ANALYTICS IN LEARNING:

FROM CONSUMER LEARNING TO ORGANIZATIONAL LEARNING

JIANG ZHIYING

(*M.Sc. Econ.*), NUS

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MARKETING

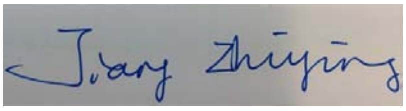
NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Jiang Zhiying

12 March 2014

ACKNOWLEDGEMENTS

At the point of completing this dissertation, I recollect my life since 06th August, 2006, the day I joined NUS Business School as a PhD candidate to pursue an academic career. I surveyed the days in retrospect and realized it was such a long and rich journey that made me grow. Thanks to the people I have encountered, for their enlightenment, encouragement, companion, and understanding.

First and for most, I’d like to thank my husband, who followed me to Singapore and gave up his own career opportunities in Europe back in 2006. Thanks for his great patience to support my eight years’ study yet faithfully believing in my potential to become a good scholar one day. At the moments of giving up, it was his firm faith in me that kept me going. Without Albert, I could not imagine walking this far.

Next, I’d like to thank my supervisor, Prof. Surendra Rajiv. It had been an honor to work with him. Rajiv had been recognized by his peers in the field as an extraordinary intelligent and profound scholar. However, one thing he constantly conveyed to me, either consciously or unconsciously, is a simple principle that genius are made of sweat. Such diligence is reflected as the mental effort one is willing to exert to explore the thorough nature of a phenomenon. It is also reflected as the mental simulations one goes through time and time again to connect the intricate web of knowledge in mind. I always remember the tease he had with me when I told him I forgot how the derivation should go. He smiled and said: “for you, it’s a problem of memorizing; for me, it’s a problem of understanding.” It was a bit painful, when I heard it for the first time, but now it has become a doctrine that will benefit my whole academic life. I would also like to thank him for the freedom he gave me to let me figure out what I really wanted without influencing me in his favor. He helped me to come up the topic of my dissertation. He had always been constructive whenever I needed his advice. He has been very hands on hands on with my first essay, but pushed me a lot to finish my second essay with more independence. When I recall the days with him, the only thing I regret was I should have initiated more discussions with him.

I would also like to thank my second supervisor Prof. Chu Junhong. Junhong entered NUS the same year as I entered the PhD program. I would like to thank her for her selfless help to any PhD students. As far as I remembered, she always stopped her work at hand whenever I dropped by her office to ask for help. She is a role model for the PhD students for her strong will, persistency, hard work, email response in light-speed, and never depleted strong self-control. After recent close work with her, I have also found her solid knowledge foundation, the rigidness of doing research yet the down to the earth humbleness. I always hug myself for being lucky to have her as my advisor, mentor and friend.

I would like to thank Prof. Trichy Krishnan. I could never forget in my first year summer, when he taught me hand by hand on basic analytics. Doing research with him made me understand that one should not give up an idea easily when facing hurdles. I want to thank Xiao Ping, for all her encouragement, sincerity and the long hours she spent to channel me back to the track. I want to thank Prof. Lim Weishi for all the delightful chat and discussions I had with her. Of course, I could not forget my dearest PhD fellow students, without whom my memory would become so plain. Thank you for all your companion and I look forward to meeting you in near future as a new force in the field. I'd like to thank the marketing department as a whole for all the supports and I'm proud to be a PhD candidate here.

Last but not least, I'd like to give my special thanks to both the internal and external examiners, who rendered constructive comments to make the dissertation a better work. Any errors that remain are my sole responsibility.

TABLE OF CONTENTS

SUMMARY

Environment changes constantly and it is learning that enables us to adapt to the external changes in a timely fashion. The topic of this dissertation is about learning. The first essay discusses consumer experiential learning with recall from two different memory systems. The second essay studies an organizational learning capability called absorptive capacity under the context of knowledge alliances.

In Essay I, we first ask ourselves an interesting question on what has been recalled in consumer's mind when forming an attitude toward a brand. Is it a previously formed overall impression or is it a vivid visualization of certain consumption episodes? A large literature in cognitive research has established the existence of both semantic and episodic memory in human brain, where semantic memory stores general knowledge and episodic memory stores personally experienced events that are context specific. In the traditional learning model, a consumer is assumed to make brand choice only based on the overall quality evaluation from semantic memory. Hence, in this paper we propose a structural model with Bayesian learning that allows recall from both semantic and episodic memory. We also attempt to empirically test the effect of idiosyncratic traits as well as situational factors triggering the type of memory recalled. We calibrate the proposed model on scanner panel data in the laundry detergent category. We find that consumers are more likely to recall past consumption experiences to form a new evaluation at the point of purchase, compared to recalling an existing belief from semantic memory.

Absorptive capacity is defined as a firm's capability to recognize the value of external knowledge, assimilate it and apply it to commercial ends. Absorptive capacity is a firm's fundamental learning capability that enables a firm to be adaptively innovative and structurally flexible to external changes. In Essay 2, we propose a 3-step structural model to

model this construct, which is widely applied but poorly measured in the literature. With our model, it is possible to use widely available alliance data to test empirically various theories about absorptive capacity. It sheds light on the determinants of each building block of absorptive capacity and gives implications to firms on how they can build and strengthen their absorptive capacity.

LIST OF TABLES

LIST OF FIGURES

Essay I

Semantic Versus Episodic Processing in Consumer Experiential Quality Learning

Abstract

When making a brand choice, a consumer needs to form an evaluation for each brand under consideration. An interesting question to ask is what has been recalled in her mind to form an attitude toward a brand. Is it a previously formed overall impression or is it a vivid visualization of certain consumption episodes? A large literature in cognitive research has established the existence of both semantic and episodic memory in human brain, where semantic memory stores general knowledge (such as brand evaluation) and episodic memory stores personally experienced events that are context specific (such as consumption experiences). In the traditional learning model, a consumer is assumed to make brand choice only based on the overall quality evaluation from semantic memory. Hence, in this paper we propose a structural model with Bayesian learning that allows recall from both semantic and episodic memory. We also attempt to empirically test the effect of idiosyncratic traits as well as situational factors (based on finding in both experimental and MRI-based studies) on triggering the type of memory being recalled. The consumer depicted in this paper is assumed to have imperfect memory, i.e., recall with forgetting errors. In fact, it is the explicit modelling of these forgetting errors that allows us to econometrically identify and distinguish between the two memory systems. We calibrate the proposed model on scanner panel data in the laundry detergent category, and find that consumers are more likely to recall past consumption experiences to form a new evaluation at the point of purchase, rather than recalling an existing belief from semantic memory.

KEYWORDS: Quality Learning, Memory-based Judgment, Dual-process Model, Semantic Memory, Episodic Memory, Structural Model

1. Introduction

Consider a consumer looking to buy a laundry detergent at a typical supermarket. Choosing a brand is definitely not that simple for her if she were a beginner and quite daunting even if she has a decent usage experience of the product category. The first issue to resolve is: liquid or powder? Then there is brand proliferation to deal with – Tide, Surf, Cheer, Bold, Fab, etc. The Tide brand (by Procter and Gamble) itself has several varieties: Tide, Tide Liquid, Tide Powder, Tide Simple Pleasures, Tide Coldwater, Tide with Bleach, Tide HE, 2X Ultra Tide Liquid and several more. The consumer can also get Tide in a variety of scents – clean breeze, mountain spring, tropical clean, meadows & rain, citrus & light, April fresh, glacier, etc. Other brands are also in multiple variants. How will she choose a particular brand? Rationality based arguments will suggest that she will look at her preference for the various brands and their prices and select the one that yields highest quality per unit price (Allenby and Rossi 1991, Chiang 1991, Chintagunta 1993). A moot question, then, is: Is a consumer "endowed" with (possibly evolving) brand preferences i.e., does our consumer arrive at the supermarket with a preference structure (with associated indifference curves) in her mind or is it "constructed" when confronted with the brand choice task?

A dominant view in behavioural decision research posits that preferences for objects of any complexity are constructed – not merely revealed – while generating a response to a judgment or choice task (Payne et al. 1992). This perspective suggests that while making brand choice, consumers construct preferences – brand evaluations/ quality assessment – at the purchase occasion by combining external information such as price/promotional cues, on-package attribute information, etc. and internal information stored in their memory obtained through prior consumption experience, word-of-mouth effects and previous exposure to advertising messages. In the context of frequently purchased consumer goods such as laundry detergent, ketchup, etc. – product categories that have been typically used in the choice modelling literature – it is reasonable to say the in-store information is not very diagnostic and consumers rely mainly on memory-based information to construct brand evaluations. Thus, in the above-mentioned example of laundry detergent, she will “construct” her preference for Tide, Surf, etc. along with their different variants, relying on information about these brands from prior consumption experience. The issue then is: *what quality-relevant information do consumers retrieve to make quality judgment that dictates their choices?*

A major strand of literature in cognitive psychology views “memory” as comprising 2 parts: (1) declarative or “explicit” memory, and (2) procedural or “implicit” memory. While implicit memory is characterized by a lack of conscious awareness in the act of recollection, explicit memory requires conscious recollection of previous experience. In the context of memory-based judgment, explicit memory is the relevant memory component. This literature again posits explicit memory being comprised of two sub-systems: (1) “episodic” memory and (2) “semantic” memory. These are conceptualized as “two information processing systems that (a) selectively receive information from perceptual systems or other cognitive systems, (b) retain various aspects of this information, and (c) upon instructions transmit specific retained information to other systems, including those responsible for translating it into behaviour and conscious awareness” (Tulving, 1972).

Episodic memory is a more or less faithful record of a person’s experience. Thus, every “item” in episodic memory represents information stored about the experienced occurrence of an episode or event. A perceptual event can be stored in the episodic system solely in terms of its perceptible properties or attributes, and is stored in terms of its autobiographical reference to the already existing contents of the episodic memory store. In contrast, inputs into the semantic memory system have two sources – perception and thought. When input is perceptual, perceptible attributes of stimulus events are important only to the extent that they permit unequivocal identification of semantic referents of the events. These properties themselves are not recorded in semantic memory. Inputs into the semantic memory system are always referred to an existing cognitive structure, that is, they always have some cognitive reference and the information they contain is information about the referent they signify rather than information about the input signal as such.

To understand the distinction between episodic and semantic memories in the context of experiential quality learning, let us re-visit the case of the consumer making a brand choice in the laundry detergent category. She may have had prior consumption experiences with a subset of brands. Taking the case of Tide HE as an example, she might remember the specific "episodes" of brand usage. She might remember that when she used Tide HE last time to wash a load of clothes consisting of mostly cotton garments, she had also added 2 tablespoons of bleach and that she was "fairly satisfied" with the outcome. She might also recall that sometime back she had used Tide HE on a heavy load of clothes of mixed fabric – cotton, silk shirts, designer georgette saris – along with fabric softener and she was "very unsatisfied" with the outcome. These are examples of recall from episodic memory system. Alternatively, she may recall the "overall evaluation" that she had about Tide and the other competing brands while making the brand choice in the current purchase occasion. This is an example of recall from semantic memory system. Note that while the semantic memory of "overall evaluation" are based on quality signals contained in prior consumption "episodes", the recalled item is the mental construct "brand evaluation" without the recall of specific episodic quality signals.

Viewed from this perspective, the extant quality-learning literature (e.g. Erdem and Keane, 1996; Mehta, Rajiv and Srinivasan 2003, 2004) models memory-based judgment based on semantic memory system alone. A consumer has a mental construct – viz., overall quality index – as well as the rules for manipulation of this construct – viz., Bayesian updating rule – in her semantic memory. As additional quality signals based on consumption episodes arrive, the consumer updates the quality index construct and stores this revised value in the semantic memory, without storing the specific signal associated with the consumption episode.

The *primary* purpose of this paper is to propose a dual-process model of memory-based judgment allowing for recall from both semantic and episodic memory systems. Mehta et al. (2004) has shown how forgetting affects the quality-learning process and hence the memory-based brand choice. Thus our *secondary* objective is to look at how the evaluations change as a result of imperfect memory. We also wish to investigate how the magnitude forgetting varies across the two memory processes. It is important to note that it is the occurrence of memory error that allows us to statistically identify the two memory processes.

We calibrate the proposed model on scanner panel data in the laundry detergent category. We find that consumers are more likely to recall past consumption experiences to form a new evaluation at the point of purchase, rather than recalling an existing belief from semantic memory. We also find, in line with cognitive literature, episodic memory is more vulnerable to forgetting than semantic memory. The model that accounts for recall from both memory systems is able to capture the effect of forgetting better and leads to less estimation bias. In addition, the proposed model also performs better in both estimation and hold-out sample in terms of predictive power.

2. Related Literature

There is a wide variety of research that suggests that individuals use different types of processes for storing and retrieving information from their memory (Chaiken 1980; Cacioppo and Petty 1982; Denes-Raj & Epstein, 1994). These differences in processing are a function of the type of memory that is active during the encoding and recall processes. Tulving (1972; 1983) coined the term episodic and semantic memory to describe the encoding processes which might lead to these differences. In processing of information using episodic memory, the person uses all the experiences stored about the product in detail while processing the information using semantic memory, they make use of the overall evaluation/impression about the product. In the literature the recall of the overall quality judgment/impression for decision making has been referred to by different names - heuristic processing (Chaiken 1980; Cacioppo and Petty 1982), attitude-based processing (Sanbonmatsu and Fazio 1990), category-based processing (Fiske and Pavelchak 1986) or holistic processing (Nisbett et al. 2001). All these are similar in concept and vary very slightly and in this paper we refer to this as *semantic processing.* Similarly, the recall of entire set of information/experiences is referred to as attribute based processing (Mantel and Kardes 1999), piecemeal based processing (Fiske and Pavelchak 1986) or analytic processing (Nisbett et al. 2001). These are also similar in concept and in the paper, we refer to this collectively as *episodic processing*.

According to Tulving (1983), accessing information from episodic memory requires conscious effort and that from the semantic memory can be accessed in a relatively easier fashion. This means that information processing and accessing reflect the differences in the involvement of the consumers and their inherent traits as well as the differences in the circumstances when the processing happens. When consumers are making a judgment, they use the memory they have encoded to help them make their decision. Depending upon their need for cognition (Srull, Lichtenstein and Rothbart 1985) or motivation towards accuracy

(Hutchinson and Alba 1991) or their expertise level (Alba and Hutchinson 1987), the individuals either carry out more elaborate processing of their memory making use of the entire set of experiences they have had or carry out a simpler processing with recall of an overall prior judgment. Nisbett et al. (2001) has suggested that this propensity varies with ethnicity and Meyers and Maheswaran (1991) has shown that this is a gender trait. Fiske and Pavelchak (1986) has suggested that people might consistently do only one type of processing. Sujan (1985) suggests that more experienced consumers will go for semantic processing. We incorporate most of these variations into our model to test the effects of these traits on different types of processing.

Recall of information from either of the types of memory leads to biases. Cook and Flay (1978), Estes (1997), and Roediger and McDermott (2000) show that forgetting is a common phenomenon and this would bias the memory being recalled. Rubin and Wenzel (1996) show that forgetting increases with passage of time which is consistent with Cook and Flay (1978) who show that there is a decay of attitude persistence with time. However, there has been limited evidence as to which type of memory is more subject to distortion. Snodgrass (1997) suggests that experiential information is the most fragile, context-dependent, and therefore more subject to distortion. Therefore, in our results, we expect episodic memory processes to be more subject to biases.

In the choice model literature, there has been increasing efforts to incorporate behavioural theories into the econometric model to understand the process better. Forward looking consumers were modelled using dynamic models (Erdem and Keane 1996). Mehta et al. (2003) modelled the consideration set formation of the consumers. The same author/s in their 2004 paper tried to look at the impact of forgetting in consumer’s brand choice decisions. In this paper we extend this stream of literature by incorporating the dual process model of memory retrieval and decision making as well as explore the effect of biases in the memory retrieval processes.

3. Model Development

In this section we discuss the modelling details on the choice decision by a consumer who may use either the semantic or the episodic memory. In section 3.1, we discuss the model primitives. In section 3.2, we describe the memory evolution of both semantic and episodic memory and how the consumer makes her choice decision based on the two memory systems. In section 3.3, we discuss how forgetting works in each of the memory systems. From section 3.4 onwards, we discuss the models from econometrician's perspective and present the likelihood function in section 3.5. Finally, in section 3.6, we compare the asymptotic properties of the posterior mean and variance across these two memory systems.

3.1 Model Primitives

Consider a product category with $j = 1, \ldots, J$ brands with the true quality of brand j being q_j. The consumer learns about the brand quality through their consumption experiences. However, even after multiple consumptions, the consumer would still be uncertain about the true quality as each consumption experience brings her only a "noisy" signal about the "true" quality.

At consumption occasion t[1], after the product is consumed, the consumer receives a signal $\lambda_{j,t}$[2]. Since consumption experience is inherently "ambiguous" (Hoch and Ha 1986)

[1] Here, the symbol t should be interpreted as the consumer's t^{th} purchase incidence in the category and not the calendar time.
[2] We assume that the consumer receives this quality signal just prior to the next purchase occasion i.e., there is an infinitesimally small time gap ι between the receipt of quality signal at consumption and the next choice task.

due to perceptual errors, inherent variability in product quality and context specific factors, the quality signal received by the consumer will be a sum of true quality and other noises, i.e.

$$\lambda_{j,t} = q_j + \eta_{j,t} \ldots \ldots (1)^3$$

where q_j is the true quality of brand j, $\eta_{j,t} \sim N(0, \sigma_\lambda^2)$ stands for the inherent quality variation. Thus the quality signal $\lambda_{j,t}$ is a random variable from $N(q_j, \sigma_\lambda^2)$.

It is to be noted that the consumer is unable to distinguish between the true quality q_j and the inherent variation in quality, $\eta_{j,t}$. Hence, as far as the consumer is concerned, the quality specific component, $\lambda_{j,t}$is a random variable from the normal distribution $\lambda_{j,t} \sim N\left(q_j, \sigma_\lambda^2\right)$.

At the beginning of the purchase history, the consumer's initial belief about product quality is, $q_{j,0} \sim N(\omega_o, \psi_0^2)\ \forall\ j$ where ω_0 is her expectation and ψ_0^2 is her uncertainty about brand's quality at t=0. With more purchases, the consumer uses realized quality signals $\hat{\lambda}_{j,t}$, to either form a new belief or to update a prior belief. At purchase occasion t, the consumer uses this latest quality belief $q_{j,t} \sim N\left(\omega_{j,t}, \psi_{j,t}^2\right)$ to form her utility function. Since the consumer is assumed to be risk neutral, thus she uses expected utility for brand choice*:*

$$E_t U_{j,t} = E\left(q_{j,t}\right) - \theta p_{j,t} \ldots \ldots (2)$$

where $p_{j,t}$ is the price of brand j and θ is the consumer's price sensitivity.

3.2 Memory Formation and Evolution

As discussed in the introduction, the consumer might use either episodic or semantic memory for her choice decision at each purchase occasion. In this section, we lay out our

[3] For notational convenience, we suppress subscript 'i' for individual consumer. We will bring it back when we layout our likelihood functions.

mathematical formulation for both memories. Specifically, in section 3.2.1, we discuss the evolution of contents in both semantic and episodic memory; in section 3.2.2, we discuss how the consumer makes her choice based on the recalled values at the purchase occasion.

