word Length</td>
<td>-0.225+</td>
<td>-0.225+</td>
<td>-0.225+</td>
</tr>
<tr>
<td>Total Number of Words</td>
<td>0.351+</td>
<td>0.351+</td>
<td>0.351+</td>
</tr>
<tr>
<td>Knowledge Text</td>
<td>0.178</td>
<td>0.178</td>
<td>0.178</td>
</tr>
<tr>
<td>Open Text</td>
<td>0.429+</td>
<td>0.429+</td>
<td>0.429+</td>
</tr>
<tr>
<td>Innovation Text</td>
<td>-0.151</td>
<td>-0.151</td>
<td>-0.151</td>
</tr>
<tr>
<td>Alliance Text</td>
<td>-0.043</td>
<td>-0.043</td>
<td>-0.043</td>
</tr>
<tr>
<td>Partner Text</td>
<td>0.143</td>
<td>0.143</td>
<td>0.143</td>
</tr>
<tr>
<td>Collaboration Text</td>
<td>-0.178</td>
<td>-0.178</td>
<td>-0.178</td>
</tr>
<tr>
<td>Constant</td>
<td>-5.734</td>
<td>3.297**</td>
<td>3.810**</td>
</tr>
</tbody>
</table>

**p<0.01; *p<0.05; +p<0.10