A Structural Model of Employee Behavioral Dynamics in Enterprise Social Media

Yan Huang

School of Information Systems and Management & i-lab,

Heinz College,

Carnegie Mellon University

yanhuang@andrew.cmu.edu

Param Vir Singh

Tepper School of Business & i-lab, Heinz College,

Carnegie Mellon University

psidhu@cmu.edu

Anindya Ghose

Stern School of Business, New York University

& i-lab, Heinz College

aghose@stern.nyu.edu

Abstract

We develop and estimate a dynamic structural framework to analyze social media content creation and consumption behavior by employees within an enterprise. We focus in particular on blogging behavior by employees. The model is flexible enough to handle trade-offs between blog posting and blog reading as well as between work-related content and leisure-related content, all of which are ubiquitous in actual blogging forums. The model incorporates the dynamics induced by these aspects in employee behavior. We apply the model to a unique dataset that comprises of the complete details of blog posting and reading behavior of 2396 employees over a 15-month period at a Fortune 1000 IT services and consulting firm. We find that blogging has a significant long-term effect in that it is only in the long term that the benefits of blogging outweigh the costs. There is also evidence of strong competition among employees with regard to attracting readership for their posts. While readership of leisure posts provides little direct utility, employees still post a significant amount of these posts because there is a significant spillover effect on the readership of work posts from the creation of leisure posts. We conduct counterfactual experiments that provide insights into how different policies may affect employee behavior. Incidentally, we find that a policy of prohibiting leisure-related activities can hurt the knowledge sharing in enterprise setting. By demonstrating that there are positive spillovers from work-related blogging to leisure-related blogging, our results suggest that a policy of abolishing leisure-related content creation can inadvertently have adverse consequences on work-related content creation in an enterprise social media setting.

***Keywords*:** Structural modeling, Dynamics, Enterprise social media, Blog posting, Blog reading, Work-related content, Leisure-related content.

1. Introduction

Over the last couple of years, we have seen several firms adopt various kinds of social media such as blogs and wikis for internal use. These social media technologies, often termed collectively as Enterprise 2.0, enable firms to accelerate organizational performance by supporting not just inward facing collaboration but also to come together internally to respond to customer support, innovation, and sales and marketing opportunities. The presence of Enterprise 2.0 tools paves the way for a future where employee-to-employee collaboration and intra-firm interaction become extensions of each other and thus a useful way of providing mutual value for the company and individual employees. Such social media tools share three fundamental properties (McAfee 2009). First, they are typically easy to learn and use. Second, on a social media platform everyone typically starts on an equal, contributing to a blank slate and external reputations matter less than internal reputations footing. Third, and perhaps the most remarkable, is that it allows individuals to engage in both work and leisure-related activities.

In this study, we focus on one specific social media tool, namely enterprise-wide blogging. Increasingly several leading organizations have systems in place in order to encourage their employees to blog (Lee et al. 2006, Aggarwal et al. 2009, Singh et al. 2010a). Prominent adopters are General Motors, IBM, HP, Microsoft, Infosys, Google, Charles Schwab etc. The general thinking within such firms is that enterprise blogging forums can be used to build the structured platform required for an environment that supports emergent innovation. When used effectively, they also may encourage participation in projects and idea sharing, thus deepening a company's pool of knowledge leading to increased number of successful innovations for new products or services (Mckinsey 2009). Thus, the success of blogs as a mechanism to provide the most updated information has increased interest in them as new sources of knowledge creation and dissemination within the enterprise (Lee et al. 2006, Huh et al. 2007, Yardi et al. 2009).

That being said, employees' motivation to blog may not always be in alignment with a firm's objectives. As noted above, a unique attribute of social media is such that it allows individuals to engage in both work and leisure-related content generation[1]. Hence, firms often observe that some employees' blog postings are not relevant to their work knowledge or expertise. If unchecked, such behavior can undermine the very goal of enterprise blogging.

[1] In this paper, by leisure-related blogs we mean to refer to all blog content that is not work-related.

Therefore, it is becoming increasingly important to understand the employees' blogging behavior in an enterprise social media setting in order to help firms devise strategies for influencing user behavior.

Typically, an employee's propensity to blog utility is constrained by the time available for such activities. In most cases, employees spend a certain amount of time on blogging activities on a periodic (daily or weekly) basis. That is, they typically have a fixed budget of time dedicated for such activities. Hence, their utility from blogging is determined by a choice between work vs. leisure-related topics, and consequently this is a trade-off they need to keep in mind. In addition, there is another type of trade-off that some employees need to make in this context - how much time to spend on content generation (i.e., posting blogs) vs. content consumption (i.e., reading blogs).

In general, there can be some reputational benefits from writing posts in an enterprise blog setting. Work-related blogging allows individuals to express their expertise to a broad audience at a low cost. Once employees are identified as "experts", they may receive indirect economic incentives, such as promotion, salary hike, etc (Kavanaugh et al. 2006, Aggarwal et al. 2009). Leisure-related blog posting, on the other hand, can help employees become popular among peers and improve self-satisfaction. There are benefits to reading blogs as well. Work-related knowledge increase may help them in becoming more productive professionally, be more informed about new ideas, open up new opportunities for collaborations, etc. (Huh et al. 2007, Yardi et al. 2009). Leisure-related information can feed to an employee's interests and allow him to relax (Singh et al. 2010a), thereby indirectly improving productivity subsequently.

From the discussion above one can imagine that the benefits of blogging are dependent on the accumulated "reputation" of an employee over time instead of simply the contemporaneous reputation accruing from current period participation. Hence, employees in an enterprise wide setting need to be forward-looking, rather than myopic, when making their decisions about the usage of enterprise 2.0 and the two kinds of trade-offs described above. Take work-related posting and reading as examples. As mentioned above, work-related posting can bring employees opinion leadership status, professional recognition, salary hike, etc. These benefits are based on reputational status, which is a long-term measure of the quality of one's contribution. An employee with a high work-related reputation accumulated over time can still obtain reputation-based benefits even if he/she does not post in the current period. In contrast, an employee with a low work-reputation will find it much harder to build a high reputational status based only on the current period posting. In other words, employees typically cannot get an immediate benefit from a single posting in a given period; rather these reputation related benefits tend to accrue over time based on their blogging contributions over a longer time horizon. Similarly, an employee's knowledge gain from reading blogs can lead to an improvement in productivity and performance only over a longer time horizon.

Another important characteristic of enterprise blogging is that the audience for the enterprise bloggers is purely internal – consisting of the set of employees themselves. Since employees also have limited time available to devote to blog reading, this causes intense competition among employees for readership in order to vie for their attention. Blogs, in general, exhibit a strong Matthew effect, which is "the rich get richer and the poor get poorer" effect. Employees with high reputation continuously tend to receive higher new readership, which further reinforces their high reputations; while employees with low reputation find it hard to attract readers' attention, and consequently, their reputation tends to further decrease over time.

Though "work reputation" and "leisure reputation" provide employees with very different benefits, these two types of reputation can have a strong interdependence. This interdependency could be either positive or negative. Readers may follow bloggers, not specific posts. Employees with high levels of leisure reputation can attract readers' attention to their blog pages, and this in turn can increase the probability of their work-related posts being read, and vice versa. If this were to occur, there could be positive spillovers between work and leisure reputation. On the other hand, it is also possible that employees with high levels of leisure reputation are seen as "not serious" people. If this were to occur, readers would not expect their work-related posts to be of high quality. In this case, leisure reputation can have a negative impact on work reputation. Because of these two countervailing effects, it becomes useful to understand the nature of the interdependency between work and leisure reputation, since they each have different business and policy implications.

Our paper aims to understand employees' motivations for reading and writing work-related and leisure-related posts within an enterprise setting and then draw policy implications based on potential incentive structures. To be more specific, in this paper, we examine the following questions: 1) How do employees in an enterprise social media setting allocate their time to the different types of blogging activities when facing the trade-offs between work- and leisure-related content and between posting and reading content? 2) Is there any interdependence between work- and leisure-related reputation accruing from blogging dynamics, and if so, are these spillovers positive or negative in nature? 3) What factors can explain the observed participation patterns resulting from inter-temporal trade-offs that employees make in such a context?

We present an empirical framework to analyze dynamics in enterprise blogging using a unique individual-level dataset that maps the blog posting and reading behavior of employees. We formulate an employee's decision on when to write or read and what to write or read (in terms of work-related or leisure-related) as a dynamic competition game in the tradition of structural dynamic competition games such as Bajari et al. (2007). Following Erickson and Pakes (1995), we focus on Markov Perfect Equilibrium as a solution concept. We exploit revealed preference arguments from economic theory to estimate the parameters in the model. There are two main components of the employee's utility function, which forms the core of this paper. The ***first component*** is reputation. The reputation incentive is proportional to the cumulative readership of an employees' blog. In each period, an employee receives utility from her cumulative readership, or "reputation status". Further, the work-related and leisure-related posts provide separate kinds of reputation incentives. Individuals also derive utility from the "knowledge state" they are in, which can be improved by reading posts (Singh et al. 2010, Yardi et al. 2009, Huh et al. 2007). This knowledge state forms the ***second component***. Similar to the reputation incentive, we allow the knowledge-based utility derived from work and leisure-related posts to be different.[2] Our model captures inter-employee interactions by allowing competition among employees for attracting readers to their blogs. Besides these, the model incorporates an employee-specific blogging time constraint. Each employee maximizes his/her utility subject to this constraint. This captures the inherent tradeoff between devoting time to work related blogging, leisure related blogging and other non-blogging activities.