3.2.1 Evolution of Memory

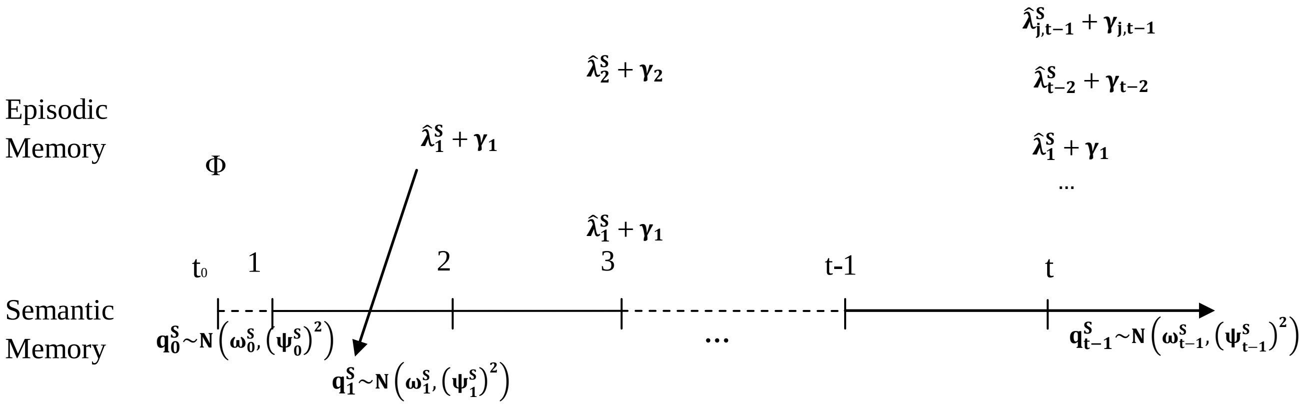
Semantic and episodic memories are two distinctive but related memory systems. Semantic memory records overall evaluations that are context free but is formed based on specific episodes. In this section, we discuss in detail the evolution of each memory system before the *t*-1th consumption occasion, which happens at a small time ι before purchase occasion *t*.

<u>Evolution of Semantic Memory</u>: Semantic memory contains the overall brand evaluations that are continuously updated as the consumer gets additional consumption signals. It does not contain any context specific information about the product quality. In addition, it also contains the rules for updating the belief by the consumer, which is assumed to follow a Bayesian updating process.

At the beginning of her consumption history, a consumer has prior beliefs about the brands based on external information such as brand name (national/store brand/private label), advertising, word-of-mouth, etc. Hence, what is stored in the semantic memory is her prior knowledge $\mathrm{q}_0^S \sim N(\omega_0^S, (\psi_0^S)^2)$ about the overall quality of the brand, which is assumed to be same across brands[4]. As she purchases more in the category, this initial prior gets updated whenever a consumption signal is received. The evolution of stored content in semantic memory is graphically presented at the bottom half of Figure1.

[4] Here we use super script 'S' to stand for *stored* values, super script 'R' to represent *recalled* values.

Figure 1: Evolution of both Semantic and Episodic Memory for brand j



Next, we discuss how this overall quality belief gets updated from consumption occasion to consumption occasion in semantic memory. Let the consumer buy brand j at purchase occasion $t-1$. Upon consumption, the consumer receives a realized quality signal $\hat{\lambda}^S_{j,t-1}$, which is used to update her prior belief in the semantic memory.

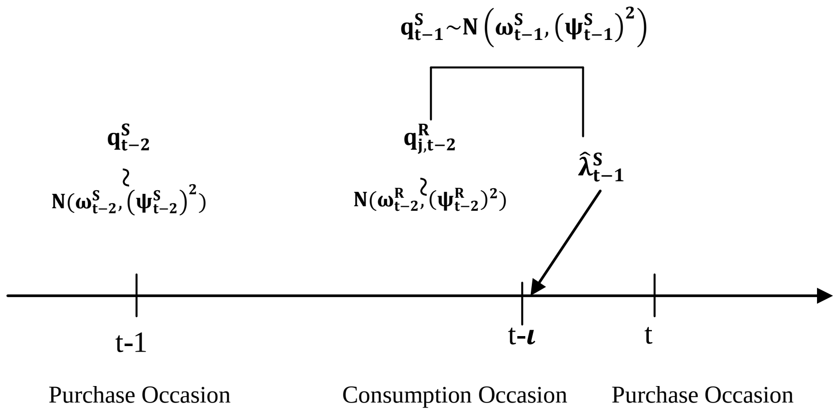
To update her belief, she also needs to recall the prior quality belief which was updated in the last consumption occasion and stored in her semantic memory. The consumer thus recalls $q^R_{j,t-2} \sim N\left(\omega^R_{j,t-2}, \left(\psi^R_{j,t-2}\right)^2\right)$ where $q^R_{j,t-2} \neq q^S_{j,t-2}$ as a result of forgetting due to passage of time. The details of how the consumer recalls the stored quality belief will be discussed in the section 3.3 for exposition purpose.

The consumer then uses this recall of the prior belief and the newly received signal $\hat{\lambda}^S_{j,t-1}$, to update her quality belief following Bayesian rule as described in equation (3). This process is detailed in Figure 2.

$$\omega^{S,SM}_{j,t-1} = \frac{\frac{\omega^R_{j,t-2}}{\left(\psi^R_{j,t-2}\right)^2} + d_{j,t-1} \cdot \frac{\hat{\lambda}^S_{j,t-1}}{\sigma^2_\lambda}}{\frac{1}{\left(\psi^R_{j,t-2}\right)^2} + d_{j,t-1} \cdot \frac{1}{\sigma^2_\lambda}};$$

$$\frac{1}{\left(\psi^{S,SM}_{j,t-1}\right)^2} = \frac{1}{\left(\psi^R_{j,t-2}\right)^2} + d_{j,t-1} \cdot \frac{1}{\sigma^2_\lambda} \ldots \ldots \ldots (3)$$

Figure 2: Belief Updating in Semantic Memory for Brand j



<u>Evolution of Episodic Memory</u>: Episodic memory is a more or less faithful record of a person's experiences. In this context, it contains all the detailed context specific information about the product experience that the consumer has received over time. Each individual episode is stored in great details in this memory. In the detergent example, the consumer finds that a particular detergent is not only "good" but remembers that this particular detergent is good for washing a particular type of clothes using a particular method of washing, i.e., this detergent is extraordinarily effective in washing white cotton clothes using the hot water cycle in a washing machine. Dubé (2004) has suggested that consumers do take into account this context specificity when considering purchase of products leading to simultaneous purchase of multiple products. Thus, the context specific details of the consumption signal get stored in the episodic memory.

The upper half of Figure1 illustrates how episodic memory evolves along a consumer's purchase history. At t = 0, since the consumer has never purchased any product in the category, her episodic memory is basically an empty set Φ. As she gains more consumption experiences with various brands, her episodic memory becomes a time-specific and context-specific log of all these product experiences. Thus, the episodic memory at purchase occasion *t* is the set of realized quality signals, $\hat{\lambda}_1^S$, $\hat{\lambda}_2^S$,..., $\hat{\lambda}_{t-1}^S$received till the date for the sampled brands. However note that in the Figure 1.1, for the purpose of ease of depiction, we do not include the purchase dummy $d_{j,t-1}$ where $d_{j,t-1} = 1$ if a consumer buys brand j at purchase occasion $t-1$ and $d_{j,t-1} = 0$ when she does not. Instead, we assume that if the consumer purchases the same brand across multiple time periods, how both the memory systems would evolve.

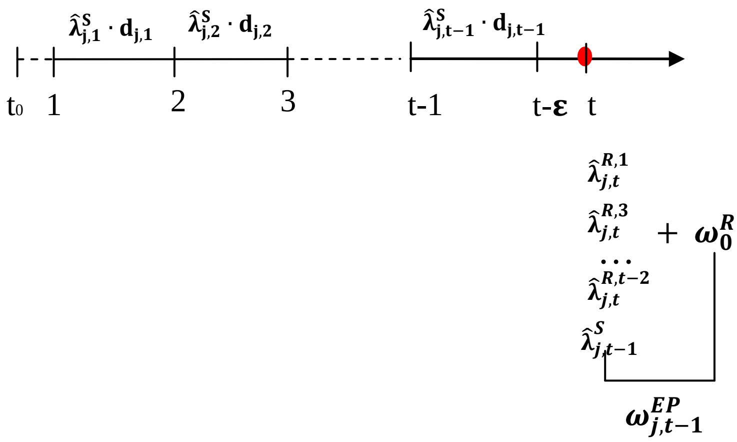
3.2.2 Memory Retrieval and Choice

Shortly after the consumption, the consumer arrives at the next purchase occasion *t*. Here, the consumer uses her product valuation to choose a brand that gives her the largest expected surplus, as described by equation (2). To make the choice, she might use the quality belief stored in her semantic memory or she might construct a new belief by recalling all her past consumption episodes. We describe the recall for each of the process in detail below.

For Semantic Memory: During the purchase occasion *t*, if the consumer is using semantic memory, she will recall the recently updated overall evaluation stored in her semantic memory (as per equation 3).Since ι is an infinitesimally small time gap, the posterior $\omega_{j,t-1}^{S,SM}$ formed as a result of previous consumption (at ι before t) can be recalled perfectly at purchase occasion *t*. This is similar to the previous learning models (Erdem & Keane 1996; Mehta, Rajiv & Srinivasan 2003, 2004) where the consumer will always recall a formed belief from her semantic memory rather than forming any new belief.

For Episodic Memory: If a consumer is using the episodic memory during the purchase occasion, she will be constructing an overall belief by retrieving all of her previously realized sequence of consumption signals together with the initial prior as shown in Figure 3. Since these consumption experiences are recalled from episodic memory, they are usage episodes with rich context information. Due to the time gaps between current period and the periods when these consumption signals were received, consumers are not able to recall these signals exactly. We use $\hat{\lambda}^{R}_{j,\tau,t}$ to denote the value of recalled consumption signals, where ‘τ’ represents the consumption occasion when the signal was received. For example, $\hat{\lambda}^{R}_{j,1,t}$ is the value of a signal that was received by consumer at consumption occasion 1 but recalled at the purchase occasion *t*. Here too $\hat{\lambda}^{R}_{j,\tau,t} \neq \hat{\lambda}^{S}_{j,\tau,t-1}$ due to the forgetting with the passage of time. Again, we shall discuss the details of this forgetting in the following section.

Figure 3: Construction of a New Belief



Let $\omega^{EM}_{j,t-1}$and $\left(\psi^{EM}_{j,t-1}\right)^2$be the new belief which is constructed by following the Bayesian rule.

$$\omega^{EM}{}_{j,t-1} = \frac{\frac{\omega_0^R}{\left(\psi_0^R\right)^2} + \sum_{\tau=1}^{t-1} \frac{\hat{\Lambda}^{R}_{j,\tau}}{(\sigma_\Lambda^R)^2} \cdot d_{j,\tau}}{\frac{1}{\left(\psi_0^R\right)^2} + \sum_{\tau=1}^{t-1} \frac{d_{j,\tau}}{(\sigma_\Lambda^R)^2}};$$

$$\frac{1}{\left(\psi^{EM}_{j,t-1}\right)^2} = \frac{1}{\left(\sigma^R_\lambda\right)^2} + \sum_{\tau=1}^{t-1} \frac{d_{j,\tau}}{(\sigma^R_\lambda)^2} \dots \dots (4)$$

Once the consumer has obtained her new belief from either of the memories, she uses the expected quality to compare across the brands to choose a brand that maximizes her expected utility as per equation (2).

3.3 Modeling of Forgetting

With the passage of time, the consumer is not able to recall these quality perceptions or consumption signals perfectly. In this section, we discuss what have been recalled by the consumer and how they use these recalled values for belief updating and formation.

Forgetting in the Semantic memory As discussed in section 3.2.1, at the consumption occasion, the consumer needs to recall the prior belief for belief updating. This prior belief $q_{j,t-2}^{S} \sim N\left(\omega_{j,t-2}^{S,SM}, \left(\psi_{j,t-2}^{S,SM}\right)^2\right)$ was stored at consumption occasion t-2. Due to the time lapse between last and current consumption, the consumer cannot recall the prior belief $q_{j,t-2}^{S} \sim N\left(\omega_{j,t-2}^{S,SM}, \left(\psi_{j,t-2}^{S,SM}\right)^2\right)$ exactly as it was stored. Instead, $q_{j,t-2}^{R} \sim N\left(\omega_{j,t-2}^{R}, \left(\psi_{j,t-2}^{R}\right)^2\right)$ is recalled at this moment. Here we use the superscript '*R*' to differentiate what was stored from what is recalled. Clearly, due to forgetting, $\omega_{j,t-2}^{R} \neq \omega_{j,t-2}^{S,SM}$ and $\left(\psi_{j,t-2}^{R}\right)^2 \neq \left(\psi_{j,t-2}^{S,SM}\right)^2$ in most cases. However, as a consumer who is aware of her imperfect memory, she knows she has recalled a different value from what was stored. She does not know the exact recall error, as otherwise she would have corrected it. With the awareness of the recall error, the consumer will give larger weight to more accurate recall and smaller weight to less accurate weight.

Forgetting in the Episodic Memory At the purchase occasion, if the consumer decides to construct a new belief, she needs to recall all the past consumption episodes. Here too, since the consumer is aware of her imperfect recall, she recalls a consumption experience, $\hat{\lambda}_{j,\tau,t}^{R}$ with uncertainty $\phi_{j,\tau,t}^{2}$ due to both quality fluctuation and forgetting.

3.4 The Econometrician's Perspective

The consumer sees the actual consumption signals and they know the recalled and the adjusted values of the brand quality. However, the econometrician does not see any of the above. As such, the econometrician has to build a sensible model by mimicking the consumer's behavior.

3.4.1 Modeling the Semantic Memory

At time ι before the purchase occasion t, a consumer recalls her overall knowledge about the brand as $q^{R}_{j,t-2}$ from her semantic memory, but the econometrician does not see $q^{R}_{j,t-2}$. However, the econometrician knows that the recalled value is nothing but the stored value plus the recall error. He also knows that recall errors grow with the passage of time. Hence, the econometrician can infer what is recalled by a consumer. The econometrician has access to the purchase history of the consumer, thus, he can infer the stored belief $q^{S,SM}_{j,t-2}$ and he does not forget. The econometrician can therefore construct the recalled prior belief $q^{R}_{j,t-2} \sim N\left(\omega^{R}_{j,t-2}, \left(\psi^{R}_{j,t-2}\right)^{2}\right)$, which is his best possible guess as constructed

$$q^{R}_{j,t-2} \sim N(q^{S}_{j,t-2} + \nu_{j,t-2}\varphi_{j,t-2}, \ \varphi^{2}_{j,t-2}) \ldots\ldots\ldots\ldots(5)$$

In equation (5), the econometrician constructs $q^{R}_{j,t-2}$ from $q^{S,SM}_{j,t-2}$ since he only knows $q^{S,SM}_{j,t-2}$ but does not see $q^{R}_{j,t-2}$. He constructs the forgetting error as $\nu_{j,t-2}\varphi_{j,t-2}$, where $\nu_{j,t} \sim N(0,1)$, is a random draw from a standard normal distribution, which allows forgetting to happen in either direction. $\varphi_{j,t-2}$ is the scale of this forgetting error that is modelled as an exponential function of time lapse between the value is stored and that is recalled.

$$\varphi_{j,t-2}^{2} = \left(\psi_{j,t-2}^{S,SM}\right)^{2}\left(e^{B^{SM}w_{t-1}} - 1\right)\ldots\ldots(6)$$

where w_{t-1} is the absolute calendar time in weeks between period *t* and period *t*-1 and B^{SM} ($B^{SM} > 0$) measures consumer's tendency to forget. $\left(\psi_{j,t-2}^{S,SM}\right)^{2}$is the posterior variance of the consumer's belief in period t-1. Similarly, $\varphi_{j,t-2}^{2}$ is the additional uncertainty brought by forgetting, as the econometrician knows that the consumer recognizes the noises added due to forgetting.

Since $q_{j,t-2}^{S} \sim N\left(\omega_{j,t-2}^{S,SM}, \left(\psi_{j,t-2}^{S,SM}\right)^{2}\right)$, the unconditional distribution of $q_{j,t-2}^{R}$ can be expressed as

$$q_{j,t-2}^{R} \sim N(\underbrace{\omega_{j,t-2}^{S,SM} + \nu_{j,t-2}\psi_{j,t-2}^{S}\sqrt{e^{B^{SE}w_{t-1}} - 1}}_{\omega_{j,t-2}^{R}}, \underbrace{\left(\psi_{j,t-2}^{S,SM}\right)^{2} e^{B^{SM}w_{t-1}}}_{\left(\psi_{j,t-2}^{R}\right)^{2}})\ldots.(7)$$

Thus, $q_{j,t-2}^{R} \sim N\left(\omega_{j,t-2}^{R}, \left(\psi_{j,t-2}^{R}\right)^{2}\right)$ is the recalled prior from the econometrician's perspective.

3.4.2 Modelling the Episodic Memory

Now we discuss the econometrician's formulation for recalled consumption signals from the episodic memory. Here too, the econometrician does not observe the recalled values but he can infer the recalled values from the stored values in the similar.

$$\lambda_{j,\tau,t}^{R} \sim N(\lambda_{j,\tau,t}^{S} + \nu_{j,t}\phi_{j,\tau,t}, \phi_{j,\tau,t}^{2})\ldots\ldots.(8)$$

where $\lambda_{j,\tau,t}^{S}$ is the value of stored signal for brand *j* received at consumption occasion τ and recalled at purchase occasion t, $\nu_{j,t} \sim N(0,1)$ and $\nu_{j,t}\phi_{j,\tau,t}$ is the recall error. Since consumers recognize the added uncertainty associated with forgetting, $\phi_{j,\tau,t}^{2}$ will therefore be

$$\phi^2_{j,\tau,t} = \sigma^2_\lambda(e^{B^{EM}W_{\tau,t}} - 1)\ldots\ldots\ldots.(9)$$

where $W_{\tau,t}$ is the actual time in weeks between purchase occasion τ and purchase occasion *t*. B^{EM} ($B^{EM} > 0$) is consumer's tendency to forget under episodic retrieval, the equivalent of B^{SM} under semantic retrieval. Allowing the forgetting tendency to be different across episodic and semantic retrieval enables us to test the argument whether episodic memory is more vulnerable to forgetting compared to semantic memory. Therefore, $\lambda^R_{j,\tau,t}$ can be specified as

$$\lambda^R_{j,\tau,t} \sim N(q_j + \nu_{j,t}\phi_{j,\tau,t},\ \sigma^2_\lambda(e^{B^{EP}W_{\tau,t}}))\ldots\ldots.(10)$$

3.5 Likelihood Function

The consumer can deterministically make her choice decision by choosing a brand that maximizes her surplus. The econometrician uses similar utility maximization as the consumer i.e. equation (2).

$$E_tU_{j,t} = E(q_{j,t}) - \theta p_{j,t} + \varepsilon_{j,t}\ldots\ldots\ldots\ldots\ldots..(11)$$

where $\varepsilon_{j,t}$ is the unobservable to the econometrician. Since it is assumed to be a Type I extreme value distributed random error that is I.I.D. across all consumers, brands and purchase occasions, the econometrician can define the consumer's choice probability for each brand conditioned on the mode of processing is

$$\mathrm{Pr}_{\mathrm{i,j,t}}[d_{i,j,t} = 1|\mathrm{SE}] = \frac{\exp(\omega_{i,j,t}{}^{SE} - \theta \cdot p_{i,j,t})}{\sum_{\mathrm{j \in J}}(\ \omega_{i,j,t}{}^{SE} - \theta \cdot p_{i,j,t})}$$

$$\mathrm{Pr}_{\mathrm{i,j,t}}[d_{i,j,t} = 1|\mathrm{EP}] = \frac{\exp(\omega_{i,j,t}{}^{EP} - \theta \cdot p_{i,j,t})}{\sum_{\mathrm{j \in J}}(\ \omega_{i,j,t}{}^{EP} - \theta \cdot p_{i,j,t})} \ldots \ldots.(12)$$

Here, $\omega_{i,j,t}$ is the knowledge to the consumer *i* and it is not directly observable to the econometrician. He also does not know the realized values of the quality signals $\{\hat{\lambda}_{i,j,k}^{S}\}_{\mathrm{k}=1}^{\mathrm{t}-1}$ that are stored by consumers. However, the econometrician knows the distributions of both actual quality signals $\lambda_{i,j,t} \sim N(q_j, \sigma_\lambda^2)$, and the random shock for forgetting errors $\nu_{i,j,t} \sim N(0,1)$. He also knows the consumer's rule for belief updating and the law of forgetting. Using these, the econometrician can construct $\omega_{i,j,t}$ as a consumer does.