[2] Note that in this paper, an employee's knowledge state measures the knowledge he/she gains from blogging internally, i.e., knowledge he/she achieves from external offline activities are not included since we do not observe that in our data.

We apply the model to a unique dataset that comprises of the complete details of blog posting and reading behavior for a large dataset of 2396 employees over a 15-month period at a Fortune 1000 IT services and consulting firm. Our results show that employees are indeed forward-looking. Blogging has a significant long-term effect in that it is only in the long term that the benefits of blogging outweigh the costs. Employees derive higher utility from readership of their work-related posts than their leisure-related posts. There is also evidence of strong competition among employees with regard to attracting readership for their posts. While readership of leisure posts provides little direct utility, employees still post a significant amount of these posts as there is a significant spillover effect on the readership of work posts from the creation of leisure posts. Further, employees derive knowledge based utility from reading both work and leisure related posts.

Using the point estimates of the parameters, we then implement counterfactual experiments to analyze the effects of two different policies. In the first experiment we disallow leisure-related activities. Interestingly, this policy does not help increase work-related posting. Instead, it discourages work-related posting. This suggests that a policy of prohibiting leisure-related activities can hurt the knowledge sharing in enterprise setting. In the second experiment, we relax employees' budget (time) constraint and examine two different policies - a "Medium-Budget" (indicating a medium cost of blogging) and a "High-Budget" (indicating a zero cost of blogging). It turns out that such policies lead to an increase both work-related and leisure-related activities. More interestingly, we find that reading (both work and leisure) increases much more significantly than posting. In addition, we see that the probabilities of both work and leisure posting increase by a larger amount than the probabilities of work and leisure reading when the firm switches from "Medium-Budget" to "High-Budget". More generally, this suggests that the tradeoff between posting and reading is different under different budget constraints.

Our study aims to makes a number of contributions. *First*, it provides insights into an emerging and important phenomenon of why employees in an enterprise social media setting incur the cost of reading and writing content when there appears to be no explicit monetary incentives for doing so. *Second*, to our knowledge we are the first study that provides a structural framework to analyze employee blogging activities in order to derive some important insights into the tradeoffs between work and leisure-related blogging and between content creation (blog writing) and content consumption (blog reading). While prior studies have investigated why individuals create content on blogs, those studies are based on surveys/questionnaires which can be affected by self-reporting bias. In contrast, our study uses actual micro-level blogging activity data from a large enterprise-wide setting to shed light on why individuals blog and model employee behavior accordingly. *Third,* we quantify the nature and magnitude of the interdependence between work and leisure reputation. Our model thus generates some policy implications for firms who may be contemplating prohibiting leisure-related blogging within the enterprise. By demonstrating that there are positive spillovers from work-related blogging to leisure-related blogging, our results suggest that a policy of abolishing leisure-related content creation can inadvertently have adverse consequences on work-related content creation.

The rest of paper is organized as follows. In Section 2 we discuss the literature relevant to this paper. Section 3 presents our model of employee blogging dynamics. Section 4 describes our data, estimation strategies and estimation results. Some counterfactual experiments are discussed in Section 5. Section 6 summarizes and concludes our study.

2. Literature Review

Our research is related to multiple streams of literatures. The first steam of relevant literature relates to the impact of social media and user-generated content. Several studies have investigated how social media or user-generated content influences variables of economic interest. Many studies have focused on how different aspects of user-generated blogs affect product sales and market structure (Venkataraman et al. 2009, Dhar and Chang 2009, Chintagunta et al. 2010). Hofstetter et al. (2009) study how content generated in blogs has a causal effect on the users' ability to form social ties and how users' social ties affects their propensity to create content.

Another stream of work is related to enterprise adoption of social media. Internally, organizational adoption of social media has been found to have improved access to knowledge experts, increased the rate of solving problems, increased the rate of new product development, reduced costs of internal communication (McAfee, 2006; Grossman et al., 2007; Gurram et al., 2008). Some recent research examines how employees' participation influences the benefits firms can get from implementing Web 2.0 applications (Levy 2009). Huh et al. (2007) reveal that blogs facilitates access to tacit knowledge and resources vetted by experts, and, most importantly, contribute to the emergence of collaboration across a broad range of communities within the enterprise. Yardi et al. (2009) study a large internal corporate blogging community using log files and interviews and find that employees expect to receive attention when they contribute to blogs, but these expectations often are not met commensurately. Singh et al. (2010a) study blog reading dynamics of employees within a large firm and find that most of the employees' time is devoted to reading and writing leisure-related posts. Aggarwal et al. (2010) study how negative posts by its employees can actually benefit a firm.

This paper is also related to the emerging literature on the dual role played by users in social media. In online settings, users need to allocate resources between content generation and content usage activities since they can take on the dual role of creators as well as consumers (Trusov et al. 2010). However, there is little research that models or quantifies the interdependencies between how content on a social media platform is created and consumed. The primary reason for such a gap in literature is that while data on content creation is easily available, data on content consumption is typically not available to researchers. However, the consumption information can be retrieved from access logs. A small but emerging stream of work has begun to look at both content creation and consumption data (Ghose and Han 2009, 2010, Albuquerque et al. 2010). Ghose and Han (2009) estimate a dataset encompassing users' multimedia content creation and consumption behavior using mobile phones and find that there exists a negative temporal interdependence between the content generation and usage behavior for a given user. Ghose and Han (2010) find evidence of dynamic learning in such two-sided forums created by mobile Internet based multimedia content. Albuquerque et al. (2010) use data from print-on-demand service of user-created magazines and find that content price and content creator marketing actions have strong effects on purchases.

Another relevant research stream is the motivation for individuals to contribute to online communities, or social media. Many studies consistently show that status is an important factor that explains knowledge creation, retention and transfer (Argote et al. 2003; Thomas-Hunt et al. 2003). At the same time, status, or reputation is also one of the most important motivations for individuals to contribute to online communities (Roberts et al. 2006). Others like Kumar and Sun (2009) and Lu et al. (2010) empirically model various kinds of inter-temporal tradeoffs that contributors to online communities have to make. Forman et al. (2008) and Moe and Schweidel (2010) examine how previously posted opinions in a ratings environment may affect a subsequent individual's posting behavior.

Finally, our paper is related to the literature in dynamic structural models. The dynamic game model, in which multiple agents make decisions simultaneously and the utility each agent gets depends on others decisions, is one specific type of dynamic structural model. It has been widely adopted applied in industrial organization research (for example, Pakes and McGuire 1994, Ericson and Pakes 1995, Bajari et al. 2007, Aguirregabiria and Mira 2007, Aguirregabiria and Ho 2009) and marketing (for example, Dube et al. 2005, Sweeting 2007, Ryan and Tucker 2008, Chung et al. 2009, Misra and Nair 2009, Kumar and Sun 2009, Ching 2010, Lu et al. 2010).[3] Of these papers, the most closely related to our work are Kumar and Sun (2009) who use a dynamic game to examine why users contribute to connected goods in social networking sites and Lu et al. (2010) who study how the social structure of individuals on a social media platform affects their willingness to share knowledge with peers. However, none of these papers examine the implications of a firm's adoption of enterprise social media on internal employee behavior nor do they examine on their incentives for creating and consuming content internally within a firm.

3. Model

3.1 Per-Period Utility

Employees *i*=1,...., *I* decide about blogging decisions on a periodic basis for time periods *t*=1,..., *T*. In enterprise blogging there are two types of posts that employees can generate while blogging represented by *j*={*w*, *l*} where *w* represents work-related posts and *l* represents leisure-related posts. In each period, an employee decides whether to read (or post a blog of type *j*). We use *p* to indicate 'post' and *r* to indicate 'read'. In other words, the action that an employee takes in each period is composed of four binary elements, i.e. $d_{itwp}, d_{itlp}, d_{itwr}$ and d_{itlr}, where $d_{itjp(r)}$ is an indicator variable which equals 1 if employee *i* posts (reads) a type *j* post at time *t*. Hence, in total, there are sixteen possible combinations of choices an employee can make. For notational convenience, we convert the four-dimensional action space to a one-dimensional action space A_i, which is defined as $A_i = \{0,1,2, \dots 15\}$, a finite set of sixteen elements. In every period, every employee chooses an action $a_{it} \in \boldsymbol{A_i}$. And $\boldsymbol{a_t} = (a_{1t}, \dots, a_{It})$ denotes the set of actions that all employees choose at time *t*.

[3] For a complete review of the literature, see Dube et al. (2005).

Note that each value of a_{it} is associated with only one combination of the four activities. For instance, $a_{it} = 0$ corresponds to the situation where $d_{itwp} = 0, d_{itlp} = 0, d_{itwr} = 0$ and $d_{itlr} = 0$; $a_{it} = 1$ indicates the situation where $d_{itwp} = 0, d_{itlp} = 0, d_{itwr} = 0$ and $d_{itlr} = 1$, etc. In other words, knowing a_{it} is equivalent to knowing $(d_{itwp}, d_{itlp}, d_{itwr}, d_{itlr})$.

We assume that an employee's per period utility function at time *t* comprises of utility from reputation (denoted by *R*), knowledge (denoted by *K*), an unobserved private shock, and everything else in the shape of an outside good. An employee's utility at time t, U_{it}, is given by:

$$U_{it} = \omega_{it}(\theta_1, \theta_2, R_{itw}, R_{itl}) + \tau_{it}(\theta_3, \theta_4, K_{itw}, K_{itl}) + O_{it} + \gamma_{it}(a_{it}). \tag{1}$$

Here ω_{it} denotes reputation-based utility and τ_{it} denotes knowledge-based utility. R_{itj} is the discounted cumulated readership employee *i* receives from type *j* posts up until period *t*. K_{itj} is the knowledge level of type *j* for an employee *i* at the end of time period t. O_{it} is the consumption of outside goods, with the utility derived from per unit consumption of outside good normalized to one. $\gamma_{it}(a_{it})$ is the action specific random shock associated to the utility that may affect an employee's decisions. Before choosing his/her actions, employee *i* receives a vector of choice-specific shocks, $\gamma_{it} = (\gamma_{it}(0), \gamma_{it}(1), \dots, \gamma_{it}(15))$. Each element in γ_{it} has type one extreme value distribution and is *i.i.d.* across individuals and actions. When employee *i* chooses an action a_{it}, the choice-specific shock associated with this particular action, i.e. $\gamma_{it}(a_{it})$, is realized and goes into individual's current period utility. The first four elements in the utility function are further explained and operationalized later. A summary of all notations and variables is given in Table 1.

Table 1: Summary of Notations

Variable Meaning and Corresponding Parameters in Parenthesis	
$log(R_{itw})(\theta_1)$	Natural log of cumulative reputation (measured by depreciated past readership and current period readership) of work-related posts for employee *i* in period *t*.
$log(R_{itl})(\theta_2)$	Natural log of cumulative reputation (measured through depreciated past readership and current period readership) of leisure-related posts for employee *i* in period *t*.
$K_{itw}(\theta_3)$	Work-related knowledge state (measured in levels. The number of levels specified after HMM estimation)
$K_{itl}(\theta_4)$	Leisure-related knowledge state (measured in levels. The number of levels

	specified after HMM estimation)
$d_{itwp}(\theta_5)$	Work-related posting decision (Average cost of posting work-related posts in terms of time)
$d_{itlp}(\theta_6)$	Leisure-related posting decision (Average cost of posting leisure-related posts in terms of time)
$d_{itwr}(\theta_7)$	Work-related reading decision (Average cost of reading work-related posts in terms of time)
$d_{itlr}(\theta_8)$	Leisure-related reading decision (Average cost of reading leisure-related posts in terms of time)
	Other notations
i, j,t	Indices of employee, post types and periods (week)
$d_{itjp(r)}$	Binary variable with 1 denoting employee *i* post (read) type *j* posts in period *t* and 0 otherwise
a_{it}, s_{it}	Employee *i*'s action and states in period *t*
$\boldsymbol{a_t}, \boldsymbol{s_t}$	Vectors of actions and states of all employees in period *t*
$\gamma_{it}(a_{it})$	Employee *i*'s choice specific private shock in period *t*
ω_{it}	Employee *i*'s reputation-based utility in period *t*
τ_{it}	Employee *i*'s knowledge-based utility in period *t*
$t_{jp(r)}$	Time cost of post (read) type *j* posts each period
O_{it}, t_o	Consumption of outside goods and its associated time cost.
y_{it}	Budget constraint assigned to individual *i* in period *t*
r_{itj}	New readership employee *i* receives in period *t* from type *j* posting
β, δ	Discount factors in life time utility function and depreciation factor in accumulation of the readership

3.1.1 Reputation

Prior work has shown that those individuals who are identified as possessing expertise are often afforded power and status within the organization (French and Raven 1959). Hence, expertise sharing can produce significant personal benefit in terms of increase in reputation within the organization (Constant et al. 1994, Thomas-Hunt et al. 2003, Lu et al. 2010). These benefits are applicable in blog settings as well, since intuitively, when employees blog they are sharing their expertise with others. Employees derive reputation benefits from the posts they write based on their area of expertise (Nardi et al 2004). We assume that an employee derives reputational benefit consistent with her readership. In the utility function, the reputation-based utility is incorporated as ω_{it} where

$$\omega_{it} = \theta_1 \log(R_{itw}) + \theta_2 \log(R_{itl}) \qquad (2)$$

We argue that employees derive reputation-based utility from their cumulative readership (accumulated over time), and not just the contemporaneous readership (readership they receive in the current period). This is because employees' readership in previous periods can be carried on to the next period with a certain discount rate. Even if one does not write any posts in the current period, he/she can still enjoy the benefits from the reputation he/she has built in the past. We apply natural log transformation here to adjust the over dispersion of the cumulative readership. It also results in a concave relation between reputation based utility and cumulative readership, which makes sense because one additional unit of readership does not provide as much additional utility for those who have very high cumulative readership than for those who have low cumulative readership.

We allow for the work and leisure reputations to enter separately into the model as they may lead to different kinds and levels of incentives. Work-related posts may express a user's expertise in work-related knowledge, which may help an employee derive indirect/direct economic incentives within the enterprise (McAfee 2006). Leisure-related posts may benefit the individual from developing a following and becoming more popular among the employee who read his/her posts. Higher reputation on posts may also help in ego-boosting or gaining higher self satisfaction (Nardi et al. 2004).

3.1.2 Knowledge

Blogs facilitate access to tacit knowledge and resources vetted by experts. The primary reason why corporations allow their employees to participate in blogging activities during their work hours is because the employee blogs act as a new source of work relevant knowledge sharing within the enterprise (Huh et al. 2007; Lee et al. 2006; Singh et al. 2010a; Yardi et al. 2009).

Employees can acquire knowledge by reading other's posts. When employees read others work posts they can become more productive, more informed about new ideas, and more aware of their (blogger) expertise, which opens up new opportunities for collaborations (Singh et al

2010). By reading other's posts, the readers also learn new ideas which open up opportunities for them to express their opinions in future posts. Leisure posts can help the reader relax and refresh. Further, individuals have an inherent need for leisure which both leisure reading and posting provide. As in the case of reputation, the two types of posts affect the type and level of incentives differently. Hence, we incorporate them separately. In the utility function, the knowledge-based utility is captured by τ_{it} where

$$\tau_{it} = \theta_3 K_{itw} + \theta_4 K_{itl} \tag{3}$$

3.1.3 Budget Constraint

Every week an employee has an exogenous time budget (y_{it}), to do all kinds of activities. Let t_{jr} be the cost of identifying and reading a type j post and t_{jp} be the cost of developing and writing a type j post. Then we have the following budget constraint:

$$y_{it} = \sum_j t_{jp} d_{itjp} + \sum_j t_{jr} d_{itjr} + t_o O_{it}.$$

Here, O_{it} is the outside good consumption and t_o is the associated coefficients that capture the per unit time cost of consuming outside good. This budget constraint allows us to capture the tradeoff that an individual would consider while deciding time to allocate to work, leisure activities, and non-blogging activities. Let us define $\theta_5 = t_{wp}/t_0\,; \theta_6 = t_{lp}/t_0\,; \theta_7 = t_{wr}/t_0\,; \theta_8 = t_{lr}/t_0$, which can be interpreted as the relative cost of participating in the four activity compared to doing non-blogging activities. Then the budget constraint can be written as:

$$\frac{y_{it}}{t_o} = \theta_5 d_{itwp} + \theta_6 d_{itlp} + \theta_7 d_{itwr} + \theta_8 d_{itlr} + O_{it}.$$

Solving for O_{it}, we get

$$O_{it} = \frac{y_{it}}{t_o} - \theta_5 d_{itwp} - \theta_6 d_{itlp} - \theta_7 d_{itwr} - \theta_8 d_{itlr} \tag{4}$$

Substituting Equations (2), (3) and (4) into the utility function (1) gives:

$$U_{it} = \theta_1 R_{itw} + \theta_2 R_{itl} + \theta_3 K_{itw} + \theta_4 K_{itl} + \frac{y_{it}}{t_o}$$
$$- \theta_5 d_{itwp} - \theta_6 d_{itlp} - \theta_7 d_{itwr} - \theta_8 d_{itlr} + \gamma_{it}(a_{it})$$

Since, y_{it}/t_o affects all choices in the same way, we drop it from the utility function and rewrite the utility function as

$$U_{it} = \theta_1 \log(R_{itw}) + \theta_2 \log(R_{itl}) + \theta_3 K_{itw} + \theta_4 K_{itl}$$
$$-\theta_5 d_{itwp} - \theta_6 d_{itlp} - \theta_7 d_{itwr} - \theta_8 d_{itlr} + \gamma_{it}(a_{it}) \qquad (5)$$

3.2 Dynamic Game

3.2.1 State Variables

Note that the first four elements in Equation (5) are all state variables that evolve according to the actions employees take in each period. The reputation states evolve as follows:

$$R_{itj} = \delta R_{it-1j} + d_{itjp} r_{itj} \qquad (6)$$

Here, r_{itj} is the number of people who read *i*'s type j post in period t. In addition, δ is a depreciation factor which is set at 0.9. This implies that the contribution of past readership to current reputation declines as time passes by. The individual will have to continue posting to maintain a high reputation.