At purchase occasion *t*, a consumer knows for certain whether she has recalled a belief from her semantic memory or she has constructed a new belief with her episodic memory, but the econometrician does not. Hence, the econometrician needs to make a probabilistic assumption on the consumer's likelihood to use the episodic memory or the semantic memory. Laboratory studies use demographic or situational variables to predict consumer's tendency for using either of these memories. In our study, we use variables such as gender, age and product knowledge to predict the likelihood of recalling the belief from semantic memory. Thus, the probability of the consumer being the semantic type is

$$\Pr[\mathrm{SM}] = \frac{\exp\,(\alpha_\mathrm{i} + \beta \cdot \mathrm{X})}{1 + \exp\,(\alpha_\mathrm{i} + \beta \cdot \mathrm{X})} \dots \dots .. (13)$$

where $\alpha_\mathrm{i} \sim \mathrm{N}(\alpha\,, \sigma_\alpha^2)$ is an individual intrinsic tendency to use semantic memory and X is the matrix of the explanatory variables. The probability of the consumer using episodic memory follows naturally, i.e. $\Pr[\mathrm{EM}] = 1 - \Pr[\mathrm{SM}]$. Hence, the purchase probability for an individual *i* to choose brand *j* at purchase occasion t can be represented as

$$\Pr\,(d_{i,j,t} = 1 | \Lambda_{\mathrm{i,t-1}}, \mathrm{V}_{\mathrm{i,t-1}}, \alpha_\mathrm{i}, \Delta) = \Pr[\mathrm{SM}] \cdot \mathrm{P}_{\mathrm{i,j,t}}^{\mathrm{SM}} + \Pr[\mathrm{EM}] \cdot \mathrm{P}_{\mathrm{i,j,t}}^{\mathrm{EM}} \dots\dots (14)$$

where $P^{SM}_{i,j,t}$ and $P^{EM}_{i,j,t}$ are the choice probability conditional on consumer's retrieval mode. $\Lambda_{i,t_i} \equiv \{\lambda_{i,1,s} d_{i,1,s}, \dots \lambda_{i,j,s} d_{i,J,s}\}_{s=1}^{t-1}$ is the string of signals that are received by consumer till purchase occasion *t*, $\Gamma_{i,t_i} \equiv \{\gamma_{i,1,s} d_{i,1,s}, \dots \gamma_{i,j,s} d_{i,J,s}\}_{s=1}^{t-1}$ is the set of context specific information that is associated with the string of signals received and $V_{i,t_i} \equiv \{\upsilon_{i,1,s}, \dots, \upsilon_{i,j,s}\}_{s=1}^{t-2}$ is a matrix of $J \times t_i$ iid standard normal random errors. Δ is the vector of population parameters $\{\beta, \theta, q_1 \dots q_1, \sigma_\lambda, \sigma_\alpha\}$. With equation (13) and (14) defined, we can now lay out the conditional individual likelihood function as

$$L_i\left(D_{i,t_i} | \Lambda_{i,t-1}, \Gamma_{i,t_i}, V_{i,t_i}, \alpha_i, \Delta\right) = \prod_{t=1}^{t_i} \prod_j^J \Pr\left(d_{ij,t} = 1 | \Lambda_{i,t_i}, \Gamma_{i,t_i}, V_{i,t_i}, \alpha_i, \Delta\right)^{d_{i,j,t}} \ldots\ldots(15)$$

The unconditional likelihood for individual *i* is specified as

$$L_i\left(D_{i,t_i} | \Delta\right) =$$

$$\int_{\alpha_i} \int_{\Lambda_{i,t_i}} \int_{V_{i,t_i}} L_i\left(D_{i,t_i} | \Lambda_{i,t_i}, V_{i,t_i}, \alpha_i, \Delta\right) \cdot$$

g1Vi,tig2Λi,tig3αig4(Γi,ti)dVi,tidΛi,tidαidΓi,ti.......(16)

Here, $g_1(\cdot)$ is the joint distribution of the random shocks and $g_2(\cdot)$ is the joint distribution of the consumption signals. In addition, $g_3(\cdot)$ is the distribution for consumer's individual tendency to use semantic memory.

Since the numerical computation for the above likelihood (16) with multidimensional integration is prohibitively expensive, we resort to simulated likelihood with R draws of $\{V_{i,t_i}, \Lambda_{i,t_i}, \alpha_i\}$. We get the estimation of the individual likelihood as follows

$$\hat{L}_i\left(D_{i,t_i} | \Delta\right) = \frac{1}{R} \sum_{r=1}^{R} L_i\left(D_{i,t_i} | \Lambda^r_{i,t_i}, \Gamma_{i,t_i}, \alpha^r_i, V^r_{i,t_i}, \Delta\right) \ldots\ldots(17)$$

To reduce the asymptotic bias in the estimate of the likelihood, we take the number of draws R=300. Once we get the individual likelihood contribution, we compute the estimate of log-likelihood for the entire sample of N consumers as

$$l\left(\left\{\mathrm{D}_{\mathrm{i,t_i}}\right\}_{\mathrm{i=1}}^{\mathrm{N}} | \Delta\right) = \sum_{\mathrm{i=1}}^{\mathrm{N}} \ln \hat{L}_i\left(\mathrm{D}_{\mathrm{i,t_i}} | \Delta\right) \text{.....(18)}$$

Finally, the parameters can be estimated by maximizing the log-likelihood function as follows:

$$\Delta_{\mathrm{MLE}} = \arg\max_{\Delta} l\left(\left\{\mathrm{D}_{\mathrm{i,t_i}}\right\}_{\mathrm{i=1}}^{\mathrm{N}} | \Delta\right) \text{.....(19)}$$

3.6 Asymptotic Property of Posterior Belief

In the above elaboration, we see that forgetting takes place in both semantic and episodic retrieval and the same set of information goes into the formation of quality belief by the consumer in each case. However, the actual process of forming this belief varies across the two types of retrieval. In the case of semantic retrieval, consumer uses the prior belief together with the latest quality signal for updating. Thus, she forgets the prior belief. In episodic retrieval, consumer retrieves all the previously received consumption signals together with the latest signal for belief updating. Thus she forgets the retrieved signals. This raises the following question: Given infinite consumptions 1) does the posterior belief converge to true value with the existence of forgetting in either semantic memory or episodic memory? 2) If not, which mechanism gives a posterior closer to the true quality?

To facilitate the illustration, we set the inter-purchase time between any two consecutive purchases to be *W* and the forgetting error ν be constant across all purchase occasions. In addition, since we have to compare across the two mechanisms, we also assume equal forgetting rates ($B^{SE} = B^{EP} = b$) across the two memories for a fair comparison.

Proposition 1: With the existence of forgetting, consumers can never be certain about her posterior quality expectation even after infinite consumptions. However, their uncertainty does approach certain constant i.e.

$\lim_{N\to+\infty}\left(\frac{1}{\psi_N^{SE}}\right)^2 = \frac{e^{bW}}{e^{bW}-1}\cdot\frac{1}{\sigma_\lambda^2}$ and $\lim_{N\to+\infty}\left(\frac{1}{\psi_N^{EP}}\right)^2 = \frac{e^{bW}}{e^{bW}-1}\cdot\frac{1}{\sigma_\lambda^2+1}$. Also, at each period $\left(\frac{1}{\psi_N^{SE}}\right)^2 > \left(\frac{1}{\psi_N^{EP}}\right)^2$ (Please see the appendix for the detailed proof).

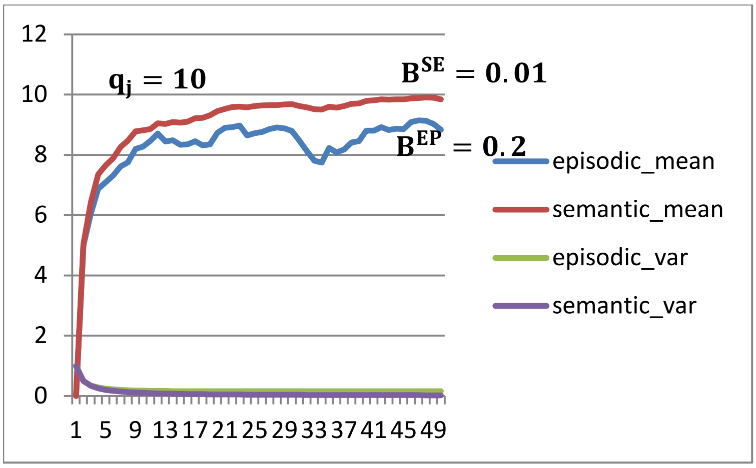
The latter is not a surprising result that at each stage, semantic memory has a larger precision than episodic memory. The reason is that signals are deposited into episodic memory with context specific information, thus leading to larger variance of the consumption signals. It is more interesting to know that even with infinite consumptions; the posterior variance is never decreased to zero, but to a limiting value. This is because, in the case of perfect recall, every consumption signal takes the same weight in updating. Hence, each signal increases the consumer's precision about the true quality with the same impact. In the presence of imperfect memory, the earlier signals are not recalled intact. Hence, they have less impact on improving the precision compared to the later signals. Therefore, with forgetting, consumer's uncertainty is never resolved completely.

Proposition2: With the existence of forgetting, the posterior mean of both semantic and episodic retrieval will never converge to the true quality even after infinite consumptions. In the case of episodic retrieval, the posterior mean will converge to a constant i.e. $\lim_{N\to+\infty} \omega_N^{EP} \to C$, but in the case of semantic retrieval, it is not converging.

Despite the same information set for both memories, i.e., the consumption signals received and the initial quality belief, the limiting posterior belief evolves in different fashions under these two memories. This is because forgetting acts differently in these two types of memory systems. In the case of semantic retrieval, forgetting occurs to the prior belief and arrival of this error not only gets accumulated in each period but also persists in the following periods. Hence, when $N\to \infty$, the accumulated errors are non-convergent, i.e., a set of errors on error. In the case of episodic retrieval, though the quality signals are imperfectly recalled from each previous period, but they get assimilated as time passes by. They do not get added to the following periods, thus limiting the magnitude of the total error. It seems that the constructed belief from episodic memory is more precise than using the prior overall belief from the semantic memory. This could be because in constructing a belief, the consumer needs to use more cognitive resources and she would do so only if the end result of taking this effort is worthwhile.

However, note that this conclusion is based on the assumption of equal forgetting tendency, namely, $B^{SE} = B^{EP}$. When $B^{SE} \neq B^{EP}$, it is difficult to say which memory is better under a limited learning setting. Figure 4 is a simulated example with $B^{EP} > B^{SE}$ and it shows that semantic memory can be better than episodic memory. The figure also shows evidence of proposition 1.

Figure 4: Simulation Plot: Evolution of Posterior Mean and Variance



4. Data, Estimation and Results

4.1 Data

For model calibration and analysis, we use the detergent category from the IRI scanner panel data (Bronnenberg, et al. 2008). The panel data is collected from both grocery and drug stores in two markets, Eau Claire, Wisconsin and Pittsfield, Massachusetts. The brands included for analysis are Tide, Xtra, Purex, Arm,All and Other, where the national brands account for a total of 76.39% market share. The detergent data set has in total 836 panelists who have at least 2 purchases in the observation span. We choose panelists whose total times of purchases range from 8 to 40. This leaves us with 144 panelists (40 male and 104 female), from which we randomly select 40 subjects as our holdout sample. The estimation sample has 1776 observations and the holdout sample has 568 observations. The summary statistics for the entire sample are given in Table 1.

Table 1:Descriptive Statistics for Detergent Category

Brands	Market Share	Mean Price (Std. err.)*
Tide	0.213	1.319(0.579)
Xtra	0.195	0.375(0.038)
Purex	0.085	0.606(0.122)
Arm	0.103	0.735(0.151)
All	0.168	0.960(0.234)
OTH	0.236	0.874(0.250)

*** The mean price is price per 16 oz detergent.**

4.2 Model Free Evidence

In this section, we provide some model free evidence to show the data has both learning and forgetting effect.

<u>Learning Effect:</u> If there is indeed some learning about the brand, then we shall see more switching at the beginning of a consumer purchase history and less switching with the progression of the purchase history. The reason is that at the beginning, when the consumer has limited knowledge to differentiate among the brands, price dictates her choice. However, with more purchases and once the consumer is better informed about the quality differences between the brands, then larger price differences are needed to induce brand switching. Hence, to examine such effect, we construct a variable called switching in the following fashion

$$\text{switching}\begin{cases} = 1 \text{ if consumer buys a different brand from last purchase} \\ = 0 \text{ if consumer buys the same brand as last purchase} \end{cases}$$

Our hypothesis then follows if there is learning, there must be more switching in the early purchase stage than in the late purchase stage.

$$\mathrm{H_0}: \overline{\mathrm{S_E}} = \overline{\mathrm{S_L}}$$

$$\mathrm{H_A}: \overline{\mathrm{S_E}} > \overline{\mathrm{S_L}}$$

In the hypothesis, the subscript 'E' stands for early purchase stage and 'L' stands for later purchase stage. Since the length of Consumer's purchase history ranges from 8 to 40 times, we use different thresholds to define early stage, such as the first 3, 4, 5 times of total purchase. We intend to use absolute times of switching as comparison statistics. However, due to different lengths of purchase history, this comparison is implausible, as switching 3 times in a late stage with 15 purchases is less frequent than 2 times in 2 purchase occasions at early stage. Hence, we use the percentage of switching as our comparison statistics.

$$\mathrm{S_E} = \frac{\text{Switching times in early stage}}{\text{Total Purchase Time in Early Stage} - 1}$$

$$\mathrm{S_L} = \frac{\text{Switching times in late stage}}{\text{Total Purchase Time in late Stage}}$$

Note that we have minus one in the denominator of early stage as the first purchase is random, we cannot say whether it is a switching or not.

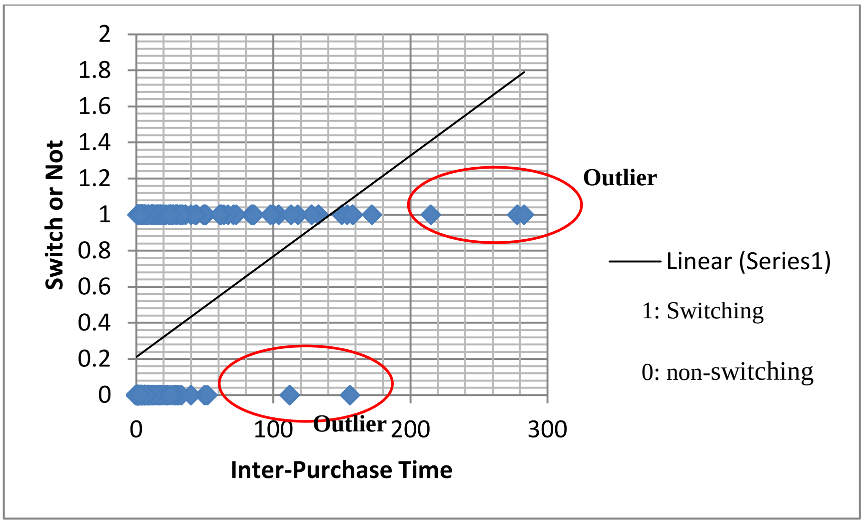
We use the paired-sample t-test to compare the means of two populations. As shown in Table 2 below, we find the alternative hypothesis is supported when early stage is defined as the first 3 or 4 purchases.

Table 2: t-test for Learning Effect

Definition of early stage	t	H_A
First 3 purchases	2.869	Supported
First 4 purchases	2.868	Supported
First 5 purchases	1.495	Not Supported

<u>Forgetting Effect</u>: as aforementioned, time lapse between purchases is the major contributor to forgetting under the context of our research. The longer the time span, the more can be forgotten. From this perspective, forgetting decreases learning efficiency and leads to brand switching. If this is indeed the case, then we should observe from the data that longer inter-purchase time is accompanied with more switching. We then plot the distribution of inter purchase time for both switching and non-switching occasions. The first time purchases are deleted from the samples, leaving us with only 1,880 data points. From Figure 5 plot we found that switching occasions are accompanied with longer inter-purchase time than non-switching occasions.

Figure 5: Plot of Switched Purchase against Inter-Purchase Time



4.3 Model identification

In our model, we have the following parameters to estimate $\{ q_1, \ q_2, q_3, \dots, q_j$, $\sigma_\lambda, B^{SE}, B^{EP}, \alpha, \sigma_\alpha, \beta\}$. To facilitate the discussion, we reiterate the meaning of the parameters here. $\{q_1, \ q_2, q_3, \dots, q_j\}$ represents the set of mean quality of the brands under analysis and σ_λ describes the noise size of the consumption signals. B^{SE} and B^{EP} are the rate of forgetting under semantic and episodic processing, respectively. α and σ_α are the mean and variance for consumer's intrinsic inclination to employ semantic versus episodic processing, whereas the β's are the demographic parameters that might help to explain consumer's preference for semantic to episodic processing.

First, we discuss how we can identify the mean quality $\{q_1, \ q_2, q_3, \dots, q_j\}$ as well as the quality variance σ_λ^2 for each brand. As we mentioned before, consumers are able to see the realized consumption signals $\hat{\lambda}_{j,t}$, thus using these signals to update their belief in a Bayesian fashion. Hence, the econometrician can estimate the mean quality and variance, should he observe a large sample of consumption signals from each brand. In our dataset, though the econometrician does not observe the realized consumption signals, he has access to a large sample of cross section choices made by consumers, and he also knows consumer's rule for belief updating. Hence, with both pieces of information the econometrician can infer the values of the consumption signals received by consumers and estimate the brand mean quality and quality variance. As usual, not all the q_j's can be identified, hence, we set one $q_j = 0$ as the base category.

Second, we see how we can identify the rate of forgetting, B^{SE} and B^{EP} from the data. When people are forgetting, but assumed to recall perfectly, the effect of forgetting is attributed to consumption signals. Thus, brand quality mean and variance are estimated with systematic bias. We are able to identify the forgetting rates B^{SE} and B^{EP} , as we assume that forgetting errors have the exponential relationship with the time lapse between the stored and recalled content. Hence, the next question is whether it is a plausible assumption to impose to allow such identification. Think of a case where consumers are recalling with perfect memory and forming their quality belief along the purchase history. Remember that at the beginning of purchase history, a consumer has the same prior belief across brands. Hence, any small price drop will motivate brand switching. However, with the progression of learning after multiple purchases, price reduction needs to be big enough (larger than quality difference) to induce brand switching. Since forgetting is lowering down learning efficiency and impeding convergence of brand quality to its true value, forgetting increases brand switching. Thus, we would expect purchases with long inter-purchase time would be accompanied with more switching than occasions with shorter inter-purchase time. This is shown in Figure 5. The model free evidence enables us to identify the rate of forgetting by using time lapse.

Last but not least, we discuss how we can identify consumer's heterogeneity in intrinsic inclination to use endowed or constructed belief, namely, N $(\alpha, \sigma_\alpha^2)$. With perfect memory, both episodic and semantic belief approaches the true value of the brand quality after infinite purchases. Moreover, at the end of each stage, semantic belief equals to episodic belief. This is self-evident as the information sets, namely the realized consumption signals, are the same at the end of each stage. In fact, it is forgetting, which varies across both processes, that allows us to identify consumer's intrinsic preferences over both processes. For example, due to different forgetting mechanisms, episodic belief predicts a choice of brand 1, but semantic belief predicts a choice of brand 2; while the actual choice is brand 2. Hence, more weight is attached to the semantic belief (Equation 13). Chintaqunta (1991) mentions that consumers are heterogeneous in their brand evaluation. Here, we argue that one of the reasons for such heterogeneity in preferences can be explained by the different memory retrieval modes employed by consumers. It is also due to consumers' heterogeneity in rate of forgetting.

Though B^{SE} and B^{EP} are the same across population, patterns of each consumer's inter-purchase time varies.

4.4 Parameter Estimates and Model Comparison

In this section, we estimate three competing models with detergent data sets. The parameters are by and large statistically significant. We compare the predictive power and goodness-of-fit of the proposed model (Model III) with two other competing specifications:

1) Model I is a Bayesian learning model assuming the consumer to have perfect memory, as Erdem and Keane (1996). Model II is learning with forgetting as Mehta, Rajiv and Srinivasan (2003). In this model, consumer is assumed to have imperfect memory, but she only involves semantic memory in the recall process. Model III is our proposed model, in which consumer can involve both semantic and episodic memory in the recall process, i.e., she can either retrieve a previously formed evaluation, or construct a judgment on the spot of purchase occasion. Whichever approach the consumer may employ, we allow forgetting happens in both approaches differently. Hence, Model III is the most flexible model, while Model II and Model I are restricted models, with Model I nested in Model II and Model II nested in Model III.
2) Since these models are nested, we use log-likelihood ratio test to see whether there is significant improvement between the models. In total, we ran 2 tests in each category. We test whether Model II is better than Model I and Model III is better than Model II. The log-likelihood values for each model in the estimated sample are reported at the bottom of Table 3. We found that the likelihood value is almost same for Model II and Model I, and this is due to the small magnitude of the forgetting variable. Despite of its significance, Model II is almost a Bayesian learning with perfect memory.