The third and fourth state variables are the knowledge of employee *i*, K_{itw} and K_{itl}. While, we observe the reading behavior among employees, the evolution of individuals' knowledge states is unobserved to us. Hence, we follow Arcidiacono and Miller (2006) and employ a hidden Markov model (HMM) framework to identify the unobserved knowledge states and their evolution. An HMM is a model of a stochastic process that cannot be observed directly but can only be viewed through another set of stochastic processes that produce a set of observations (Rabiner 1989). We treat the knowledge evolution to be an unobserved stochastic process. Further, we treat the combined four actions to be the state-dependent observed process. Specifically, we consider knowledge to be a discrete unobserved state and assume that knowledge evolution follows a Markov process. This Markovian transition of knowledge states is consistent with literature on knowledge evolution (Singh et al 2010b). We further enforce that the probability of transitioning to a high knowledge state is higher if an individual reads a post than otherwise. The two knowledge state variables are assumed to evolve independent of each other.

We define s_{it} as the set of the state variables for an employee at period t. Then $s_{it} = (R_{itwp}, R_{itlp}, K_{itw}, K_{itl})$ is set of values of the state variables for an employee i at time t. We further define $s_{-it} = (R_{-itw}, R_{-itl}, K_{-itw}, K_{-itl})$ as the set of state variables of *i*'s peers. Then the strategy profile for *i* depends on $s_t = (s_{it}, s_{-it})$. [4]

3.2.2 Sequence of Events

The specific sequence of events in our model is as follows:

1. Employees observe their states s_{it} and their peers' states s_{-it}.
2. Employees receive a set of choice-specific random shocks (γ_{it}).
3. Employees make expectation of the readership (r_{itj}) they may get in the current period if they write a post.
4. The employee probability of moving up a knowledge state is just a function of his/her current knowledge state and reading action. The employee has perfect information about the knowledge state transition probabilities.
5. Employees make decisions as to what to write or read in the current period.
6. Employees receive utility because of their decisions.
7. Employee states evolve to s_{it+1} because of their and their peers' decisions.

3.2.3 Long Term Utility Function

We model the employee's writing and reading decisions as a dynamic optimization problem. The employee's tasks are to decide when and whether to read work post, read leisure post, write work post, and write leisure post to maximize the sum of the discounted expected future utility U_{it} over the infinite horizon.

---

[4] One notable point about this strategy profile is that ideally, s_t represents the states of all employees. In our case, however, there are 2396 employees. Tracking every employee's state would make our model intractable To deal with this problem, we introduce a simplifying but more realistic assumption that employees make their decisions in accordance with their own state and the average state of other employees (Aguirregabiria and Ho 2010). This is a reasonable assumption because in reality, it is infeasible for employees to track all other employees' states and make decisions accordingly. Instead, employees may only want to get a general idea about what the other employees' states are and make decision based on his/her own state and the moments of others' states. In other words, strategy is a function of employees' own states and the moments of states of the competing group.

$$max_{a_{it}} E(\sum_{t=1}^{\infty} \beta^t U_{it}(a_{it}, s_t, \gamma_{it}) | s_t, \gamma_{it})) \tag{7}$$

The variable β is the common discount factor. The operator $E_t[\cdot]$ denotes the conditional expectation operator given the employee's information at time t. There are three components of the model that need to be emphasized. The employee in our model maximizes his life time utility, which makes the model dynamic. Since, the present actions and state of an employee affect his/her future utility by affecting the states, the employee is forward looking. The employee balances his/her time between reading and writing, and work and leisure due to the budget constraint. Hence, his/her actions are interdependent. At the same time, the utility of an employee is a function of the decisions made by his/her peers (through r_{itj}) making it a multi-agent dynamic game. As will be explained later, the readership that the employee gets for his/her posts is not only a function of his/her states but also a function of his/her peers' states.

3.2.4 Equilibrium concept

Following Ericson and Pakes (1995) we focus on Markov Perfect Equilibrium (MPE) as a solution concept. We assume that the behavior is consistent with MPE. In an MPE, each employee's behavior depends only on the current states and his current private shock. Formally, a Markov strategy for an employee i is a function $\sigma_i: S \times \Gamma_i \rightarrow A_i$. A profile of Markov strategies is a vector, $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_I)$, where $\boldsymbol{\sigma}: S \times \Gamma_1 \times \dots . \times \Gamma_k \rightarrow A$. Here, we drop the time index because the strategy profile is time invariant. If behavior is driven by a Markov strategy profile $\boldsymbol{\sigma}$, employee i's expected utility given state **s** can be written recursively as Bellman Equation:

$$V_i(\boldsymbol{s}; \boldsymbol{\sigma}) = E_\gamma \left[U_i(\boldsymbol{\sigma}(\boldsymbol{s}, \boldsymbol{\gamma}), \boldsymbol{s}, \gamma_i) + \beta \int V_i(\boldsymbol{s}'; \boldsymbol{\sigma}) dP(\boldsymbol{s}' | \boldsymbol{\sigma}(\boldsymbol{s}, \boldsymbol{\gamma}), \boldsymbol{s}) \middle| \boldsymbol{s} \right].$$

Here, V_i is a value function which reflects the expected value for employee i at the beginning of a period before private shocks are realized. Following the literature, a profile $\boldsymbol{\sigma}$ is a Markov perfect equilibrium if, given the opponent profile $\boldsymbol{\sigma}_{-i}$, each employee i prefers its strategy $\boldsymbol{\sigma}_i$ to all alternate strategies $\boldsymbol{\sigma}_i{}'$. That is, for $\boldsymbol{\sigma}$ to be MPE

$$V_i(\boldsymbol{s}; \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_{-i}) \geq V_i(\boldsymbol{s}; \boldsymbol{\sigma}_i', \boldsymbol{\sigma}_{-i}).$$

3.3 Identification and Normalization

There are several identification issues that need to be addressed before the model can be consistently estimated. First, the distribution of private shocks is assumed to be known for identification. We assume that the private shocks are extreme value type 1 distributed. Second, in the HMM the probability of knowledge increase should be higher for an individual as a result of reading than otherwise. We enforce this by assuming that reading has a non-negative impact on knowledge increase. Third, we cannot identify t_o and the cost-specific parameters together. Hence, we normalize t_o =1. Fourth, we assume that the knowledge state of an employee is a private state and make an exclusion restriction. An employee's actions, while depend on his/her own knowledge state, are independent of his/her peer's knowledge states. Such exclusion restrictions are commonly imposed in structural models (Bajari et al 2009).

We do not know the initial values of the state variables, which raise the well known “initial conditions” problem. The first observation in our sample may not be the true initial outcome of an employee's blog content generation and reading behavior. If one does not control for initial choice history, the implicit assumption is that every employee has the same beliefs across both kinds of activity (read and write) and across both kinds of blogs (work and leisure). However, it is possible that a employee that has engaged in an activity multiple times in the past would have more informed priors than another user who has engaged very little in that activity. Hence, one needs to account for the heterogeneity of priors in the sample. We follow an approach that is similar in spirit to that used in Erdem et al. (2008). We use a part of our data (16 weeks) as the pre-estimation sample to estimate the distribution of the state variables.

4. Empirical Estimation

4.1 Data

Our data comes from a large, Fortune 1000 information technology services, business process outsourcing, and consulting firm. Fortune named this firm one of the fastest growing companies in 2009 (Fortune 2009). Its annual revenues in year 2009 were a few billion dollars. It is a US- based firm with significant presence and operations in several other countries across multiple continents: Europe, Asia, and Americas being the major areas. To influence more knowledge and information sharing across as well as within locations, the firm has undertaken several measures. Prominent among these measures is the use of Web 2.0 technologies such as blogs for use within the enterprise.

These blogs are hosted on an internal platform and are not accessible by people outside the organization. Every employee is allowed to host his/her own blog on this platform. These blogs are accessible to all the employees of the firm across the entire hierarchy. The identity of the blogger is also revealed on the blog. Bloggers classify their posts into one of several categories (for example, software testing, movies, history, knowledge management, senior management, etc). To be able to measure the knowledge sharing aspect, the firm tracks who (which employee) reads which blog and at what time. Since the blogs are only internally accessible, the firm did not impose any restrictions on the kind of posts that can be written by the employees. To analyze the type of content that is being shared in the internal blogosphere, the firm broadly classifies the blog article categories into two topics: Work-related (*w*) and Leisure-related (*l*). Table 2 presents the sub-categories that constitute each topic.

Table 2: Blog Post Classification

Topics	**Leisure Related**	**Work Related**
Sub Categories	Fun; Movies-TV-Music; Sports; Puzzles; Chip-n-putt; Religion-Spiritual-Culture; History-Culture; Photography; Arts; Poetry-Stories; Books; Geographies	FLOSS; Technology; Testing; Domains; Corporate Functions; Knowledge Management; Project Management; Business Development; Senior Management; Practices-Programs-Accounts

Our data consists of blogging activities of 2396 employees over a 15 month (64 weeks) period. We have data on exact timestamps of blog reading and posting activities. For the purpose of estimation, we define a period as one week. This provides us data for 64 periods in total. We treat the first 16 weeks as the hold-out period, and estimate the model using data from the remaining 48 weeks.

High-level descriptive statistics of our data are shown in Table 3. These employees wrote 26075 posts during the 64 week period, indicating that not every employee generates a blog posting in every period. Out of these 9967 posts are work-related and 16108 posts are leisure related. There were 26831 readings of work posts and 37265 readings of leisure posts by these employees during the 64 week period.

Table 3: Overall Sample Statistics

Overall Statistics	
Number of Employees	2396
Period length	1 week
Number of periods	64
Total posts written	26075
Work-related posts written	9967
Leisure-related posts written	16108
Work-related post reading	26831
Leisure-related post reading	37265

The descriptive statistics for the key variables used to construct the model variables are presented in Table 4. Both reading and writing exhibit a long tail distribution where a few employees read and write a lot while the majority of others participate little. On average, every week, 6% of employees post a work-related blog, 11% employees post a leisure-related blogs, 17% read work-related blogs and 24% read leisure–related blogs. A given work-related post by an employee is read by 40.03 employees on an average. A given leisure-related post by an employee is read by 107.48 employees on an average.

Table 4: Descriptive Statistics of Key Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
R_{itw}	16.28	91.34	0	4043.83
R_{itl}	86.19	305.22	0	7438.58
K_{itw}	0.12	0.33	0	1
K_{itl}	0.16	0.37	0	1
d_{itwp}	0.06	0.18	0	1
d_{itlp}	0.11	0.26	0	1
d_{itwr}	0.17	0.38	0	1
d_{itlr}	0.24	0.43	0	1

4.2 Estimation Strategy

The model parameters that need to be estimated are: $\boldsymbol{\rho} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8)$. To estimate the structural model, we follow the two-stage estimation procedure suggested by Bajari et al. (2007) (BBL hereafter) for this scenario. There are several other estimation procedures suggested by for example Pakes et al. (2007), Pesendorfer et al. (2003), and Aguirregabiria and Mira (2007). However, all of them apply only to the case of discrete state dynamic games. However, the BBL estimator applies to both the discrete and continuous state cases. We combine the BBL estimator with the Bajari et al. (2009) and Arcidiacono and Miller (2006) to estimate the model.

In the first stage, we estimate the policy functions ($\sigma_i: S \times \Gamma_i \to A_i$), the state transition probabilities $P: A \times S \to \Delta S$, and the value functions. Our choice variables are discrete. And the utility function defined above implies additive separability and conditional independence between U_i and Υ_i. Define the choice-specific value function as

$$v_i(a_i, \boldsymbol{s}) = E_{\Upsilon_{-i}}[U_i(a_i, \boldsymbol{\sigma}_{-i}(\boldsymbol{s}_{-i}, \boldsymbol{\Upsilon}_{-i}), s_i) + \beta \int V_i\left(s', \boldsymbol{\sigma}\right) dP\left(s_i' | a_i, \boldsymbol{\sigma}_{-i}(\boldsymbol{s}_{-i}, \boldsymbol{\Upsilon}_{-i}), \boldsymbol{s}\right)\Big)]$$

With this notation, action a_i is employee i's optimal choice when

$$v_i(a_i, \boldsymbol{s}) + \gamma_i(a_i) \geq v_i\left(a_i', \boldsymbol{s}\right) + \gamma_i\left(a_i'\right), \quad \forall a_i' \in A_i$$

Hotz and Miller (1993) showed how to recover the choice-specific value functions by inverting the observed choice probabilities at each state:

$$v_i\left(a_i', \boldsymbol{s}\right) - v_i(a_i, \boldsymbol{s}) = ln(Pr(a_i' | \boldsymbol{s})) - ln(Pr(a_i | \boldsymbol{s}))$$

Given this, we only need to estimate the probability distribution of actions at each state, i.e. conditional choice probability (CCP), from the data. Note that we face two challenges: (1) reputation states are continuous; and (2) knowledge states are unobserved. As discussed earlier, we follow Arcidiacono and Miller (2006) and employ an HMM to identify the knowledge states, the Markovian transition probabilities of the knowledge states as a function of actions and the most probable knowledge state for each individual at every time point. However, several of state variables are continuous (reputation states and the average states of peers). Therefore, $\Pr(a_i | s_i)$ will be a function, rather than discrete values. Following Bajari et al. (2009), we use the sieve logit method to estimate CCP, with a second degree orthogonal polynomial as bases.

The estimation strategy suggested by BBL is inspired by Hotz and Miller (1993). The estimation process uses forward simulation to estimate an employee's value functions for a given strategy profile (including the equilibrium profile) given an estimate of state transition probabilities. Let $V_i(\boldsymbol{s};\boldsymbol{\sigma},\boldsymbol{\rho})$ denote the value function of employee i at state $\mathbf{s}$, assuming i follows Markov strategy $\boldsymbol{\sigma}_i$ and all its peers follow $\boldsymbol{\sigma}_{-i}$. Then

$$V_i(\boldsymbol{s};\boldsymbol{\sigma};\boldsymbol{\rho}) = E\left[\sum_{t=0}^{\infty}\beta^t U_{it}(\boldsymbol{\sigma}(\boldsymbol{s_t},\boldsymbol{\gamma_{it}}),\boldsymbol{s_t},\gamma_{it};\boldsymbol{\rho})\middle|\boldsymbol{s_0}=\boldsymbol{s};\boldsymbol{\rho}\right].$$

If the policy profile used in this step is the policy profile estimated in the first stage then its the resultant value over a number of simulated paths is an estimate of the payoff ($\hat{V}_i(\boldsymbol{s};\hat{\sigma}_i,\hat{\boldsymbol{\sigma}}_{-i};\boldsymbol{\rho})$) from playing $\hat{\sigma}_i$ in response to all peers playing $\hat{\sigma}_{-i}$.

In the second stage, we use the estimates from first stage combined with the equilibrium conditions of the model to estimate the underlying structural parameters.

Let $z \in Z$ index the equilibrium conditions, so that each z denotes a particular $(\boldsymbol{i},\boldsymbol{s};\boldsymbol{\sigma'}_{\boldsymbol{i}})$ combination. Let us further define:

$$g(z;\boldsymbol{\rho},\aleph) = V_i(\boldsymbol{s};\boldsymbol{\sigma_i},\boldsymbol{\sigma}_{-i};\boldsymbol{\rho};\aleph) - V_i(\boldsymbol{s};\boldsymbol{\sigma'}_{\boldsymbol{i}},\boldsymbol{\sigma}_{-i};\boldsymbol{\rho};\aleph) \qquad (8)$$

Here, $\aleph$ reflects that $\boldsymbol{\sigma}$ and state transitions are parameterized by $\aleph$. The equilibrium condition is satisfied at $\boldsymbol{\rho},z$ if $g(z;\boldsymbol{\rho},\aleph) \geq 0$. This is estimated through simulated minimum distance estimator.

4.3 Equilibrium Behavior

4.3.1 Work Posting vs. Leisure Posting

As discussed earlier, organizations adopt Enterprise 2.0 systems to facilitate knowledge sharing by employees. Firms typically want to encourage work posting and discourage leisure posting. Hence, it becomes interesting to investigate how the decisions to engage in work or leisure posting are affected by the reputation and knowledge of an employee in both work and leisure.

Spillover from Leisure to Work and from Reading to Posting

The key to understand the relationship between work and leisure posting is to first investigate how work and leisure reputations affect the readership of an employee's new post. In

Figure 1 (Figure 2), we plot how the readership one can gain from new work (leisure) post varies with his/her reputation states. In constructing these figures we assume all other state variables are at their mean values. The figures highlight several interesting insights which we discuss briefly here.