When comparing Model II with Model III, we calculate the LR statistics and get 166.690, which is highly significant for chi-square test. Hence, the proposed model is a better representation of consumer choice decision than the competing models.

3) In terms of predictive power, we use hit rates in both estimation sample and holdout sample as a measure. As shown in Table 4, Model III has the highest hit rates among all the models in both estimated and holdout sample. Model III has even a higher hit rate in the holdout sample than in the estimated sample. Hence, we conclude that Model III is superior to Models I and II in predicting individual consumer’s choice decision.

Table 3: Parameter Estimates for Competing Models Detergent Category

Parameter	Explanation	Model I	Model II	Model III
$\mathbf{q}_{\mathbf{Tide}}$	True quality of TIDE	**4.482 (0.447)**	**4.906 (0.446)**	**9.469 (1.342)**
$\mathbf{q}_{\mathbf{XTRA}}$	True quality of XTRA	**4.114 (0.365)**	**4.189 (0.364)**	**8.786 (1.271)**
$\mathbf{q}_{\mathbf{PUREX}}$	True quality of PUREX	**1.985 (0.369)**	**1.985 (0.371)**	**4.463 (0.912)**
$\mathbf{q}_{\mathbf{Arm}}$	True quality of ARM	**3.516 (0.436)**	**3.516 (0.436)**	**7.483 (1.234)**
$\mathbf{q}_{\mathbf{ALL}}$	True quality of ALL	**3.669 (0.367)**	**3.669 (0.368)**	**6.974 (0.967)**
$\boldsymbol{\sigma}_{\boldsymbol{\lambda}}$	Standard deviation for quality signals	**1.933 (0.315)**	**1.932 (0.455)**	**0.960 (0.059)**
$\mathbf{b}^{\mathbf{S}}$	Forgetting rate for semantic retrieval		**3.679E-12(1.627E-22)**	**1.789E-10(1.571E-23)**
$\mathbf{b}^{\mathbf{E}}$	Forgetting rate for episodic retrieval			**81.062 (35.536)**
$\boldsymbol{\alpha}$	Tendency to use semantic retrieval			**-5.414 (1.139)**
$\boldsymbol{\sigma}_{\boldsymbol{\alpha}}$	Standard deviation of $\boldsymbol{\alpha}$			**0.109 (0.048)**
$\boldsymbol{\theta}$	Price coefficient	**0.141 (0.077)**	**0.141 (0.077)**	**0.121 (0.062)**
$\boldsymbol{\beta}_{\mathbf{gender}}$	Gender coefficient			0.979 (0.666)
$\boldsymbol{\beta}_{\mathbf{age}}$	Age coefficient			-1.280 (1.256)
$\boldsymbol{\beta}_{\mathbf{pk}}$	Product knowledge variable			**12.554 (2.360)**
Log-likelihood		-4128.379	-4128.379	-4005.069

* The numbers in bold are significant at 0.05

Table 4: Hit rates in both Estimation and Holdout Samples

	Model I	Model II	Model III
Estimation Sample	0.487	0.487	0.532
Holdout Sample	0.481	0.481	0.556

4.5 Results and Discussion

We discuss the results of our model from the following five general perspectives.

True quality of brands: From the parameter estimates in Table 3, we found that true quality estimated by our proposed model is remarkably higher than the competing models. After a careful examination of the estimates in Table 3, together with the estimates in Mehta, Rajiv and Srinivasan (2003), we realize ignoring forgetting can lead to systematic under-estimation of brand true quality. Remember in the section of parameter identification, we mentioned that the econometrician can identify the true quality as he can infer the value of the consumption signals received by consumers through their brand choices as well as the price difference between the brands under comparison. Treating forgetful consumers as ones with perfect memory leads to erroneously over-weighting recalled prior or consumption signals. Hence, the econometrician will infer an under-valued consumption signal, thus under-estimate true brand qualities. Such under-estimation is positively correlated with the size of forgetting parameter and the inter-purchase time. When forgetting parameters are larger, the bias of the estimates is also larger. That is why in the detergent data set, true qualities estimated by the proposed model are much larger than those by the competing models as the forgetting rate parameters are larger in the proposed model.

Noise in consumption signals*:* Another observation from Table 3 is we found a decreased noise in consumption signals from Model I to Model II and to Model III. The estimator σ_λ is 1.933 for Models I and II, but 0.960 for Model III. This result is consistent with Mehta, Rajiv and Srinivasan (2003). The reason is that when consumers are forgetting but treated as if they have perfect memory, the noise due to forgetting is attributed to production volatility, σ_λ, thus overestimation.

Choice of Episodic versus Semantic Memory: One of our contributions in this paper is that we ask a fundamental question whether belief is endowed or constructed. In the proposed model, we allow consumer to choose between a constructed preference through recalling from episodic memory and an endowed belief from semantic memory. The population mean of consumer's intrinsic tendency to use an endowed belief—α is -5.414, showing that on average consumers are more likely to recall their past consumption experiences to form their preference than recalling a formed attitude. Product knowledge is significant, indicating that with the progression of learning, consumers tend to use a formed attitude. Gender and age are not significant, though of the correct sign. The standard deviation of individual's tendency to use semantic processing is 0.109.

Forgetting under Each Mode*:* As we mentioned before, with perfect memory, it is irrelevant to distinguish between semantic and episodic processing, since the information set for both memories is the same. In fact, it is forgetting that acts differently across the two memories that allows us to identify the two systems. For example, the rates of forgetting, such as B^{SE} and B^{EP}, are different. As shown in Table 3, forgetting rate for semantic memory is remarkably smaller than that for episodic memory. This is consistent with the literature that episodic memory is more vulnerable to recall errors. The inclusion of these two memory systems definitely provides more flexibility to parameter estimation; it avoids biased estimators due to under-estimated forgetting size as well. For example, Model II in the detergent category has a significant but almost negligible forgetting effect B^{SE} =3.679E-12. However, once we allow episodic processing, the forgetting effect becomes prominent, $B^{EP} = 81.062$.

5. Conclusions

In this paper, we empirically tested what constitutes a consumer's memory at the point of purchase. A more fundamental question here is whether preferences are endowed or constructed. More specifically, where does the so called intrinsic brand value in consumer choice model, i.e., the intercepts in the utility function, come from? The classical learning model depicts the evolution of consumer's intrinsic brand value along her purchase history in a Bayesian fashion. In the model, the consumer uses her consumption experience to update her prior belief about the brand to the posterior belief. At the purchase occasion, she uses this posterior belief as the brand evaluation. The assumption made in classic learning model is that only overall evaluation (posterior belief) is retrieved, while the usage specific experiences are ignored at the purchase occasion. However, a large literature in semantic and episodic memory shows the existence of the two distinctive memories. Hence, we propose a learning model that is more in line with human behaviour. It allows consumer to recall either general evaluation or certain usage experiences to form an evaluation.

Our findings show that consumers are more likely to use recalled consumption episodes to form a new belief than using an existing belief. Also we find forgetting varies across both memory systems. Episodic memory is more vulnerable to semantic memory. We also empirically tested the effect of demographic variables as well as situational factors on the type of memory being recalled. Since the model includes both memory processes it will definitely be more flexible in fitting the data and have higher predictive power as well. The implication would be for marketers to initiate marketing strategies at purchase occasion to trigger recall of positive episodes and at consumption occasion to make positive usage more salient in memory when they are encoded.

Essay 2

Unpacking Absorptive Capacity under the Context of Knowledge Alliances: A Dynamic Co-evolution Model

Abstract

Absorptive capacity is defined as a firm's capability to recognize the value of external knowledge, assimilate it and apply it to commercial ends. Absorptive capacity is a firm's fundamental learning capability that enables a firm to be adaptively innovative and structurally flexible to external changes. Ever since Cohen and Levinthal's seminal work in 1990, this construct has been used in more than 900 peer-reviewed papers. However, the measurement and operationalization of absorptive capacity has been problematic in empirical research, due to the complexity of the construct itself. Three aspects contribute to the difficulty of measuring absorptive capacity: the intangible process-based nature, the broad multidimensional nature and the dynamic sequential nature of the construct.

To conquer the above issues when operationalizing absorptive capacity, we employ a 3-step structural model so that each dimension of the construct is measured directly and sequential relationships between dimensions are delineated. In addition, our model allows a firm to interact with its knowledge environment, depicting a co-evolution process of both the firm and the environment it is embedded in. This study provides empirical evidence and sheds light on the determinants of each building block of absorptive capacity. Hence, it gives implications to firms on how they can build and strengthen their absorptive capacity.

We find when managing their knowledge base, firms show a pattern of cyclic, rather than simultaneous, ambidexterity that they switch between explorative and exploitative partnerships. We also find that past knowledge enhances firm's assimilation ability in general, but too much past knowledge impedes firm's knowledge assimilation. In terms of utilization efficiency, firms are far more efficient in utilizing capital than using knowledge stock.

KEYWORDS: Absorptive Capacity, Knowledge Alliances, Knowledge Portfolio, Knowledge Production

1. Introduction

In a high-velocity environment due to market volatility or technology changes, some firms manage to survive constantly from such turbulences, whereas others fail to do so. It is believed (Kedia & Bhagat, 1988; Koza & Lewin, 1998) that firms with strong receptivity to external changes also have high absorptive capacity, a fundamental learning capability that enables an organization to be adaptively innovative and structurally flexible. Absorptive capacity refers to "*an organization's ability to recognize the value of new, external information, assimilate it and apply it to commercial ends*", a construct coined by Cohen and Levinthal in their seminal work (Cohen and Levinthal, 1990).

The concept of absorptive capacity has been widely used in more than 900 peer-reviewed papers (Lane, Koka and Pathak, 2006). However, the measurement and operationalization of absorptive capacity have been problematic in empirical research, due to the complexity of the construct itself. Simple proxies, such as R&D intensity (Cohen and Levinthal, 1990; Meeus, Oerlemans, and Hage, 2001), patents (Ahuja and Katila, 2001), co-authorship (Cockburn and Henderson, 1998), firm age (Rao & Drazin, 2002; Sorenson & Stuart, 2000), patent citation overlap (Mowery et al., 1996), etc. have been used to depict absorptive capacity. The problem of measuring absorptive capacity with simple proxies is, firstly, the impacts of absorptive capacity become conjectures. The effects of absorptive capacity are merely hypothesized as they are not directly measured. For example, it is believed that R&D investment can enhance absorptive capacity and absorptive capacity will increase a firm's innovation capability. Hence, firms with high R&D investment are believed to have high absorptive capacity, thus strong innovation capability. However, the impact of absorptive capacity on firms is hypothesized, not tested or measured. Secondly, absorptive capacity is a multi-dimensional construct, which is too rich to be fully captured by a single proxy or measure. The consequence of doing so is it becomes unclear if these proxies are measuring the same aspect of the construct, as questioned by Zahra and George (2002). The questionable appropriateness and validity of these measures also lead to inconsistent conclusions in the literature. For example, Tsai (2001) found that R&D intensity has positive impact on innovation, but Mowery et al. (1996) found R&D intensity is not a good predictor of interorganizational learning. The second problem with using simple proxies is that absorptive capacity is a process, or a routine, not a resource such as R&D investment or number of R&D personnel, which can be easily acquired from the external market. Using measures manifesting the resource aspect of a firm to depict absorptive capacity is an essential betrayal of this concept itself (Wernerfelt, 1984, Lichtenthaler,2009).

To avoid the above issues, researchers (Nichools-Nixon 1993; Lane and Lubatkin, 1998; Szulanski, 1996) have used multiple measures to capture the multidimensional nature of absorptive capacity. For example, Szulanski (1996) uses 9 items to measure the working unit's ability to value, assimilate and apply new practices inside a firm. Lane and Lubatkin (1998) use 5 different measures to separately describe the three dimensions of absorptive capacity in a dyadic relationship. Most of these measures are obtained through surveys of firms. Using multiple measures recognizes the multidimensional aspect of absorptive capability, and is definitely an improvement over simple proxies. It also emphasizes the fact that absorptive capacity is a capability rather than a resource. Nevertheless, such operationalization still disregards two important aspects of the construct. The first aspect is *the sequential relationship between the three dimensions and the second is the cumulative and path-dependent nature of absorptive capacity***.** Both features need to be adequately captured and accommodated in a modeling framework to fully understand the power and impact of absorptive capacity on firm's adaptability and performance in environmental turbulence.

There is a sequential relationship between the three dimensions of absorptive capacity, namely “to *recognize*, *assimilate* and *apply* new knowledge”, and the order of the sequence cannot be changed. One dimension must be built upon the previous dimension. To be in detail, the type or the characteristic of new knowledge recognized or sourced in the first stage will determine the efficiency and efficacy of knowledge assimilation in the second stage. The knowledge assimilated in the second stage determines the success of the knowledge application or transformation in the third stage. In the reconceptualization of absorptive capacity, Zahra and George (2002) suggest to combine knowledge identification and assimilation as potential absorptive capacity and knowledge application and exploitation as realized absorptive capacity. This reconceptualization implicitly recognizes such sequential relationships between the dimensions. However, the aforementioned multi-measures treat these dimensions as independent that enhancing the capability of one dimension has no impact on other dimensions.

***Cumulated and path-dependent absorptive capacity* -** In their original work, Cohen and Levinthal (1990) mentioned that absorptive capacity associated with a particular product or production process is firm specific. Therefore, it cannot be purchased immediately from the market. Rather, there exist considerable time lags before certain level of absorptive capacity can be reached. This reveals the evolutionary and cumulative nature of absorptive capacity. The buildup of absorptive capacity therefore cannot be depicted with a single-wave cross-sectional data, unless longitudinal survey data is obtained. Hence, using multiple items to measure absorptive capacity through a single-wave cross-sectional data only reflects the state of a firm’s absorptive capacity at a particular time. It does not and cannot explain the mechanism whereby such levels are built up.

In this study, we delineate the three dimensions of absorptive capacity under the context of knowledge alliances. Absorptive capacity is about internalization of external knowledge. In the past decade, due to increasing competition intensity, the short time from identification of a problem to its arrival does not allow a firm to internally develop the capability to solve the problem. Instead, the majority of such firms form a learning-alliance to speed up solution search or capability development. An extensive literature (Inkpen & Dinue, 1998, Larsson et al, 1998, Simonin, 1999, Schildt et al, 2012) has discussed how a firm uses alliance to form its knowledge base and get access to technology-based capabilities. Knowledge collaboration has become one of the most widely cited motives for firms to form an alliance (Mowery et al, 1996, Grant & Baden-Fuller, 2004, Lavie & Rosenkopf, 2006).

Hence, our research objective is to *propose a holistic model that will depict absorptive capacity in a dynamic and co-evolutionary fashion by describing each of its dimensions separately, yet recognizing the relational sequence between these dimensions* under the context of knowledge alliance**.** We use a 3-stage structural econometric model to describe in full details the co-evolution of a firm's absorptive capacity as well as the knowledge environment where the firm resides. In doing so, not only do we restore the original scope of view this construct attempts to express, but also allow the usage of widely available alliance data, considering the difficulty of obtaining longitudinal survey data. We explicitly model the knowledge screening process, which is a firm's strategic behavior but missing in the literature. Conditioned on the wide recognition of the importance of absorptive capacity in the literature, we provide a framework that enables us to test empirically what the antecedents of absorptive capacity are. Hence, it answers the question how a firm's absorptive capacity is built up from period to period. This is one of our main contributions.

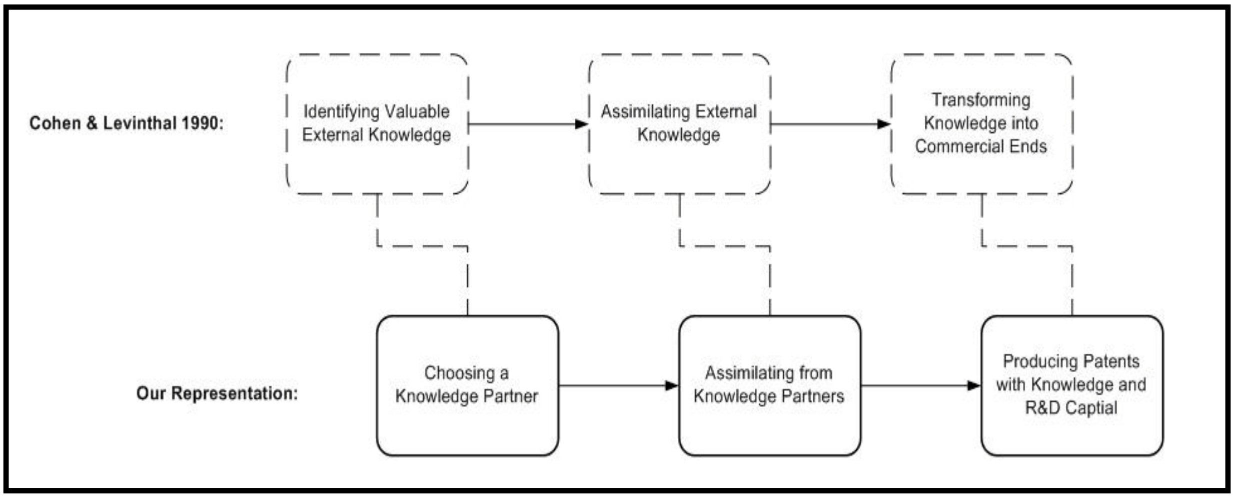
We find when choosing a knowledge partner, firms are more likely to build an explorative partnership versus an exploitative partnership. Instead of following Hannan and Freeman's (1984) argument of structural inertia, firms in pharmaceutical industry are better described as with ambidexterity proposition. However, we find that firms are only inter-temporally, more precisely, cyclically ambidexter that they maintain the balance between explorative and exploitative partnerships from period to period. We also find that knowledge similarity shows a bi-polar pattern that partner firms with either very different or very similar knowledge structure are more likely to be chosen. This forms our second main contribution.

We empirically test the effect of past knowledge on knowledge assimilation. In line with Cohen and Levinthal's (1990) conjecture, we find that the stock of past knowledge enhances knowledge assimilation. However, too much knowledge stock impedes knowledge assimilation. Similar pattern is also found for knowledge similarity. For knowledge transformation, we find that firms are less efficient in knowledge utilization than in capital utilization. Further, we find that the effort needed to produce an extra patent in a year drops quickly and levels off at 7 or more patents.

2. Conceptual Framework

In this part, we discuss how we unpack absorptive capacity into three sequential dimensions in the setting of knowledge alliances. Figure 6 presents our conceptual framework and shows how the original definition of absorptive capacity is represented under the context of knowledge alliances. In our model, identifying new valuable external knowledge, the first dimension of Cohen and Levinthal's original definition, is translated into choosing a knowledge partner. The second dimension, assimilating external knowledge, is reflected as assimilating knowledge from the knowledge partners. The third dimension, applying knowledge to commercial ends, is modeled as transforming knowledge into innovative knowledge product, namely patents here. Next, we shall substantiate the appropriateness of the correspondence between Cohen and Lavinthal's definition and our representation under the setting of knowledge alliances.

Figure 6: Conceptual Framework



2.1 From Identifying New Knowledge to Choosing a Knowledge Partner

Identification of new and valuable external knowledge is the first dimension of absorptive capacity. Cohen and Levinthal (1990) believe such capability allows a firm to accurately predict the nature and commercial potential of technological advances in certain industry. To foster such capability, they emphasize the importance of knowledge structure possessed by a firm. They use the metaphor of individual learning to illustrate that cumulated prior knowledge helps a firm to assimilate new knowledge through better understanding and association. Technological knowledge is usually tacit knowledge that cannot be articulated as descriptive manual or readily obtained from market. Instead, it is cumulated through considerable organizational learning effort. Accumulation of such related prior knowledge allows a firm to detect and evaluate the intermediate technology advances that hint the merits of new technological development. From this perspective, the volume of cumulated prior knowledge matters. In addition, to increase the prospect of such technological opportunities, a firm should also have fairish level of knowledge diversity so that the incoming knowledge can be related to what is already known. Hence, knowledge depth sets the sensitivity and knowledge diversity sets the range of a firm's radar system for knowledge detection. It is these two attributes that form the structure of a firm's knowledge base, which determines a firm's likelihood to extract knowledge with promising potential from rapidly advancing environment.