First, the readership for a new post that an employee may receive increases with his/her reputation. On average, the log readership that a new work post may receive increases from 2.7 to 4.6 (70%) as work reputation increases from 0 to 10 and the leisure reputation is at its minimum value of 0. In the same way, on average, the log readership that a new leisure post may receive increases from 3.5 to 4.6 (31%) as leisure reputation increases from 0 to 10 and the work reputation is at its minimum value of 0. This finding highlights the presence of a strong "rich get richer" effect. Once an employee builds up reputation, his/her new posts will attract more readers. Interestingly, we also observe that readership of new work-related posting by a given employee increases with that employee's leisure reputation. Leisure reputation could impact work readership in two countervailing ways. First, a blogger with a high leisure reputation attracts people who are interested in leisure-related posts, and in the process of doing so, they can get exposed to the work-related posts by this blogger. This leads to a positive spillover effect from leisure reputation to work readership. Second, there could be a negative signaling effect. A high leisure reputation could be harmful, because this might signal lack of commitment to work-related blogging. Thus, potential readers will have a lower expectation of the quality of their work-related blogs.

From Figure 1, we can see that when an individual's work reputation is low, the positive spillover effect is the dominant effect, and hence leisure reputation has a positive interdependence with readership of a new work-related post. As employee's work readership goes up, the negative signaling effect dominates the positive spillover effect, and hence in this region, leisure readership negatively affects readership of a new work-related posting. On average, as leisure reputation of the blogger increases from 0 to 10, the log readership that a new work post may receive increases (decreases) from 2.7 to 3.7 (37%) when work reputation is 0. In contrast, on average, as leisure reputation of the blogger increases from 0 to 10, the log readership that a new work post may receive decreases from 3.3 to 2.6 (21%) when work reputation is 10. In comparison, if we look at the relation between new leisure readership and reputation states (Figure 2), work reputation seems to have no impact on readership of a new leisure-related post. It seems that the two countervailing effects, positive spillover and negative signaling cancel out each other. In other words, reputation on work related posts would have no effect on a blogger's ability to attract readers to his/her leisure related posts.

Comparing the two planes in the Figure 1, we can see that work-related posts written by people with high work knowledge attract more readers than those written by people with low work knowledge. This relationship also holds for leisure-related posts and the leisure knowledge of their authors as shown in Figure 2. These two relationships reveal the interdependence between posting and reading. Employees in high work (leisure) knowledge states are able to write higher quality posts, which are more attractive to readers.

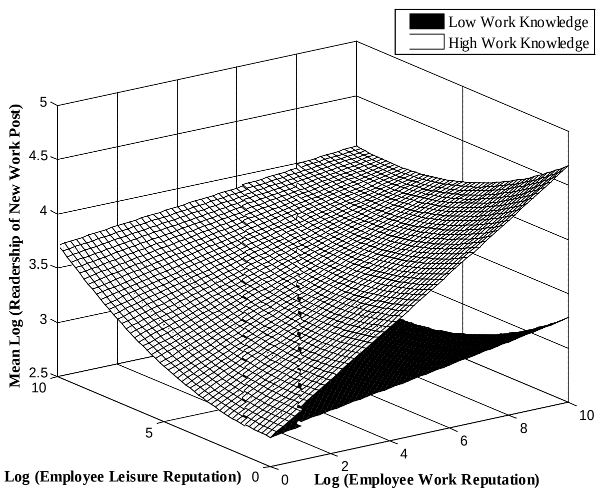

Figure 1: Readership for New Work Post

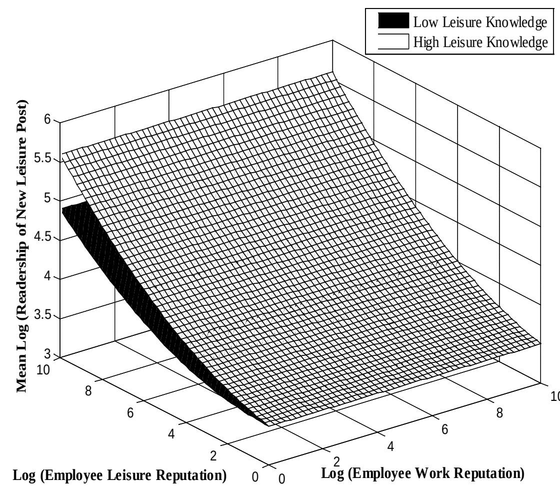

Figure 2: Readership for New Leisure Posts

Decisions on Work Posting and Leisure Posting

Figure 3 (Figure 4) plots the probability of work (leisure) posting for an employee as a function of his/her both work and leisure reputation and work (leisure) knowledge state. There are several interesting findings in these figures.

First, Figure 3 shows that at a given level of leisure reputation, employees are more likely to post work-related blogs when their work reputation is high than otherwise. This is because individuals are forward-looking and expect higher readership when they have higher levels of reputation. The same reasoning also applies to probability of leisure posting (Figure 4). Further, leisure posting always increases with leisure reputation, thereby confirming the rich get richer effect. However, leisure posting increases when work reputation is low but decreases when work reputation is high. This shows the interdependence among work and leisure which was also highlighted by the readership spillover. Basically, what we find is that while work reputation does not affect the readership of leisure posts, in contrast, leisure reputation increases work readership when work reputation is low and decreases work readership when work reputation is high. While leisure readership always increases with leisure reputation, the marginal benefit from new leisure readership when work reputation is high may not be sufficient to outweigh the negative effect of loss of new work readership. Further, more knowledgeable employees have a higher probability of posting. This finding is again consistent with the earlier discussion where more knowledgeable employees receive higher readership.

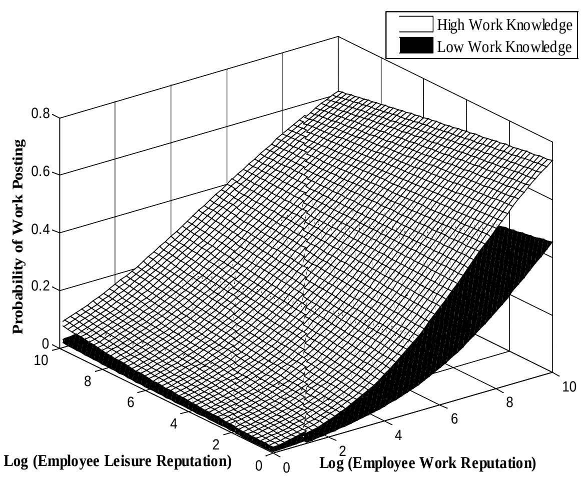

Figure 3: Probability of Work Posting

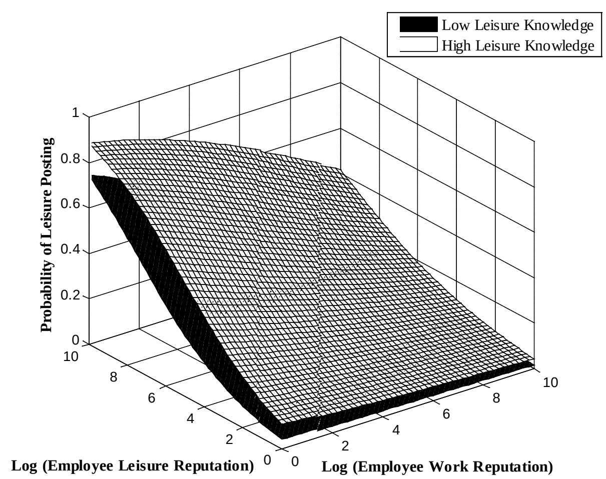

Figure 4: Probability of Leisure Posting

4.3.2 Knowledge States and Reading Decisions

Employees may also face a trade-off between work and leisure when choosing which blogs to read. And again, firms typically would want their employees to read work-related blogs. Hence, a good understanding of how employees' knowledge states would affect their reading decisions is also important. Using the HMM, we find that the model with two knowledge levels for each type (leisure and work) fits our data best. We define the two levels as "high knowledge" and "low knowledge". The choice-specific state transition probabilities of the two types of knowledge states are given in Table 5 and Table 6. Note that here we assume that an employee's work (leisure) knowledge state is only affected by his/her work (leisure) reading decisions. There are several interesting observations in the state transition probabilities.

First, an employee has to continuously update himself of the new trends, technologies, issues, etc to sustain himself in a high knowledge state. There are several reasons as to why an individual moves down a state if he/she does not participate in reading. The primary reason is that given the high tech industry context of this research setting, the technologies, algorithms, new ways to solve problems evolve very quickly leaving the current knowledge of an individual outdated very soon. Further, individuals can also forget what they know over time which is documented across several studies on individual learning behavior (Argote et al 2003). Second, we see that it is harder to get to a high work knowledge state in comparison to leisure knowledge state from a low value of the corresponding state. Third, once an individual is in a high state it is easier for him/her to stay in the high leisure knowledge state than in the high work knowledge state. The matrices indicate that it is much harder to move up in work knowledge state than in leisure knowledge from blog reading.