If as aforementioned, knowledge alliances become the major form of organizational learning, then the process of choosing a knowledge partner is a process how a firm shapes its knowledge base. The firm injects into itself with new knowledge by establishing a link with a partner who carries such knowledge. Consequently, a firm attempts to sculpture a knowledge base that is structurally responsive to knowledge identification through managing the portfolio of its knowledge partners. Therefore, if choice of knowledge partners shapes a firm's knowledge base and if structure of a firm's knowledge base determines its performance in assessing new knowledge, then a firm's capability to identify new valuable knowledge can be represented by its decision quality in choosing knowledge partners.

2. 2 From Assimilating Knowledge to Assimilating from Knowledge Partners

The second dimension of absorptive capacity is knowledge assimilation. Knowledge assimilation is a process where external knowledge is internalized, embedded and becomes active in an organization's memory system. Simply put, knowledge assimilation transforms "know-what" to "know-why", which allows a firm to learn the intricate web of relationship underlying various technological or non-technological phenomena. Under the context of knowledge alliance, assimilation of external knowledge is equivalent to interorgnizational learning, where knowledge transfers from the knowledge partner to the focal firm. We argue that two factors determine a firm's knowledge assimilation efficacy: the firm's assimilation capability and knowledge pool where the firm resides. Assimilation capability focuses on the firm's internal factors, whereas knowledge pool emphasizes how external environment can impact the firm's learning efficacy. A metaphor will help to elaborate these two factors more lucidly. The assimilation capability indicates whether a firm is a stone or a sponge, and a sponge absorbs much better than a stone. However, the external environment also matters as a sponge absorbs much more in a sea than what it would in a desert. In our model, the alliance portfolio a firm has in certain period constitutes the knowledge pool in which a firm resides. In each period, the maximum a firm could assimilate is the size of its knowledge pool. The size of this knowledge pool is nothing but the total of the knowledge carried by each member in that pool. The actual knowledge assimilated depends on the firm's assimilation capability, which is normalized to be a portion ranging between [0, 1].

In this section, we illustrate how we will model a firm's capability of knowledge assimilation. We take into account not only a firm's internal factors but also environmental factors. By doing so, we are able to track a firm's dynamic learning behavior, just as Cohen and Levinthal (1990) mentioned in their original paper that absorptive capacity is cumulated, path-dependent. In addition, we are also able to capture how a firm and its environment evolve through interacting with each other. Knowledge transfer is bidirectional, when the focal firm is absorbing knowledge from its partners, the partners also absorb from the focal firm as well as firms in other networks, should they get involved in more than one network. The evolution of each member in a network composes the evolution of the network. Therefore, the process we model is a co-evolution of the focal firm and its knowledge environment.

2.3 From Applying Knowledge to Producing Patents

After recognizing and assimilating new knowledge, the third dimension of absorptive capacity is to apply the new knowledge to commercial ends, a process from “know-why” to “know-how”. Applying new knowledge to commercial ends is nothing but a production process that transforms knowledge inputs and other inputs to outputs of innovative products. Hence, a natural way to measure a firm’s efficiency or capability in utilizing absorbed knowledge is to employ a knowledge production function. The concept of knowledge production function is initially mentioned by Grilliches (1979). Unlike production function for conventional goods where labor and capital are two key production input factors, the production of innovative products is characterized as not labor intensive but technology or knowledge intensive. Hence, knowledge becomes one of the most important production inputs. The challenge of a knowledge production function is how to measure the knowledge possessed by a firm, which is intangible. Different proxies have been used to measure it, such as number of R&D personnel, total investment in innovation, etc. The first and second stages of our model allow endogenous generation of knowledge stock of a firm through the learning process.

Although the third dimension of absorptive capacity is about producing and commercializing innovative products to the market, we use the number of patents produced by the focal firm as the output indicator. Is it appropriate to use patent production to represent a firm’s capability in knowledge utilization? Logically speaking, a patent is a knowledge invention. Patent is a realization of R&D effort when the technical requirements of an idea are successfully fulfilled. Hence, patent itself is a sufficient reflection of a firm’s capability in knowledge utilization. Patent does not necessarily lead to innovation and not all innovations are patentable. However, according to an extensive survey across 19 different industries conducted by Arundel and Kabla (2001), 79.2% of innovations are patented in the pharmaceutical industry. Hence, the majority of the innovation effort could be captured by patents produced. Numerous researches also find a positive relationship between the number of patents possessed by a firm and corporate performance in terms of profit and sales (Schere, 1965, Griliches et al. 1991, Ernst, 1995, 2001). Therefore, patent as the knowledge invention and the intermediate product that leads to innovative products is sufficient to represent a firm's knowledge utilization capability.

3. Econometric Model

In this part, we elaborate how each dimension of absorptive capacity is modeled. In Section 3.1, we discuss factors that affect a focal firm's choice of a partner in a particular period, a decision that determines the composition of alliance portfolio in the next period. In Section 3.2, we discuss, based on the choice decision made in the first stage, how a focal firm assimilates knowledge from the partners in its alliance portfolio. In the last section, we describe how both knowledge capital and R&D investments as production inputs are transformed into innovative outputs, such as the number of patents.

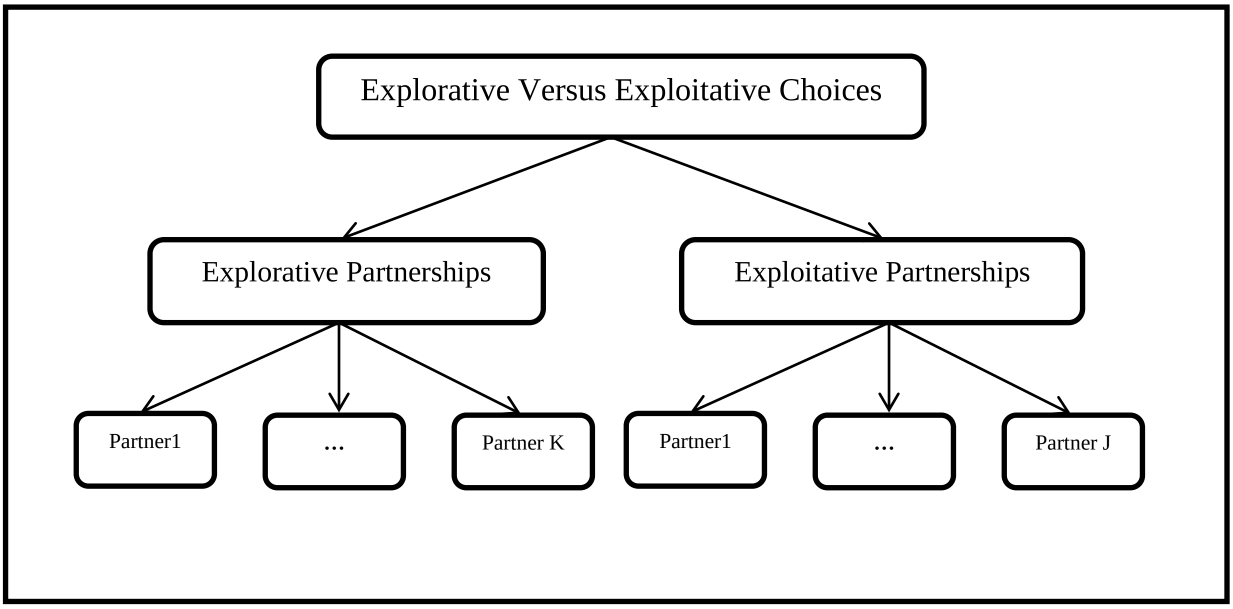
3.1 Choice of Knowledge Partner

As aforementioned, the focal firm builds its knowledge base through alliancing with knowledge partners. When choosing a particular firm as its knowledge partner, the first decision a firm should make is how the acquired new knowledge will shape the composition of its technology portfolio. Let $T_{i,t} = \{T_{i,1,t}, T_{i,2,t}, \dots, T_{i,N,t}\}$[5], an *N*- dimension row vector be the technology portfolio possessed by firm i in period t, where $T_{i,n,t}$ stands for the cumulative

[5] Please see Table8 for definition of notations.

number of alliances that has been made till period t based on technology n. The firm's decision then becomes a choice between extending the width of its technology portfolio by exploring new technology areas or increasing the depth of the technology portfolio by exploiting currently engaged technology areas. As in March (1991), exploitation is about refinement and efficiency, whereas exploration is about search, variation, and discovery. In this study, we define an explorative partnership as an alliance built on technology areas that a firm has never been engaged in. Similarly, an exploitative partnership for alliances builds on a firm's currently engaged technological areas. Exploration and exploitation are two fundamentally different approaches of organizational learning. Explorative partnership can form a technology portfolio with large diversity that outruns environmental selection process. However, explorative partnership is also associated with high risk, and failed explorative alliances waste a firm's scarce resources and undermine its learning efficiency. Exploitative partnership allows a firm to build closely on existing knowledge base and improve short-term efficiency. However, a too narrow technology base weakens a firm's adaptive flexibility and receptivity to changes. Essentially, the firm is to achieve a balance between the width and the depth of its technology portfolio. Hence, when choosing a partner, the focal firm i will first decide whether it will explore new technology areas or exploit current technology areas based on its expectation of future composition of the technology portfolio after obtaining such a knowledge partner. Then conditioned on this decision, it will choose the most suitable partner among all partners who possess the technology desired by the focal firm. Please see the decision tree in Figure 7 for the choice process of a particular focal firm.

Figure 7: Decision Tree of Partner Choice



We use a nested multinomial logit model (Train, 2009) to describe the above decision process, where the dependent variable is focal firm i's choice of a particular partner. Let U_{ijt} be the utility firm *i* will obtain should firm *j* be chosen as a partner in period *t*.

$$\mathrm{U_{i,j,t}} = \mathrm{W_{i,k,t}} + \mathrm{V_{k,j,t}} + \varepsilon_{\mathrm{i,j,t}} \qquad (1)$$

As shown in Equation (1), such utility is decomposed into three parts, where $\mathrm{W_{ikt}}$ is utility contributed by choosing nest *k* (*k* = 1, 2). There are two nests here, 'Explorative' nest which contains partners with new technologies that are not possessed by focal firm i and 'Exploitative' nest which contains partners with technologies that firm i is currently engaged in. The attributes under $\mathrm{W_{i,k,t}}$ are different across nests but same across partners within the same nest. $V_{\mathrm{k,j,t}}$ is additional utility obtained from partner j, which has not been included in the nest utility, and $\varepsilon_{\mathrm{i,j,t}}$ is idiosyncratic utility that is not observed by the econometrician, and is assumed to be independently and identically type-I extreme value distributed.

$W_{i,k,t} = \{\alpha_k, DOC_{i,k,t}, DOC^2_{i,k,t}\}$ is a vector that contains nest specific intercept α_k and $DOC_{i,k,t}$. $DOC_{i,k,t}$ is a variable that measures degree of concentration of firm *i*'s technology portfolio, should firm i choose any firm from Nest *k* to form a knowledge alliance based on certain technology area. $DOC_{i,k,t}$ is constructed as

$$DOC_{i,k,t} = \frac{\sum_{n=1}^{N} T^2_{i,n,t}}{(\sum_{n=1}^{N} T_{i,n,t})^2} \ldots \ldots (2)$$

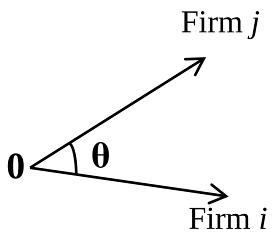
As aforementioned, $T_{i,n,t}$ is the cumulative number of alliances that have been made based on technology *n* till period *t*. A firm's degree of concentration ranges between $[\frac{1}{\sum_{n=1}^{N} T_{i,n,t}}, 1]$, and the higher the degree of concentration, the less diversified a firm's technology portfolio is. Degree of concentration is a commonly used measure that takes into account both the width and the length of a portfolio, where length refers to the number of alliances under each technology area. When concentration is low, a firm is more likely to pursue partners that lead to telescopic technology portfolio. However, when concentration is too high, a firm is less likely to pursue such partners. To test such ambidexterity hypothesis (Eisenhardt and Martin, 2000, He and Wong, 2004), we use a quadratic functional form of degree of concentration in the model and expect an inverted U shape. Since $DOC_{i,k,t}$ measures the degree of concentration, thus $FOV_{i,k,t} = 1 - DOC_{i,k,t}$ measures technology diversity. We give a vivid name to such diversity measure as Field Of Vision.

$V_{k,j,t} = \{TS_{i,j,t}, TS^2_{i,j,t}, CP_{j,t}, X_{j,t}\}$ is a vector that contains partner specific attributes, where $TS_{i,j,t}$ is the technology similarity between focal firm i's alliances' technology portfolio and firm *j*'s technology portfolio. We use cosine similarity to measure the technology similarity as follows

$$TS_{i,j,t} = \cos\theta = \frac{P_{i,t}{}' \cdot P_{j,t}}{\|P_{i,t}\|\|P_{j,t}\|} \ldots\ldots (3)$$

Figure 8: Technology Similarity

	P1	P2	P3	P4	P5	…
Firm *i*	0	1	0	1	1	…
Firm *j*	1	0	0	0	1	…
…	…	…	…	…	…	…



where $\mathbf{P_{i,t}}$ is firm *i*'s alliance portfolio's technology vector, whose dimension is the number of patent classes that can be found in the dataset. The element takes value 1, if firm *i*'s alliances have at least one patent that belongs to the patent class, and takes value 0 otherwise (See Figure 8 for a graphical representation). Since $P_{i,t}$ is always in the first quadrant, $\cos\theta \in (0,1]$. The smaller the cosine value is, the larger the angle between the two firms' technology vectors, and hence, the larger the difference is between the two firms' knowledge structures. When $\cos\theta = 1$, it means the angle between the two firms' technology vectors is 0 degree. Instead of using technology vector of focal firm *i*, we use technology vector of focal firm *i*'s alliance portfolio, which counts the patent classes possessed not only by focal firm *i*, but also all of its alliances. The reason is that when choosing a particular firm as a partner, the focal firm *i* needs to scrutinize if such knowledge is obtainable from its current alliance portfolio. If yes, there is no need to set up a new alliance that creates knowledge redundancy.

The relationship between technology similarity and the probability of firm *j* being chosen is also likely to be an inverted U shape. The rationale is that it is easier for both parties to learn and communicate if they have certain degree of common background or understanding. However, when such similarity is too much, firm *j* is less likely to be chosen due to knowledge redundancy. When such similarity is too little, firm *j* is less likely to be chosen either because of the high risk and high learning costs associated with too different technologies. We also use $CP_{j,t}$ cumulative number of patents belonging to firm *j* till period *t*, to approximate firm *j*'s track of quality in knowledge creation. We expect the larger the cumulative number of patents possessed by firm *j*, the more likely it will be chosen. In addition, $X_{j,t}$ is a vector of control variables such as the size of firm *j* as measured by its revenue.

Hence, the probability that firm j is chosen as the partner is the probability nest *k* is chosen ($P_{i,k,t}$) times the probability that firm j is chosen conditional on that nest k is chosen ($P_{j|k,t}$), as shown in equation (4).

$$P_{i,j,t} = P_{j|k,t} \cdot P_{i,k,t} \dots \dots (4)$$

where the probability that nest *k* is chosen is

$$P_{i,k,t} = \frac{\exp\,(W_{i,k,t} + \lambda_k I_{i,k,t})}{\sum_{m=1}^{K} \exp\,(W_{i,m,t} + \lambda_l I_{i,m,t})} \dots \dots (5)$$

and conditional on a chosen nest *k*, the probability that firm *j* will be chosen is

$$P_{j|k,t} = \frac{\exp\left(\frac{V_{k,j,t}}{\lambda_k}\right)}{\sum_{j\in k} \exp\left(\frac{V_{k,j,t}}{\lambda_k}\right)} \dots \dots (6)$$

The inclusive value of choosing nest *k* is defined as

$$I_{i,k,t} = \ln \sum_{j \in k} \exp\left(V_{k,j,t}/\lambda_k\right) \ldots\ldots (\mathbf{7})$$

3.2 Assimilation of Knowledge

In each period, by choosing the partners, the focal firm forms a pool of knowledge where it is embedded. Knowledge exchange takes place along the dyadic link between focal firm i and its partner *j*. In this section, we describe how knowledge gets accumulated through the dual evolution of knowledge pools belonging to both the focal firm and the partner firms.

Let i_t be the set of partners that ally with firm i (note: here i could be either a focal firm or a non-focal firm) in period t. $K_{i,t}$ is the knowledge stock possessed by firm i at period t, which is formulated as the sum of firm i's last period knowledge $K_{i,t-1}$ plus the new knowledge $k_{i,t}$ acquired in period *t* as shown in Equation (8). Let f be the rate of forgetting which captures knowledge obsolescence. Hence, in Equation (8), the stock variable is $K_{i,t-1}$ and the flow variable is $k_{i,t}$.

$$K_{i,t} = (1 - f) \cdot K_{i,t-1} + k_{i,t} \ldots\ldots (8)$$

The knowledge flow $k_{i,t}$ is modelled as

$$k_{i,t} = \sum_{j \in i_t} AC_{i,j,t} \cdot K_{j,t-1} \ldots\ldots (9)$$

$K_{j,t-1}$ is the knowledge stock of firm *j* in the last period (at the beginning of period *t* before acquiring any new knowledge), where $j \in i_t$ is one of firm *i*'s partners at period *t*. $AC_{i,j,t}$ is firm *i*'s assimilation capability when learning from partner *j*. $AC_{i,j,t}$ ranges between [0, 1], where 0 means no learning at all and 1 means full assimilation. Firm *i* cannot learn more than what is available from firm j . Hence, $K_{j,t-1}$ measures the maximum that firm *j*can offer to firm *i* and $AC_{i,j,t}$ measures how efficient firm $\boldsymbol{i}$ is in terms of learning from firm j. Thus, the new knowledge added to firm $\boldsymbol{i}$ is simply a summation of knowledge that is learnt from all members in firm *i*'s alliance portfolio.

Since $\mathrm{K_{j,t-1}}$ can be generated recursively once an initial value is set, we only need to specify the assimilation capability $\mathrm{AC_{i,j,t}}$ in Equation (10). As $\mathrm{AC_{i,j,t}}$ ranges between 0 and 1, hence, it can be modelled as

$$\mathrm{AC_{i,j,t}} = \frac{\exp\,(\delta + \mathrm{K_{i,t-1}} + (\mathrm{K_{i,t-1}})^2 + TS_{j,t}^{i} + (TS_{j,t}^{i})^2 + \mathrm{PC_{i,j,t}})}{1 + \exp\,(\delta + \mathrm{K_{i,t-1}} + (\mathrm{K_{i,t-1}})^2 + TS_{j,t}^{i} + (TS_{j,t}^{i})^2 + \mathrm{PC_{i,j,t}})} \ldots\ldots (10)$$

$K_{j,t-1}$ is a firm $\boldsymbol{i}$'s prior knowledge. As mentioned in Cohen and Leventhal (1990), prior knowledge facilitates organizational learning. However, too much prior knowledge will also create organizational inertia, thus hindering the assimilation of new knowledge (Hannan and Freeman, 1984). We put a quadratic term to capture the decreasing marginal effect of prior knowledge on learning.

$TS_{j,t}^{i}$ is the technology similarity between firm i and firm j. It is defined in a similar fashion as $\mathrm{TS_{i,j,t}}$ in equation (3). The only difference is that $TS_{j,t}^{i}$ measures similarity between firm i and firm j, instead of technology similarity between <u>firm i's alliance portfolio</u> and firm j as $\mathrm{TS_{i,j,t}}$ does. Since learning happens along the dyadic line between firm i and firm j, we directly measure the technology similarity between firm i and firm j. Again, we propose that certain degree of common knowledge background can facilitate learning and increase learning efficiency as found by Lane and Lubatkin (1998). However, we also believe that too much similarity in technology will hinder learning. $PC_{i,j,t}$ is a dummy variable indicating whether firm i and firm j have ever cooperated by period t. We expect that past cooperation will contribute positively to assimilation capability due to established link in past time as well as the familiarity with each other's routines.