Table 5: Work Knowledge State Transition Probabilities

t\t+1	Low	High
Low	0.99 (0.89)	0.01 (0.11)
High	0.63 (0.19)	0.37 (0.81)

The probabilities outside (inside) parenthesis indicate the probabilities without reading (without reading).

Table 6: Leisure Knowledge State Transition Probabilities

t\t+1	Low	High
Low	0.88 (0.77)	0.12 (0.23)
High	0.52 (0.05)	0.48 (0.95)

The probabilities outside (inside) parenthesis indicate the probabilities without reading (without reading).

Given the choice-specific state transition probability matrices, we next discuss employees reading decisions. We notice that there is little interdependence between work knowledge and leisure knowledge. Probability of work reading is only affected by employees' work knowledge level, but not by leisure knowledge level, and vice versa. Here, we use two sets of comparison to summarize the relation between current knowledge levels and the reading decisions. Generally speaking, when we fix the work and leisure reputation status, and control for the knowledge level of other type, people with high work (leisure) knowledge level have higher probability to read work- (leisure-) related posts. This is probably because it is easier for employees with high work (leisure) knowledge to maintain the high knowledge state by reading in the current period than for those with low knowledge state to move up by reading posts.

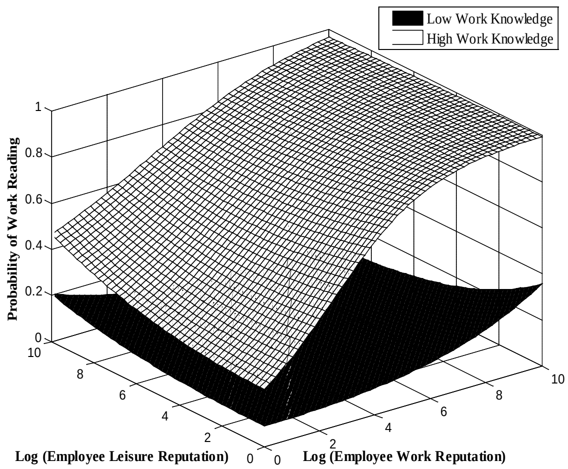

Figure 5: Probability of Work Reading

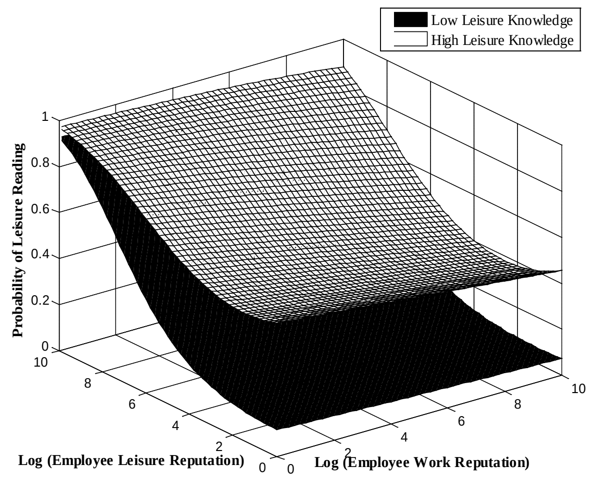

Figure 6: Probability of Leisure Reading

We also investigate how employees' reputation states affect their reading decisions (Figure 5 and Figure 6). The results show that employees with high work (leisure) readership tend to read more work- (leisure-) related posts. This occurs because as employees are forward-looking, they will consider the fact that high knowledge states can help them produce high quality posts that will attract more future readers.

4.3.3 Competitive Dynamics

As discussed in the previous section, the employees in this social media setting compete with each other to attract readership for their posts. Hence, their actions are interdependent. To illustrate the competitive dynamics, we first show, in Figure 7 and Figure 8, how the readership of a new post varies with an employee's own state and the mean state of his/her peers.

Figure 7 and Figure 8 show that the readership of a new work post increases with average work reputation of peers, whereas the readership of the new leisure post decreases with average leisure reputation of peers. These can be explained by two countervailing effects in competitive dynamics. When peers' work readership is high, the proportion of readers that an employee can

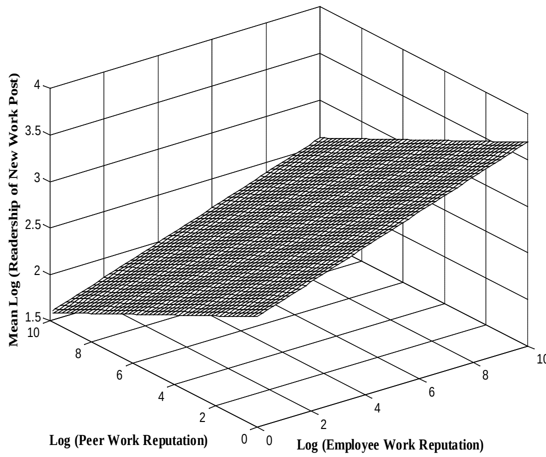

Figure 7: Readership for New Work Post

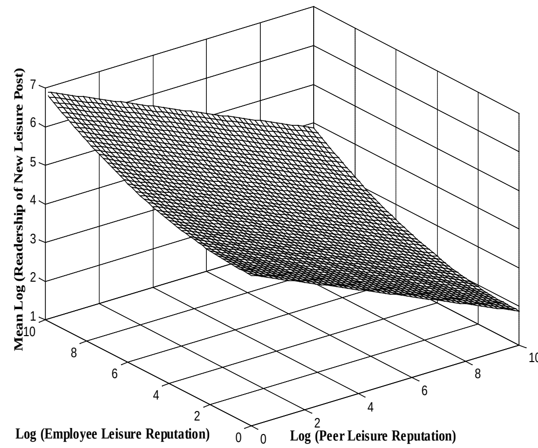

Figure 8: Readership for New Leisure Post

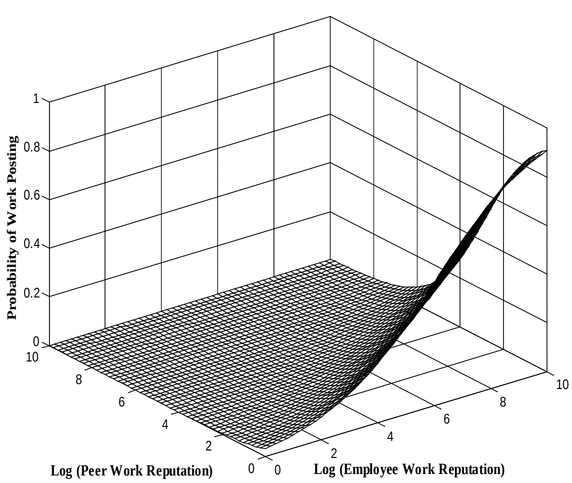

Figure 9: Probability of Work Posting

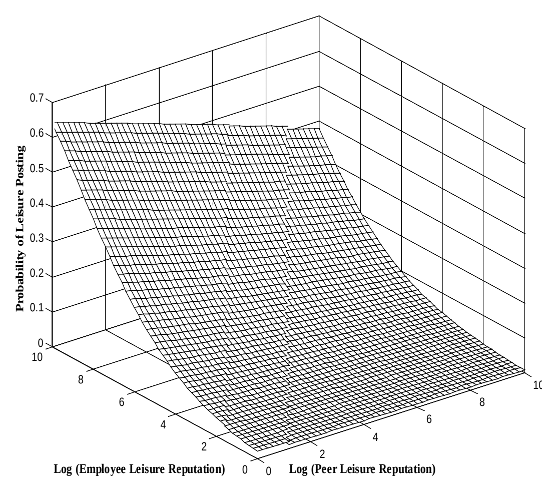

Figure 10: Probability of Leisure Posting

attract from his/her new work-related post is smaller. At the same time, employees are more likely to read work-related posts which increases the overall size of the readership pie itself. In work-related posts, the latter positive effect dominates the former negative effect, and therefore, employees get more readership as the mean work readership increases. In contrast, with leisure-related posts, the former effect dominates the latter effect, and thus employees get lower readership as the mean leisure readership increases. Figure 9 and Figure10 show the probabilities of work and leisure posting vs. peer work and leisure reputation. Consistent with the readership dynamics, we find that bloggers are more likely to post when their reputation is higher. Further, the probability of posting work (leisure) decreases (increases) with peers' work (leisure) reputation.

4.4 Second stage results

The results for the second stage (structural parameters) are presented in Table 7. For estimation purposes, the discount factor β is set at 0.95.[5] The results show that θ_1 and θ_2 are positive and significant. This indicates that employees gain positive utility from work-related reputation and leisure-related reputation. This set of results verifies that reputation gain provide employees incentives to write blogs. However, the magnitude of utility derived from work-related reputation is almost 4 times of utility derived from leisure-related reputation, indicating that in the enterprise blog environment, people appreciate work-related reputation more than leisure-related reputation.