3.3 Production of Innovative Products

Production of innovative products, like any other production process, is nothing but a process that transforms inputs into outputs. The difference is when producing a knowledge intensive product, such as patents, knowledge capital and R&D investment become the two most important factors.

We use Cobb-Douglas production function to measure the focal firm's efficiency in transforming knowledge into innovative products.

$$P(K,C)_{i,t} = A \cdot K_{i,t}^{\alpha} \cdot C_{i,t}^{\beta} \dots\dots (11)$$

where A is the total productivity factor, which measures the effect on total output that is not caused by the measured inputs. $K_{i,t}$ stands for knowledge capital of firm i at period t and $C_{i,t}$ represents capital investment for R&D. α and β are output elasticity for knowledge and capital inputs respectively. They measure a firm's efficiency in transforming the inputs into innovation. To linearize the equation

$$lnP(K_{i,t}, C_{i,t}) = lnA + \alpha \cdot lnK_{i,t} + \beta \cdot lnC_{i,t} + \varepsilon_{i,t} \dots\dots (12)$$

Cobb-Douglas production function is a continuous function, where as long as inputs are non-zero, outputs will always be positive. However, the continuous production process is not visible to the econometrician. Innovation usually comes in the form of finished products or registered patents, namely in a discrete form. For example, every experiment or research discussion contributes to the creation of the innovation. However, such progress is unobservable. Let $P(K_{i,t}, K_{i,t})$ be the continuous intermediate product from the inputs to an innovation, which is latent in the model. The intermediate product becomes an innovation once it reaches certain minimum quality threshold. Hence,

$$q_{i,t} = 0 \quad \text{if } \theta_0 < lnP(K_{i,t}, C_{i,t}) \leq \theta_1$$

$$q_{i,t} = m \quad \text{if } \theta_{m-1} < lnP(K_{i,t}, C_{i,t}) \leq \theta_m \qquad \text{for m=2,…., M-1}$$

$$q_{i,t} = M \quad \text{if } \theta_{M-1} < lnP\left(K_{i,t}, C_{i,t}\right) \leq \theta_M \ldots\ldots (13)$$

where $q_{i,t}$ is the number of innovations that firm *i* produces in period *t*. $\theta_{0,}\ \theta_{1,\ldots,}\ \theta_M$ are the minimum quality thresholds to produce different number of the innovative products. Therefore, the probability for firm *i* to have *m* innovations in period *t* can be expressed as

$$\Pr\left[q_{i,t} = m\right] = \Pr\left[\theta_{m-1} < lnP\left(K_{i,t}, C_{i,t}\right) \leq \theta_m\right]$$

$$= F\left[\theta_m - \left(lnA + \alpha \cdot lnK_{i,t} + \beta \cdot lnC_{i,t}\right)\right]$$
$$- F\left[\theta_{m-1} - \left(lnA + \alpha \cdot lnK_{i,t} + \beta \cdot lnC_{i,t}\right)\right] \ldots\ldots (14)$$

Where F(·) is the cumulative density function of $\varepsilon_{i,t}$. Since θ_m -*lnA* cannot be identified separately, we therefore normalize $\boldsymbol{-lnA}$ to 0.

$$\Pr[q_{i,t} = 0] = F[\theta_1 - (\alpha \cdot lnK_{i,t} + \beta \cdot lnC_{i,t})] \ldots\ldots (15)$$

$$\Pr[q_{i,t} = M] = 1 - F[\theta_{M-1} - (\alpha \cdot ln K_{i,t} + \beta \cdot ln C_{i,t})] \ldots\ldots (16)$$

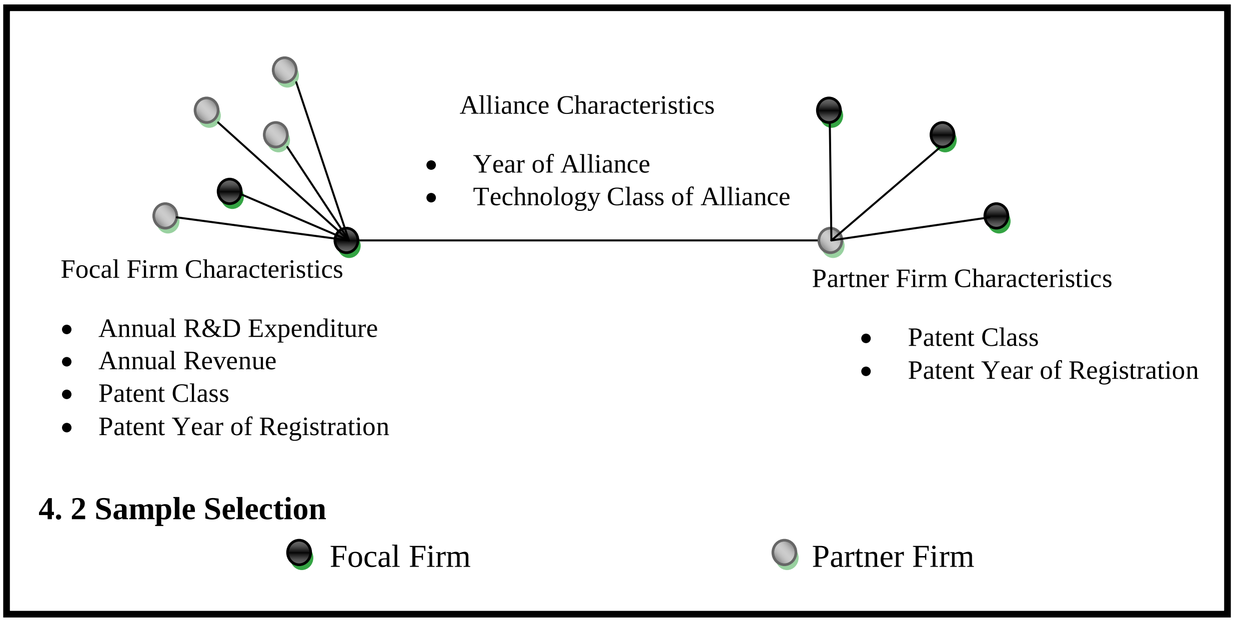
Apparently, the distribution of $\varepsilon_{i,t}$ determines the cumulative density function in the above functions. Also, from Equations (15) and (16) we know that θ_0 and θ_M are not identified. By building a production function in such fashion, we are able to address the issue whether the firm is efficient in transforming such knowledge into final commercial ends.

4. Data

4.1 Data Structure

The structure of the data that we use to calibrate our model is shown in Figure 9. The black dots are focal firms, which are our research objects, whereas the gray dots represent partner firms, which are the knowledge partners of the focal firms. The difference between a focal firm and a partner firm is that we capture the whole network topology of a focal firm in the observation span. However, we do not capture the network topology of a partner firm. In this sense, a partner firm will only enter the dataset if it allies with a focal firm, thus all partner firms' knowledge partners have to be focal firms. Nevertheless, the knowledge partner of a focal firm can be either a focal firm or a partner firm. In the dataset, we have attributes that describe focal firm's characteristics, such as the focal firm's annual R&D expenditure, annual revenue, year of patent registration, and patent class. For partner firms, we have the patent data to describe their productivity and areas of expertise. In addition, we have information that describes the dyadic relationship between the two firms, such as the year that an alliance is formed and the technology area that particular alliance intends to develop. Such information allows us to delineate how a focal firm's alliance portfolio evolves from time to time.

Figure 9: Data Structure



4. 2 Sample Selection

To draw our sample we first listed all firms available in OSIRIS directory with the following NAIC codes: 325412 (pharmaceutical preparation manufacturing), 325413 (in-vitro diagnostic substance manufacturing), and 325414 (biological product (except for diagnostic) manufacturing). Next, we retained only the listed and delisted firms that were incorporated before year 2000. Then we omitted the firms whose sales and employee records were missing in Compustat database for the first half of our study time period. These steps resulted in 519 firms. Data pertaining to alliances were collected from the Recombinant Capital (Recap) database, which provides a comprehensive listing of biotechnology companies, their alliances, their valuations, and information about clinical trials. Empirical studies frequently use alliance data from Recap, and existing research has demonstrated that Recap data are comparable in quality to those obtained from BioScan (Cowan and Jonard, 2009; Gopalakrishnan, Scillitoe, and Santoro, 2008). After checking the alliance records of the 519 firms between 1990 and 2000 in Recap database, we removed the ones whose alliance record was missing in Recap. By this way we cross validated the firm names with Recap database as well. Our final sample consists of 331 firms. Further, we removed firms that have no annual R&D expenditure information throughout the dataset. Since we are studying the evolution of firm's alliance portfolio, we retain firms with at least two alliances. This leaves us with 107 focal firms in total.

We next obtained the patents issued to these firms between 1990 and 2000 by the U.S. Patent and Trademark Office (USPTO) from the Patsnap database (www.patsnapglobal.com). The U.S. patent classification system consists of more than 100,000 patent subclasses, aggregated into approximately 400 three-digit patent classes. We used the three-digit patent classes to identify patents within the classes relevant to biopharmaceuticals, chosen according to the USPTO Technology Profile Reports and patent descriptions found in Lim (2004)[6]. Thus, we tested our hypotheses on a sample of 107 biotechnology firms between 1990 and 2000 that produced 5,567 patents and established 1,841 alliances during the study period.

4.3 Descriptive Statistics

For the included focal firms, on an average they have 15.093 (SD = 14.227) alliances in the study span. We have also 143 partner firms with an average of 6.788 (SD = 4.323) alliances. From this perspective, we capture a more complete network of the focal firms than the partner firms, as is expected. Each focal firm has 13.066 (SD = 11.715) different partners contributing to 15.093 alliances. That means repetitive cooperation is not uncommon when choosing a knowledge partner. For partner firms, they have 5.904 (3.480) different partners (either focal or partner firms) contributing to 6.788 times of alliance.

[6] The classes chosen are 424, 435, 436, 506, 514, 530, 536 and 585. Please refer to http://www.uspto.gov/patents/resources/classification/classescombined.pdf for the description.

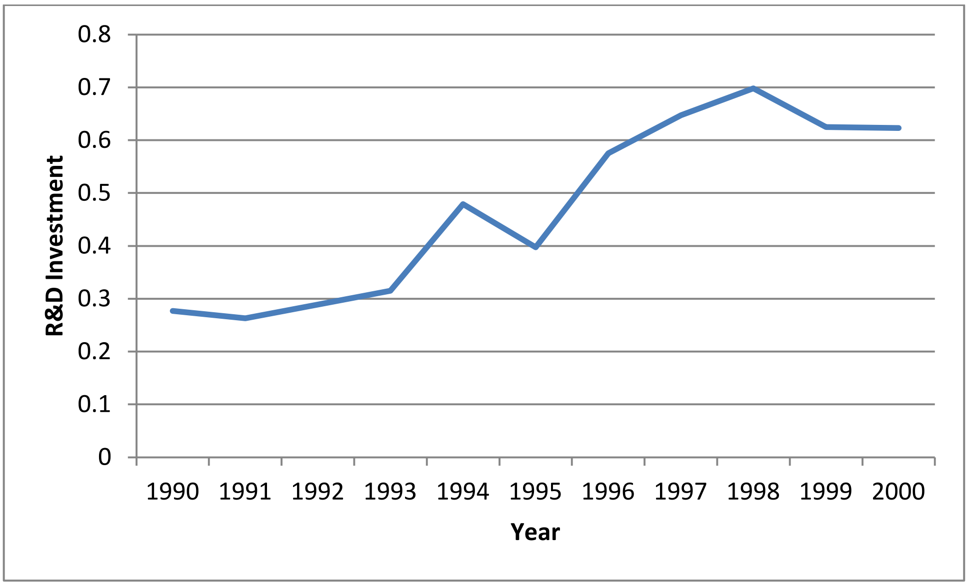
We use the annually registered patents under each firm to measure the productivity of either a focal firm or a partner firm. The number of patents that are produced by focal firms in each year is 5.051 (SD = 4.762) on an average, and the corresponding number is 4.909 (SD = 4.154) for the partner firms. The difference between the focal and partner firms is not statistically significant with a p-value 0.396.

Only the focal firm's annual R&D annual expenditure is collected with a mean of $0.644 million (SD = $0.875 million) across the study span. The annual inflation adjusted R&D expenditure marks an increase from $0.277 million in 1990 to $0.623 million in 2000. Table 2 and Figure 10 exhibit the inflation adjusted annual R&D investment, illustrating how the firm's annual R&D expenditure evolves.

Table 5: Annual R&D Expenditure by Focal Firms ($million)

Year	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
Annual R&D Expenditure	0.277	0.274	0.310	0.348	0.543	0.463	0.690	0.772	0.871	0.797	0.822
Inflation Rate	1	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.03
Inflation Adjusted Annual R&D Expenditure	0.277	0.263	0.289	0.315	0.479	0.397	0.575	0.647	0.698	0.625	0.623
Rate of Increase	0	-0.051	0.099	0.090	0.520	-0.171	0.450	0.125	0.078	-0.105	-0.003

Figure 10 Mean Annual Inflation-adjusted R&D Expenditure, 1990-2000 ($ millions)



5. Variable Operationalization and Estimation

As discussed in the data section, in the alliance dataset, we only observe the knowledge partners that were chosen, not the candidate partners that had been actually considered by a focal firm. Namely, we do not observe the choice set from which the final choice is made. Since knowledge partner selection is one important aspect that we aim to capture in our model, we could not investigate this decision process without defining a proper choice set.

We use a nested multinomial logit model to describe such process, where the firm first decides its engagement between old versus new technological areas (choice of a nest) and then decides which firm to form a knowledge partnership with (choice of a firm) conditioned on the choice of technological area. In the alliance dataset, we have information on the technological area that a particular knowledge partnership is formed for. Hence, the history of a firm's knowledge alliance also illustrates the evolution of its technology portfolio. With an alliance made by focal firm i in year t, we observe the corresponding technological area that the alliance is based on. For exposition purpose, we denote such technological area as T_N and the technology set possessed by firm i till year t as T_t^i . Should this technological area be new ($T_N \notin T_t^i$), then any other firms that have ever engaged in technology T_N by year t form the "explorative" nest for this choice occasion. However, we do not observe the technology should a firm decide to be exploitative for the same alliance occasion. We assume that 1) In the following alliance history of firm i, if we are able to find an alliance that is based on technology $T_O \in T_t^i$, then the first such technology T_O is the technology on which firm i form the "exploitive" nest at year t. 2) If in the following alliance history of firm i, no such technology T_O can be found then the exploitative nest is an empty set at year t. We believe it is appropriate to form a nest based on such assumption. If T_O and T_N are indeed two competing technology areas considered by firm i in year t and T_O as the most likely chosen technology from T_t^i, firms should have the urgency to form the alliance in near future. However, if none of the technology area that belongs to T_t^i has ever shown in the rest history of firm i, then we could conjecture that exploiting the current technology portfolio T_t^i is not a priority, thus not considered. The same logic applies, if the alliance is based on an old technology $T_O \in T_t^i$. Then the exploitative nest is formed by firms that have engaged in T_O by year t. Similarly, the explorative nest is formed on the next new technology area T_N in the following history after year t. Should there not exist such new technology, then the explorative nest is an empty set.

Equations (1) and (16) form our final models of estimation. To calibrate the model, we use maximum likelihood method to first estimate out the nested multinomial logit model (Equation 4) based on the alliance choice data. We then estimate the rank ordered logit model (Equation 14) with the patent and the R&D annual expenditure data set.

6. Results

6.1 Choice of Knowledge Partners

As shown in Table 6, the constant α_1 for explorative nest is 12.958 (standard error or SE = 2.062), indicating that firms under study are in general quite explorative that they prefer explorative partnership to exploitative partnership. Knowledge alliance outside of the current technology field is more likely to be formed. As expected, field of vision has an inverted U shape relationship with the chance that a firm chooses to be explorative. This means that with the increase of technology diversity, a firm is more likely to explore new technology area. However, when the technology portfolio becomes too diverse, then the firm will have lower tendency to engage in the new areas. Similarly, degree of concentration also has an inverted U-shape as expected. This means that when the degree of concentration is low, a firm will form alliances that make its technology portfolio more focused. However, when the portfolio turns to be too narrow, then the firm is less likely to stay in the same area. Our results show that despite structure inertia and lock in effect, an ambidexterity proposal may describe firms' behaviors better.

When choosing a particular knowledge partner conditioned on a selected technology area, the focal firm will assess the technological fit between the candidate firm and itself. It chooses a firm that has the right level of technology similarity aligned with the focal firm alliance technology portfolio. The focal firm will not bring in a knowledge partner to his alliance portfolio with complete technology overlap, which leads to knowledge redundancy. However, neither will it bring in one that is not adjacent to any member of the focal firm's alliance portfolio, as it renders a lot of cooperative barriers for partners without any common backgrounds. However, opposite to our expectation, technology similarity has a U-shaped relationship with the chance a partner will be chosen. This means a partner firm is less likely to be chosen if it has a lot technology similarity with the focal firm's alliance technology portfolio. However, when such similarity exceeds certain threshold, such firm is more likely to be chosen as a partner. Consistent with the sign of the intercept, since firms are in general explorative, technologically dissimilar partners are chosen more. However, for the exploitative area, technologically similar partners are more likely to be chosen. This might explain why technology similarity first decreases partner choice likelihood at the beginning but increases once certain threshold is reached. Cumulative number of patents as a quality measure for choice partner is expected to have a positive sign that a partner with more patents is more likely to be chosen. However, the result shows opposite sign that partners with fewer patents are more likely to be chosen. This is due to the fact that most of the partner firms in our data are small bio-tech firms, which are relatively new and hence have fewer registered patents, as compared with large multinational companies. Hence, under such context, cumulative patents might not be a good quality measure for small and new firms. The degree of independence of firms within explorative and exploitative nests are 0.464 (s.e. = 0.068) and 0.629 (s.e. = 0.085), respectively, which are consistent with utility maximization theory for values between 0 and 1.

6.2 Knowledge Assimilation

In line with Cohen and Levinthal, (1990), where they assert past knowledge not only facilitates learning of future knowledge in the same discipline, but also accumulates one's ability of learning, the so called second order of learning. Hence, the more past knowledge a firm possesses, the easier it is for the firm to learn new knowledge. We use a firm's knowledge stock to explain a firm's ability to internalize or assimilate new knowledge. We find that a nonlinear relationship between past knowledge stock and assimilation capability. Initially, increase of knowledge stock indeed increases a firm's assimilation capability; however, too much past knowledge decreases a firm's assimilation capability. One plausible explanation for such phenomenon is the lock-in effect brought by knowledge stock that when the firm is overloaded with past knowledge, it is difficult to appreciate or engage into a new knowledge regime. The knowledge similarity between the focal firm and the partner firm will first facilitate knowledge transfer. As expected, when there is too much similarity between the two firms, assimilation decreases as there is basically not so much to learn. Past partnership is statistically significant with the expected sign that firms that have prior alliance experience find it easier to learn from each other due to established understanding and familiarized routine. The effect of forgetting is significant; however, the magnitude of knowledge obsolescence is rather small.

6.3Knowledge Transformation

We use a production function to measure the efficiency of how production factors such as knowledge stock and R&D investment can be transformed into the knowledge output, which is measured by the number of patents registered in each year. The quality threshold stands for the level of quality that needs to be produced to have the corresponding number of innovative outputs. For the ease of exposition, we plot the quality thresholds in Figure 2.2. As can be seen from the figure, the quality threshold line is relatively steep at the beginning and becomes almost a straight line after 8 patents per year. This is consistent with the pattern of a typical learning curve that marginal effort to produce an extra patent is large at the beginning. However, such effort decreases and levels off once the learning routine and capability are established. Producing an additional output becomes more or less a replication of previous output and the effort level increases proportionally. In terms of efficiency in transforming knowledge stock and R&D expenditure, the output elasticity for R&D expenditure doubles the output elasticity of knowledge capital. This means that the firms under study are more efficient in transforming monetary capital into innovative output compared to knowledge capital realization.

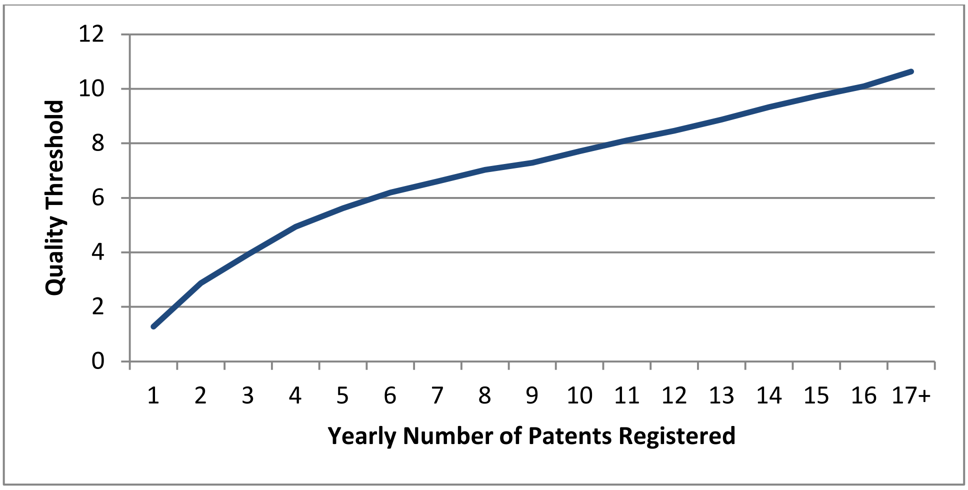
Our main findings are 1) firms show an inter-temporally or cyclically ambidexter when maintaining the balance between explorative and exploitative partnerships. 2) Knowledge similarity shows a bi-polar pattern that partner firms with either very different or very similar knowledge structure are more likely to be chosen. 3) Past knowledge does enhance knowledge assibilation, but too much past knowledge impedes knowledge assimilation. 4) Firms are less efficient in knowledge utilization than in capital utilization.

Table 6: Model Estimation Results

Stage1: Choices of Knowledge Partner		
Variable	Explanation	Estimates (Std. Error)
α_1	Constant for explorative nest	12.958(2.062)
$FOV_{i,t}$	Degree of technology concentration under explorative nest	19.177(4.073)
$FOV^2_{i,t}$		-32.195(4.711)
$DOC_{i,2,t}$	Degree of technology concentration under exploitative nest	43.917(5.917)
$DOC^2_{i,2,t}$		-28.029(4.431)
$TS_{i,j,t}$	Technology similarity between firm j and firm i's alliance portfolio.	-4.793(0.621)
$TS^2_{i,j,t}$		7.858(0.550)
$CP_{j,t}$	Cumulative number of patents of firm *j*	-0.183(0.0266)
λ_1		0.464(0.068)
λ_2		0.629(0.085)
Stage2-3: Knowledge Assimilation and Creation		
Variable	Explanation	Estimates (Std. Error)
$\boldsymbol{\delta}$	Intercept for assimilation capability	0.355 (0.208)
$K_{i,t-1}$	Knowledge stock of firm *i*	7.544 (2.239)
$K^2_{i,t-1}$		-22.024 (1.097)
$TS'_{i,j,t}$	Technology similarity between firm *i* and firm ***j***	104.130 (0.656)
$TS^{2\prime}_{i,j,t}$		-40.210 (0.028)

$PC_{i,j,t}$	$PC_{i,j,t}$ =1if there is past cooperation between firm *i*and firm ***j***	14.512 (0.029)
f	Forgetting factor f	0.001(1.149E-09)
θ_1	Quality threshold for producing 1 patent	1.276 (0.028)
θ_2	Quality threshold for producing 2 patents	2.876 (0.114)
θ_3	Quality threshold for producing 3patents	3.931 (0.147)
θ_4	Quality threshold for producing 4 patents	4.941 (0.173)
θ_5	Quality threshold for producing 5 patents	5.618 (0.193)
θ_6	Quality threshold for producing 6 patents	6.193 (0.207)
θ_7	Quality threshold for producing 7 patents	6.602 (0.213)
θ_8	Quality threshold for producing 8 patents	7.032 (0.220)
θ_9	Quality threshold for producing 9 patents	7.289 (0.225)
θ_{10}	Quality threshold for producing 10 patents	7.711 (0.233)
θ_{11}	Quality threshold for producing 11 patents	8.105 (0.240)
θ_{12}	Quality threshold for producing 12 patents	8.462 (0.246)
θ_{13}	Quality threshold for producing 13 patents	8.869 (0.256)
θ_{14}	Quality threshold for producing 14 patents	9.325 (0.271)
θ_{15}	Quality threshold for producing 15 patents	9.724 (0.287)
θ_{16}	Quality threshold for producing 16 patents	10.091 (0.306)
θ_{17}	Quality threshold for producing 17 and more patents	10.629 (0.336)
$\boldsymbol{\alpha}$	Knowledge efficiency	0.538 (0.118)
$\boldsymbol{\beta}$	Capital efficiency	1.151 (0.054)

Figure 11: Quality Threshold for Annual Number of Patents Registered



7. Conclusion

In this research, we propose a unified framework to model absorptive capacity. We unpack the three dimensions of absorptive capacity through directly measuring each dimension of the construct. In such way, we recognize the multidimensional nature of absorptive capacity. Also, direct measurement allows us to test the actual impact of absorptive capacity other than the hypothesized effect that is usually claimed by extent research. We also take into account the sequential relationship from one dimension to another dimension so that we can see how the previous dimensions will affect the following dimensions. Last but not least, we allow interaction between firm and its immediate knowledge environment and depict a dynamic co-evolution process that abides by the original intension of Cohan and Levinthal (1990) that absorptive capacity is cumulated and path-dependent.

In the empirical validation part, we find when choosing a knowledge partner, firms are more likely to build an explorative partnership than an exploitative partnership. Instead of following Hannan and Freeman's (1984) argument of structural inertia, firms in pharmaceutical industry are better described with ambidexterity proposition. However, we find that firms are only inter-temporally, more precisely, cyclically ambidexter that they maintain the balance between explorative and exploitative partnerships from period to period. We further find that knowledge similarity shows a bi-polar pattern that partner firms with either very different or very similar knowledge structure are more likely to be chosen.

The effect of past knowledge on knowledge assimilation is empirically tested. In line with Cohen and Levinthal’s (1990) conjecture, the stock of past knowledge enhances knowledge assimilation. However, too much knowledge stock impedes knowledge assimilation. Similar pattern is also found for knowledge similarity. For knowledge transformation, we find that firms are less efficient in knowledge utilization compared with capital utilization. Further, we find that the effort needed to produce an extra patent in a year drops quickly and levels off at 7 and more than patents.

With our proposed framework, the widely available alliance data can be used to test empirically the antecedents that determine each dimension of firm’s absorptive capacity. Our conclusion is made on a data with relatively simple structure. However, a much richer data can also be well accommodated with our framework. The framework first sheds light on knowledge partner selection criteria and how the knowledge pool composed of different knowledge partners can contribute to knowledge assimilation. It then reveals different efficiency in inputs (knowledge versus capital) utilization. Our research can be used as an industry benchmark on absorptive capacity building. It also gives implications to firms on diagnosis of the strength of each dimension when they are building absorptive capacity from period to period. To sum up, our work provides a basic framework to measure absorptive capacity with empirical data. It can accommodate investigation of various factors that impact firm's absorptive capacity, thus help firms to identify the room for improvement when developing absorptive capacity.

8. Limitation and Future Research

In this study, we propose to model absorptive capacity in a unified framework. The focus is more on how the three dimensions could be modeled and combined in a meaningful way. However, structural modeling of each dimension would also render good implications. There is much more to do to improve the model.

In terms of partner choice, we made a somewhat strong assumption that choice is unilateral, while in reality it is not. Alliance itself requires dyadic consensus. Secondly, we do not have any demographic data on partner firms, which might influence the value of a particular partner firm, such as the size, the location, etc. We do not have any market performance data of the innovative products, otherwise we could put into the model technological potential, which is an important consideration for knowledge selection or partner selection process. In terms of knowledge assimilation, we assume alliance is the only source where firms can absorb external knowledge, which is not necessarily the case. Knowledge absorbed from internal development also contributes to absorptive capacity. When it comes to knowledge transformation, we do not embed any structure on what determines transformation efficiency. Despite the fact that our model is the first in the area to take into account the sequential relationship between the dimensions, we did not model how knowledge utilization could echo back to affect partner choice in the next period. Such feedback effect definitely exists, as alliance process itself is a learning process that instructs future decision. In this study, we propose a general unified framework to model absorptive capacity. However, we believe that modeling absorptive capacity by building microstructure generates implications and insights that cannot be obtained through reduced-form modeling.

All the above mentioned limitations are good venues to follow up.

BIBLIGRAPHY

Ahuja, G., Katila, R. (2001), "Technological Acquisition and the Innovation Performance of Acquiring Firms: A longitudinal Study." *Strategic Management journal*, 22, 197-220.

Ajzen, I., Fishbein, M. (1980), *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, New Jersey: Prentice-Hall

Alba, J.W., Hutchinson, W. J., (1987) "Dimensions of Consumer Expertise", *Journal of Consumer Research*, 13(4), 411-454.

Allenby, G., M. and Rossi, P., E. (1991), "Quality Perceptions and Asymmetric Switching Between Brand", *Marketing Science*, 10(3), 185-204.

Anderson, N., H. (1978), "Cognitive Algebra: Integration Theory Applied to Social Attribution", in *Cognitive Theories in Social Psychology-Papers from Advances in Experimental Social Psychology*, ed. Berkowitz, Leonard, 1-101. London: Academic Press

Arundel, A., Kabla, I., (1998), "What percentage of innovations are patented? Empirical Estimates for European firms" *Research Policy*, 27, 127-141.

Braun, K. A. (1999), "Post-experience Advertising Effects on Consumer Memory", *Journal of Consumer Research*, 25(4), 319-334.

Cacioppo, J., T., Petty, R., E. (1982), "The Need For Cognition", *Journal of Personality and Social Psychology*, 42, 116-131

Carlston, D., E. (1980) "The Recall and Use of Traits and Events in Social Inference Processes", *Journal of Experimental Social Psychology.* 16, 303-328

Chaiken, S., (1980), "Heuristic versus systematic information processing and the use of source versus message cues in persuasion", *Journal of Personality and Social Psychology*, 39, 752-756.

Chiang, J., (1991), "A Simultaneous Approach to the Whether, What and How Much to Buy Questions", *Marketing Science*, 10(4), 297-315.

Ching, A., Ishihara, M., (2010), "The Effects of Detailing on Prescribing Decisions under Quality Uncertainty," *Quantitative Marketing and Economics*, 8(2), 123-165.

Chintagunta, P., K., (1993) "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households", *Marketing Science*, 12(2), 184-208.

Cohen, W. M., Levinthal, D. A. (1990) "Absorptive Capacity: A New Perspective on Learning and innovation", *Administrative Science Quarterly*, 35, 128-152.

Cook, T.D. B.R Flay. (1978), "The Persistence of Experimentally Induced Attitude Change", *Advances in Experimental Social Psychology.* 11, 1–57.

Denes-Raj, V., Epstein, S., (1994). “Conflict between intuitive and rational processing: When people behave against their better judgment”, *Journal of Personality & Social Psychology*, 66, 819-829.

Dubé, Jean-Pierre. 2004. “Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks”, *Marketing Science*, 23(1), 66-81

Eisenhardt, K. M., J. A. Martin. (2000) “Dynamic capabilities: What are they?” *Strategic Management Journal*, 21, 1105-11121.

Erdem, T., Keane, M., (1996), “Decision Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets”, *Marketing Science*, 15(1), 1-20.

Ernst, H., (1995), “Patenting Strategies in the German Mechanical Engineering Industry and their Relationship to Firm Performance” *Technovation* 15(4), 225-240.

Ernst, H. (2001), “Patent applications and subsequent changes of performance: evidence from time-series across-section analyses on the firm level”, *Research Policy* 30, 143-157.

Estes, W. K. (1997), “Processes of memory loss, recovery, and distortion”, *Psychological Review*, 104 (1), 148–169

Fishbein, M., Ajzen, I., (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading, Mass.: Addison-Wesley Pub.

Fiske, S., Pavelchak, M., A., (1986) “Category-Based versus Piecemeal-Based Affective Responses- Developments in Schema-Triggered Affect”, In *Handbook of Motivation and Cognition- Foundations of Social Behavior*, Ed. Sorrentino, Richard M. and Higgins, E. Tory, 167-203. John Wiley & Sons

Grant, R.M. , Baden-Fuller, C. (2004) “A knowledge Accessing Theory of Strategic Alliances” *Journal of Management Studies*, 41, 61-84.

Griliches, Z., (1997), “Issues in Assessing the Contribution of Rsearch and Development to Productivity Growth”, *The Bell Journal of Economics*, 10 (1), 92-116.

Griliches, Z., (1990), “Patent Statistics and Economic Indicators: a Survey” *Journal of Economic Literature*, 18 (4), 1661-1707.

Hannan, M. T., Freeman J. H., (1984), “Structural Inertia and Organizational Change” American *Sociological Review*, 49 149-164.

He, Z. L., Wong, P. K., (2004) “Exploration vs. Exploitation: An empirical Test of the Ambidexterity Hypothesis” *Organization Science,* vol. 15(4), 481-494.

Hoch, S., J., Ha, Young-Won., (1986), “Consumer learning: Advertising and the ambiguity of product experience”, *Journal of Consumer Research*, 13, 221-233.

Hutchinson, W., J., Alba, J., W., (1991), “Ignoring Irrelevant Information: Situational Determinants of Consumer Learning”, *Journal of Consumer Research*, 18(3), 325-345

Inkpen, A. C., Dinue, A., (1998) “Knowledge Management Processes and International Joint Ventures,” *Organization Science*, Vol.9, 4, 454-468;

Kedia, B. L., & Bhagat, R S. (1988), “Cultural constraints on transfer of technology across nations: Implications for research in international and comparative management.” *Academy of Management Review*,13, 559-571.

Koriat, A., Goldsmith, M., Pansky, A., (2000) “Toward a psychology of memory accuracy”, *Annual Review Psychology*, 51, 481-537

Koza, M. P., & Lewin, A. Y. (1998) “The co-evolution of Strategic Alliances.” *Organization Science*, 7, 255-264.

Lane,P.J., Koka, B.R., & Pathak, S. (2006) “The Reification of Absorptive Capacity: A Critical Review and Rejuvenation of the Construct” *Academy of Management Review*,4, 833-863.

Lane,P.J., Salk, J. E., &Lyles, M.A. (2001) “Absorptive capacity, learning and performance in international joint ventures.” *Strategic Management Journal*, 22: 1139-1161

Lane,P.J., & Lubatkin, M. (1998) “Relative Absorptive Capacity and Interorganizational Learning” *Stragegic Management Journal*, 19, 461-477

Lasson, R., Bengtsson, L., Henriksson, K, Sparks, J. (1998) “The interorganizational Learning Dilemma: Collective Knowledge Development in Strategic Alliances,” *Organization Science*, 9(3), 285-305.

Lavie, D., Rosenkorpf, L. (2006) “Balancing exploration and exploitation in alliance formation” *Academy of Management Journal*, 49, 797-818

Levin, I., P., Gaeth, G., J., (1988) “How Consumers Are Affected by the Framing of Attribute Information before and after Consuming the Product”, *Journal of Consumer Research*, 15(3), 374-378.

Lichtenthaler, U. (2009) “Absorptive Capacity, Environmental Turbulance, and the Complementarity of Organizational Learning Process”, *Academy of Management Journal*, 52(4), 822-846

Lovett, M., J., (2008), “Unstable Consumer Learning Models: Structural Models and Experimental Investigation,” Unpublished Doctoral Dissertation, Duke University.

Mantel, S., P., Kardes, F., R. (1999) “The Role of Direction of Comparison, Attribute-Based Processing, and Attitude-Based Processing in Consumer Preference,” *Journal of Consumer Research*. 25(4) 335-352

Meeus, M. T. H., Oerlemans, L., A. G., & Hage, J. (2001) “Patterns of Interactive Learning in a High-tech Region.” *Organization Studies*, 22, 145-172.

Mehta, N., Rajiv, S., and Srinivasan, K., (2003) “Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation”, *Marketing Science*, 22(1), 58-84

Mehta, N., Rajiv, S., and Srinivasan, K. (2004), “Role of forgetting in memory-based choice decisions: a Structural Model”, *Quantitative Marketing and Economics*, 2, 107-140.

Meyers-Levy, J., Maheswaran, D., (1991) “Exploring Differences in Males’ and Females’ Processing Strategies”, *Journal of Consumer Research*, 18, 63-70

Mowery, D.C., Wxley, J.E. Silverman, B.S. (1996) “Strategic Alliances and Interfirm Knoweledge Transfer” *Strategic Management Journal*, 17, 77-91

Mullainathan, S., (2002) “A Memory Based Model of Bounded Rationality”, *Quarterly Journal of Economics*, CXVII (3):735-774

Nichools-Nixon, C. (1993) “Absorptive Capacity and Technology Sourcing: Implications for the Responsiveness of Established Firms”, unpublished Ph.D. dissertation, Purdue University.

Nisbett, R., E., Peng, K., Choi, I., Norenzayan, A., (2001) “Culture and Systems of Thought: Holistic vs. Analytic Cognition”, *Psychological Review*, 108(2), 291-310

Payne, D. G.; Elie, C. J.; Blackwell, J. M. Neuschatz, J. (1996) “Memory illusions: recalling, recognizing, and recollecting events that never occurred”, *Journal of Memory and Language*, 35,261-285.

Payne, J., W., Bettman, J., R., Coupey, E., Johnson, E., J., (1992) “A Constructive Process View of Decision Making: Multiple Strategies in Judgment and Choice,” *Acta Psychologica*, 80(1-3), 107-141

Park, J., W., Hastak, M., (1994) “Memory-Based Product Judgments: Effects of Involvement at Encoding and Retrieval”, *Journal of Consumer Research*, 21(3), 534-547

Rao, H., Drazin, R. (2002), “Overcoming Resource Constraints on Product Innovation By Recruiting Talent From Rivals: A Study of the Mutual Fund Industry,1984-94”. *Academy of Management Journal*, 45, 491-507.

Roediger, H., L., McDermott, K. B.,(2000), “Tricks of memory”, *Current Directions in Psychological Science*, 9,123-127

Rubin, D. C., Wenzel, A. E., (1996) “One hundred years of forgetting: A quantitative description of retention”, *Psychological Review*, 103,734-760

Sanbonmatsu, D., M., Fazio, R., H., (1990) “The Role of Attitudes in Memory-Based Decision Making”, *Journal of Personality and Social Psychology*, 59(4), 614-622

Scherer, F. M., (1965) “Corporate Inventive Output, Profits and growth”, the Journal of Political Economy 73(3), 290-297.

Schildt, H., Keil, T., Maula, M. (2012) “The temporal Effects of Relative and Firm-Level Absorptive Capacity on Interorganizational Learning” Strategic Management Journal, 33(10) 1154-1173.

Seggie, S.H. and Griffith, D.A. (2009), “What does it take to get promoted in Marketing Academia? Understanding exceptional publication productivity in the leading marketing journals,” *Journal of Marketing*, 73, 122-132.

Simonin, B. L. (1999) “Ambiguity and the Process of Knowledge Transfer in Strategic Alliances” *Strategic Management Journal*, 20, 595-623

Snodgrass, G. 1997. “The Memory Trainers”, in *Mind and Brain Sciences in the Twenty-First Century*, Ed. Robert L. Solso, Cambridge, MA: MIT Press.

Sorenson, J. B., & Stuart, T. E. (2000), “Aging, Obsolescence, and Organizational Innovation.” *Administrative Science Quarterly*, 45, 81-112.

Szulanski, G., (1996), “Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm”, *Strategic Management Journal, 17, 27-43*

Trope, Y. (1978) “Inferences of personal characteristics on the basis of information retrieved from one’s memory”, *Journal of Personality and Social Psychology*, 36, 93–106.

Tsai, W. P. (2001), “Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance.” *Academy of Management Journal*, 44, 996-1004.

Train, K. (2009) “Discrete Choice Methods with Simulation” 2nd Edition, Cambridge University Press, 78

Tulving, E., (1972) “Episodic and Semantic Memory”, In *Organization of Memory*, eds. E. Tulving W. Donaldson, 381-403. New York: Academic Press

Tulving, Endel. 1983. *Elements of Episodic Memory*. New York: Oxford University Press

Wernerfelt, B. (1984) “A Resource Based View of the Firm” *Strategic Management Journal*, 5, 2, 171-180

Wyer, R., S., Jr. Srull, T., K., (1989) *Memory and Cognition in its Social Context.* New Jersey: Lawrence Elbaum Associates, Publishers.

Zahra, S. A., & George, G. (2002), “Absorptive Capacity: A Review, Reconceptualization, and Extension” *Academy of Management Review*, 27,185-203

Appendix A

Table 7: Table of Notations for Essay1

Model Primitives	
Notation	Definition
S	Superscript S stands for values when they are stored
R	Superscript R stands for values when they are recalled
SE	Superscript SE stands for semantic processing
EP	Superscript EP stands for episodic processing
^	^ stands for realized value that is observed by consumer but not by econometrician
q_j	true quality of brand *j*
$\lambda_{j,t}$	Quality signal brand *j* at purchase occasion *t*
$\eta_{j,t}$	Noise due to inherent quality variation of brand *j* at period *t*
σ^2_λ	Volatility due to inherent quality variation
$q_{j,0}$	Consumer's initial quality perception about brand *j*
ω_0	the expectation of the initial quality perception of brand *j*
σ^2_λ	the variance of the initial quality perception of brand *j*
$p_{j,t}$	price for brand *j* at period *t*
Memory Formation and Evolution	
q^S_0	Initial quality perception that is stored
$\hat{\lambda}^S_{j,t-1}$	Realized quality signal of brand *j* that is received and observed only by consumer
$\hat{\lambda}^R_{j,\tau,t}$	quality signal that was received in period τand recalled in period *t*
$q^S_{j,t-2}$	Stored value of perceived quality for brand *j* in period *t-2*
$\omega^{S,SM}_{j,t-1}$	Updated posterior mean that is immediately stored in semantic memory at consumption period *t-1* for brand *j*
$\psi^{S,SM}_{j,t-1}$	posterior variance that is immediately stored in semantic memory at consumption period *t-1* for brand *j*
$q^R_{j,t-2}$	Recalled quality perception of brand *j* at consumption occasion *t-2*
$\omega^R_{j,t-2}$	Recalled prior mean at consumption period *t-1* for brand *j*

$\psi^{R}_{j,t-2}$	Recalled prior variance consumption period *t-1* for brand *j*
$d_{j,t-1}$	=1 if brand *j* was purchased at purchase occasion *t-1* =0 if brand *j* is not purchased at purchase occasion *t-1*
$\omega^{EM}_{j,t-1}$	posterior mean that is constructed through recalling from episodic memory
$\phi^{2}_{j,\tau,t}$	Posterior variance for signals of brand *j* received in period $\tau$and recalled in period *t*
Modelling of Forgetting	
Notation	Definition
$\upsilon_{j,t-2}$	$\upsilon_{j,t-2\sim N(0,1)}$ determines the direction of recall error
$\varphi_{j,t-2}$	is the scale of recall error for brand *j* at period *t-2*
B^{SM}	forgetting rate of semantic memory
B^{EM}	forgetting rate of episodic memory
Choice Probability	
$Pr_{i,j,t}$	probability that brand *j* is chosen by consumer *i* at purchase occasion *t*
Pr[SM]	probability that consumer uses semantic memory
α_i	consumer *i*'s tendency to use semantic memory
α	population's mean in its tendency to use semantic memory
σ^{2}_{α}	Variance of population tendency to use semantic memory
$\Lambda_{i,ti}$	string of signals that are received by consumer till purchase occasion *t*
$\Gamma_{i,ti}$	string of context specific information received by consumer till purchase occasion *t*
$V_{i,ti}$	a matrix of J x t_i iid standard normal random errors
Δ	vector of population parameters

Table 8: Table of Notations for Essay2

Stage1: Choices of Knowledge Partner	
Notation	Definition
$U_{i,j,t}$	Utility obtained by firm *i* should firm *j* is chosen as knowledge partner in period *t*
$W_{i,j,t}$	A vector of variables that describe the nest *k* of firm *i* at period *t*
k	1:explorative nest 2: exploitative nest
α_k	Intercept for nest *k*
$FOV_{i,1,t}$	Field of vision of firm *i* at period *t* under explorative nest
$DOC_{i,2,t}$	Degree of technology concentration of firm *i* at period *t* under exploitative nest
$V_{k,j,t}$	A vector of variables that describe the firm *j* at period *t* nest *k* at period *t*
$TS_{i,j,t}$	Technology Similarity between firm *j* and firm *i*'s alliance portfolio in period *t*.
$CP_{j,t}$	Cumulative number of patents of firm *j* in period *t*
$P_{i,t}$	Firm *i*'s alliance portfolio's technology vector in period*t*
$X_{j,t}$	The vector of control variables of firm *j* in period *t*
$P_{i,j,t}$	Choice probability of firm *j* being chosen as a partner by firm *i* in period *t*
$P_{i,k,t}$	Probability that nest *k* is chosen by firm *i* in period *t*
$P_{j\|k,t}$	Probability that firm *j* is chosen conditioned that nest *k* is chosen by in period *t*
λ_k	Degree of independence between the alternative firms under nest *k*
$I_{i,k,t}$	Inclusive value of nest *k* in period *t*
Stage2: Knowledge Assimilation and Creation	
Notation	Definition
$K_{i,t-1}$	Knowledge stock of firm *i* till period *t*
$k_{i,t}$	Knowledge inflow of firm *i* in period *t*
i_t	Set of partners ally with firm *i* in period *t*
$AC_{i,j,t}$	Firm *i*'s capability in assimilating knowledge from firm *j* in period t
δ	Intercept for assimilation capability
$TS^{i}_{i,j,t}$	Technology similarity between firm *i* and firm *j* in period *t*
$PC_{i,j,t}$	$PC_{i,j,t}$=1if there's past cooperation between firm *i* and firm j till period t
f	Rate of forgetting

Stage3: Knowledge Transformation	
A	Total productivity factor
$C_{i,t}$	R&D investment by firm *i* in period t
$P(K_{i,t}, C_{i,t})$	The continuous intermediate product produced by firm***i*** in period t
α	Efficiency of knowledge utilization
β	Efficiency of capital utilization
$q_{i,t}$	Number of patents registered by firm *i* in period t
θ_i	Quality threshold for producing *i*patents *i=1,2,…16*
θ_{17+}	Quality threshold for producing more than 17 patents
Variable Operationalization	
Notation	Definition
T_N	New Technology area that firm *i* has never explored by period *t*
T_O	New Technology area that firm *i* has explored by period *t*
T^i_t	Set of technologies firm *i* has explored by period *t*

Appendix B

Proof of Proposition1

The posterior precision under semantic retrieval is the sum of recalled prior precision and the signal precision. By replacing the prior precision recursively, we can get

$$\left(\frac{1}{\psi_N^S}\right)^2 = \frac{1}{(\psi_{N-1}^{SR})^2} + \frac{1}{\sigma_\lambda^2} = \frac{1}{e^{bW}} \cdot \frac{1}{\psi_{N-1}^2} + \frac{1}{\sigma_\lambda^2} = \frac{1}{e^{bW}} \cdot \left(\frac{1}{(\psi_{N-2}^{SR})^2} + \frac{1}{\sigma_\lambda^2}\right) + \frac{1}{\sigma_\lambda^2}$$

$$= \frac{1}{e^{bW}} \cdot \left(\frac{1}{e^{bW}} \cdot \frac{1}{\psi_{N-2}^2} + \frac{1}{\sigma_\lambda^2}\right) + \frac{1}{\sigma_\lambda^2} = \frac{1}{e^{b2W}} \cdot \frac{1}{\psi_{N-2}^2} + \frac{1}{e^{bW}}\frac{1}{\sigma_\lambda^2} + \frac{1}{\sigma_\lambda^2}$$

$$\cdots$$

$$= \frac{1}{e^{bW}} \cdot \left(\frac{1}{e^{bW}} \cdot \frac{1}{\psi_{N-2}^2} + \frac{1}{\sigma_\lambda^2}\right) + \frac{1}{\sigma_\lambda^2} = \frac{1}{e^{b2W}} \cdot \frac{1}{\psi_{N-2}^2} + \frac{1}{e^{bW}}\frac{1}{\sigma_\lambda^2} + \frac{1}{\sigma_\lambda^2}$$

$$= \frac{1}{e^{bNW}} \cdot \frac{1}{\psi_0^2} + \sum_{n=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{1}{\sigma_\lambda^2}$$

Posterior precision for episodic retrieval is the weighted sum of prior precision and recalled signal precision. Hence, after expanding the recalled signal precision, we get

$$\left(\frac{1}{\psi_N^E}\right)^2 = \left(\frac{1}{\psi_0^2}\right)^R + \left(\frac{1}{\sigma_{\lambda,N+1}^1}\right)^2 + \left(\frac{1}{\sigma_{\lambda,N+1}^2}\right)^2 + \cdots + \left(\frac{1}{\sigma_{\lambda,N+1}^{N-1}}\right)^2 + \frac{1}{(\sigma_\lambda^2 + 1)}$$

$$= \frac{1}{e^{bNW}} \cdot \frac{1}{\psi_0^2} + \sum_{\tau=0}^{N-1} \left(\frac{1}{\sigma_{\lambda,N+1}^\tau}\right)^2$$

$$= \frac{1}{e^{bNW}} \cdot \frac{1}{\psi_0^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{1}{(\sigma_\lambda^2 + 1)}$$

Now we examine the asymptotic property of the posterior precision. For both memories, the first term is the same and when N $\rightarrow\infty$, it becomes 0.

$$\lim_{N\to\infty}\frac{1}{e^{bNW}}\cdot\frac{1}{\psi_0^2}=0$$

The second term becomes the sum of a geometric sequence. For semantic memory,

$$\sum\nolimits_{\tau=0}^{N-1}\frac{1}{e^{b\tau W}}\cdot\frac{1}{\sigma_\lambda^2}=\frac{1}{\sigma_\lambda^2}+\frac{1}{e^{bW}}\cdot\frac{1}{\sigma_\lambda^2}+\frac{1}{e^{b2W}}\cdot\frac{1}{\sigma_\lambda^2}+\frac{1}{e^{b3W}}\cdot\frac{1}{\sigma_\lambda^2}+\cdots+\frac{1}{e^{bNW}}\cdot\frac{1}{\sigma_\lambda^2}$$

$$=\frac{1}{\sigma_\lambda^2}\left(1+\frac{1}{e^{bW}}+\frac{1}{e^{b2W}}+\frac{1}{e^{b3W}}+\cdots+\frac{1}{e^{b(N-1)W}}\right)$$

$$=\lim_{N\to\infty}\sum\nolimits_{\tau=1}^{N}\frac{1}{e^{b\tau W}}\cdot\frac{1}{\sigma_\lambda^2}=\frac{e^{bW}}{e^{bW-1}}\cdot\frac{1}{\sigma_\lambda^2}$$

It is the same for episodic memory

$$\sum\nolimits_{\tau=0}^{N-1}\frac{1}{e^{b\tau W}}\cdot\frac{1}{\sigma_\lambda^2}=\frac{e^{bW}}{e^{bW}-1}\cdot\frac{1}{\sigma_\lambda^2}$$

Proof of Proposition 2:

The posterior mean of quality from semantic retrieval is

$$\omega_N^S=\frac{\frac{\omega_{N-1}^R}{(\psi_{N-1}^R)^2}+\frac{\lambda_N}{\sigma_\lambda^2}}{\frac{1}{(\psi_{N-1}^R)^2}+\frac{1}{\sigma_\lambda^2}}=\frac{\frac{\omega_{N-1}}{(\psi_{N-1}^R)^2}+\frac{\nu\cdot\psi_{N-1}\sqrt{e^{bW}-1}}{\psi_{N-1}^2e^{bW}}+\frac{\lambda_N}{\sigma_\lambda^2}}{\frac{1}{(\psi_{N-1}^R)^2}+\frac{1}{\sigma_\lambda^2}}$$

$$=\frac{\frac{\omega_{N-1}}{(\psi_{N-1}^R)^2}+\frac{\lambda_N}{\sigma_\lambda^2}}{\frac{1}{(\psi_{N-1}^R)^2}+\frac{1}{\sigma_\lambda^2}}+\frac{\frac{\nu\cdot\psi_{N-1}\sqrt{e^{bW}-1}}{\psi_{N-1}^2e^{bW}}}{\frac{1}{(\psi_{N-1}^R)^2}+\frac{1}{\sigma_\lambda^2}}$$

And that under episodic retrieval is

$$\omega_N^E = \frac{\frac{\omega_0^R}{(\psi_0^R)^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\lambda_{N+1}^{N-\tau}}{\sigma_\lambda^2}}{\frac{1}{(\psi_0^R)^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{1}{\sigma_\lambda^2 + 1}}$$

$$= \frac{\frac{\omega_0^R}{(\psi_0^R)^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\lambda^{N-\tau} + \nu \cdot \sigma_\lambda \cdot \sqrt{e^{b\tau W} - 1}}{\sigma_\lambda^2}}{\frac{1}{(\psi_0^R)^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{1}{\sigma_\lambda^2 + 1}}$$

Since we know that when N approaches +∞, the posterior variance equals a constant (from Proposition 1), when we analyze the limit of $\boldsymbol{\omega_N^S}$, the denominator is a constant, and we only need to look at the numerator.

$$\frac{\omega_N^S}{\psi_N^2} = \frac{\omega_{N-1}^R}{(\psi_{N-1}^R)^2} + \frac{\lambda_N}{\sigma_\lambda^2} = \frac{1}{e^{bW}} \cdot \frac{\omega_{N-1}^S}{\psi_{N-1}^2} + \frac{\nu \cdot \psi_{N-1}\sqrt{e^{bW} - 1}}{\psi_{N-1}^2 e^{bW}} + \frac{\lambda_N}{\sigma_\lambda^2}$$

$$= \frac{1}{e^{bW}} \cdot \left(\frac{\omega_{N-2}^R}{(\psi_{N-2}^R)^2} + \frac{\lambda_{N-1}}{\sigma_\lambda^2} \right) + \frac{\nu \cdot \psi_{N-1}\sqrt{e^{bW} - 1}}{\psi_{N-1}^2 e^{bW}} + \frac{\lambda_N}{\sigma_\lambda^2}$$

$$= \frac{1}{e^{bW}} \cdot \left(\frac{1}{e^{bW}} \cdot \frac{\omega_{N-2}^S}{\psi_{N-2}^2} + \frac{\nu \cdot \psi_{N-2}\sqrt{e^{bW} - 1}}{\psi_{N-2}^2 e^{bW}} + \frac{\lambda_{N-1}}{\sigma_\lambda^2} \right) + \frac{\nu \cdot \psi_{N-1}\sqrt{e^{bW} - 1}}{\psi_{N-1}^2 e^{bW}} + \frac{\lambda_N}{\sigma_\lambda^2}$$

$$= \frac{1}{e^{b2W}} \cdot \frac{\omega_{N-2}^S}{\psi_{N-2}^2} + \frac{1}{e^{bW}} \cdot \frac{\nu \cdot \psi_{N-2}\sqrt{e^{bW} - 1}}{\psi_{N-2}^2 e^{bW}} + \frac{\nu \cdot \psi_{N-1}\sqrt{e^{bW} - 1}}{\psi_{N-1}^2 e^{bW}} + \frac{1}{e^{bW}} \cdot \frac{\lambda_{N-1}}{\sigma_\lambda^2} + \frac{\lambda_N}{\sigma_\lambda^2}$$

....

$$\frac{\omega_N^S}{\psi_N^2} = \frac{1}{e^{bNW}} \cdot \frac{\omega_0}{\psi_0^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\lambda_{N-\tau}}{\sigma_\lambda^2}$$

$$+ \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\nu \cdot \psi_{N-1-\tau}\sqrt{e^{bW}-1}}{\psi_{N-1-\tau}^2} \ldots .(A)$$

and

$$\frac{\omega_N^E}{\psi_N^2} = \frac{\omega_0^R}{(\psi_0^R)^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\lambda_{N+1}^{N-\tau}}{\sigma_\lambda^2}$$

$$= \frac{1}{e^{bNW}} \cdot \frac{\omega_0}{\psi_0^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\lambda_{N-\tau}}{\sigma_\lambda^2} + \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{\nu \cdot \sigma_\lambda \cdot \sqrt{e^{b\tau W}-1}}{\sigma_\lambda^2} \ldots ..(B)$$

The first two terms of equations A and B are converging. Let Er_s and Er_Ebe the last term in equation (A) and (B) respectively.

$$Er_s = \nu \cdot \sqrt{e^{bW}-1} \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{1}{\psi_{N-1-\tau}}$$

Let

$$S_n = \sum_{\tau=0}^{N-1} \frac{1}{e^{b\tau W}} \cdot \frac{1}{\psi_{N-1-\tau}}$$

$$= \frac{1}{e^{b(N-1)W}} \cdot \frac{1}{\psi_0} + \frac{1}{e^{b(N-2)W}} \cdot \frac{1}{\psi_1} + \cdots + \frac{1}{e^{bW}} \cdot \frac{1}{\psi_{N-2}} + \frac{1}{\psi_{N-1}}$$

Then

$$S_{n+1} = \sum_{\tau=0}^{N} \frac{1}{e^{b\tau W}} \cdot \frac{1}{\psi_{N-1-\tau}}$$

$$= \frac{1}{e^{bNW}} \cdot \frac{1}{\psi_0} + \frac{1}{e^{b(N-1)W}} \cdot \frac{1}{\psi_1} + \cdots + \frac{1}{e^{bW}} \cdot \frac{1}{\psi_{N-1}} + \frac{1}{\psi_N}$$

i.e.

$$S_{n+1} = \frac{1}{e} S_n + \frac{1}{\psi_N}$$

Proof: Assume $\exists$ n s.t. $n > N, |S_n - X| < \varepsilon$, where **ε** is any positive real value.

$$\therefore X - \varepsilon < S_n < X + \varepsilon$$

Also if $|S_n - X| < \varepsilon$ then $|S_{n+1} - X| < \varepsilon$

$$\because S_{n+1} = \frac{1}{e} S_n + \frac{1}{\psi_N}$$

$$\therefore \left| \frac{1}{e} S_n + \frac{1}{\psi_N} - X \right| < \varepsilon$$

$$\left(\frac{1}{e} - 1\right) X + \frac{1}{\psi_N} - \frac{\varepsilon}{e} < \frac{1}{e} S_n + \frac{1}{\psi_N} - X < \left(\frac{1}{e} - 1\right) X + \frac{1}{\psi_N} + \frac{\varepsilon}{e}$$

$$\therefore |S_{n+1} - X| = \left| \frac{1}{e} S_n + \frac{1}{\psi_N} - X \right| > \left| \left(\frac{1}{e} - 1\right) X + \frac{1}{\psi_N} - \frac{\varepsilon}{e} \right| \ldots . (C)$$

Since $\rightarrow \infty$, $\frac{1}{\psi_N} \rightarrow C$. So from equation (C), we know that $|S_n - X|$ is not smaller than any positive value ε. Hence, there does not exist such n and S_n is not a converging sequence.

Now we look at $Er_E = \frac{\sigma_\lambda \ \nu}{\sigma_\lambda^2 + 1} \cdot \sum_{\tau=0}^{N-1} \frac{\sqrt{e^{b\tau W} - 1}}{e^{b\tau W}}$

Let $\{a_n\} = 0, \frac{\sqrt{e^{bW} - 1}}{e^{bW}}, \frac{\sqrt{e^{b2W} - 1}}{e^{b2W}}, \ldots, \frac{\sqrt{e^{b(N-1)W} - 1}}{e^{b(N-1)W}}$

$$\frac{a_{n+1}}{a_n} = \frac{\frac{\sqrt{e^{bnW} - 1}}{e^{bnW}}}{\frac{\sqrt{e^{b(n-1)W} - 1}}{e^{b(n-1)W}}} = \frac{1}{e^{bW}} \cdot \sqrt{e^{bW} + \frac{e^{bW} - 1}{e^{b(n-1)W} - 1}} < 1$$

Hence, $\{a_n\}$ is a converging sequence to 0 and sum of which is a constant.