From HMM, we know that both work knowledge and leisure knowledge have two state levels. The output of HMM also implies that employees have higher probability to stay in high level at the end of a period if they read in the current period, and have lower probability to stay in high if they do not read. The knowledge states actually count for both past knowledge base and addition knowledge acquired in current period. Individuals derive knowledge-based utility from reading other's posts as indicated by positive and significant θ_3 and θ_4.

Table 7: Structural Parameter Results

Parameter	Estimates	Standard error
θ_1	4.357***	0.254
θ_2	1.157***	0.363
θ_3	0.522***	0.027
θ_4	0.018***	0.002
θ_5	1.530***	0.054
θ_6	0.834***	0.024
θ_7	3.486***	0.169
θ_8	0.726***	0.193

***p<0.01

[5]The qualitative nature of the results are robust to several other values of the discount factor.

The parameters θ_5 to θ_8 are positive and significant. Note that in our utility function, we have a negative sign in front of these four parameters, hence, the parameters are expected to be positive. Also note that θ_5 to θ_8 are normalized costs, which are relative measure compared to the cost of consuming outside goods. Hence, the costs can be compared with each other. Reading and writing work-related post is more time consuming than reading and writing leisure posts. One possible explanation could be people who read or write work-related blog spend a lot of time digesting or creating the knowledge he/she learns from the blogs.

4.5 Model Performance

To evaluate the overall fit of the estimated model, we use the estimated parameters to simulate the blogging dynamics and then compare the simulated data moments with real data moments. The results are reported in Table 8. From the comparison we can see that simulated data moments are all reasonably consistent with the real data moments, indicating that our model does a good job in fitting the observed data.

Table 8: Model Fit

Key Statistics	Real Data Moments	Simulated Data Moments
Average Work Reputation	16.28	16.89
Average Leisure Reputation	86.19	87.93
Average Work Knowledge	0.12	0.12
Average Leisure Knowledge	0.16	0.15
Probability of Work Posting	0.06	0.07
Probability of Leisure Posting	0.11	0.11
Probability of Work Reading	0.17	0.18
Probability of Leisure Reading	0.24	0.23

4.6 Robustness Checks

We performed several tests to check the robustness of our results. First, our theory and model emphasizes that there are dynamics in user behavior. While our results support this, one concern could be that it may be a consequence of our model not accounting for user-specific unobserved heterogeneity. To test the extent of this concern, we performed individual-specific fixed effects as well as random effects estimations of our CCP. We performed a Hausman test which indicated that random effect estimates are consistent and efficient (Hausman 1978). We then compared the individual-specific random effects estimation with the pooled estimation which is the CCP we used in our stage 2. We then performed a Lagrange-Multiplier test which indicated that pooled CCP is appropriate in our case (Breusch and Pagan 1980). These tests indicate that unobserved cross sectional heterogeneity does not greatly influence our results.

Second, we had to make an assumption that knowledge states are private which allowed us to construct the CCP independent of peers' knowledge states. This assumption helped to simplify our estimation significantly. It is possible that in reality peers' knowledge states may affect the CCP. Accounting for all peers' knowledge states in an HMM would make it intractable. However, we follow our earlier assumption that individuals care only about their peer's mean states and not individual states of every peer. We treated all the peers as one individual and the average knowledge state as this individual's knowledge state. Since the average knowledge state varies between zero and one, we discretized this into 20 segments.[6] We allowed the average knowledge state transitions to be affected by moments of the peers' reading activities (mean and variance). This allowed us to incorporate peer knowledge states into the CCP. We find that peer knowledge states do not significantly predict any of the choices, indicating that peer knowledge states in fact do not enter into an individuals' decision making in any significant manner.

5. Policy Experiments

Enterprise blogs are adopted primarily to promote work-relevant knowledge sharing. Hence, firms typically encourage employees to engage in work-related blogging but discourage leisure-related blogging. Firms can implement some policies to incentivize employees' blogging activities. In this section, we explore the effects of two such policy interventions: 1) How do employees respond to the policy where leisure-related blogging is prohibited? 2) How do employees adjust their blogging behavior when the budget constraint on blogging is relaxed?

[6] Note that the number of segments for discretization is chosen arbitrarily. This method can be applied to a different number of segments too.

In the first experiment, we set the cost of leisure-posting to be extremely high. One would think that this policy will eliminate leisure-blogging activities due to the existence of the budget constraint. Hence, users will switch from leisure-related activities to work related activities and both work posting and work reading will increase. However, our simulation results suggest a very different story. Figure 11 compares probability of work posting and work reading before and after the policy is implemented. As expected, the probability of work reading increases from a base level of 18% to 24%. However, the probability of work posting decreases from 7% to 2% once the policy is implemented.

This "counterintuitive" result is actually consistent with the spillover effect we discussed in the results section. Recall that, employees with low work reputation, who are the majority of users in our sample, are the most responsive to the spillover from leisure reputation to work reputation. A moderate to high level of leisure readership greatly increases the probability for this group to post work-related blogs. When leisure activities are eliminated, those who used to have low work readership but moderately high leisure readership can no longer benefit from the spillover effect. Consequently, their probability of work posting will drop significantly. Although the probability of work reading increases, which leads to a bigger readership pie shared by all the employees who post in each period, a very high fraction of the increased readership accrues to those users who have built a high work reputation (due to Matthew effect).

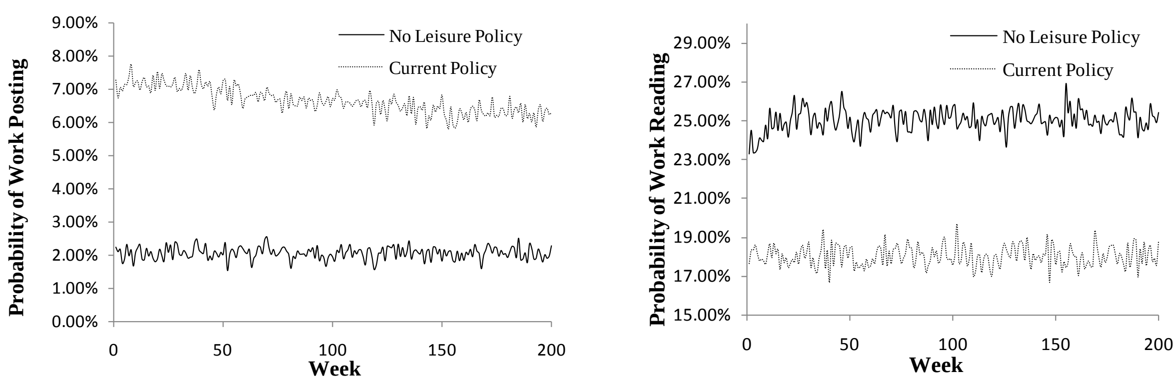

Figure 11: Comparison of Current and No Leisure Policy

Firms typically want larger and more diversified knowledge pool. We see that a policy of prohibiting leisure-related blogging can actually hurt the extent of knowledge creation and sharing within the firm. A dramatic decrease in probability of work posting means that less new knowledge will be added to the knowledge pool and fewer people will be willing to post work-related content. Even though under "No-leisure" policy, people tend to read more work-related blogs, at the firm level, the quantity and diversity of the knowledge transferred in the enterprise blog system both decrease. Thus, we see an adverse consequence of prohibiting leisure-related blogging.

In the second experiment, the firm encourages employees to blog by relaxing the budget constraint on blogging. For example, this can be operationalized by giving employees the liberty to blog during regular office hours without any penalties. High tech companies like Google are pioneers in these kinds of employee-friendly policies in the hope that this can lead to increased innovation and productivity.

Relaxing employees' budget constraint on blogging is equivalent to cutting down total costs of blogging. Here, we try two levels of cost reduction, denoted by "Medium-Budget" and "High-Budget". The results are shown in Figure 12. Under the "Medium-Budget" policy, employees only need to pay half of the total cost incurred by participating in the four blogging activities. Under the "High-Budget" policy, employees incur no cost of blogging[7]. The latter policy may not be very realistic due to its extreme nature. However, by examining this extreme case, and combining it with the case when there is a medium level of cost cutting, we can understand how users may react to different level of budget relaxations.

From Figure 12, we can see posting activities are not responsive to the "Medium-Budget" policy. When the "Medium-Budget" policy is introduced, the probabilities of work and leisure posting increase only marginally from a base level of 7% to 8% and from 11% to 12%, respectively. In contrast, the probabilities of work and leisure reading have significant increases from the base level, going from 19% to 70% and 23% to 51%, respectively. This suggests that the marginal benefit of posting is lower than the marginal benefit of reading when the budget constraint is relatively tight. When the "High-Budget" policy is introduced, the probabilities of work and leisure posting increase from a base level of 7% to 40% and from 11 to 32%,

[7] Note from Table 7 that under the current policy if an employee participates in all four activities in a period, he/she will incur a total cost of 6.576 (the sum of the mean values of parameters θ_5 to θ_8. Under the Medium-Budget policy the employee does not incur a cost up to 3.288 (half of the maximum possible cost of 6.576). For example, if in a period an employee reads a work post, he only incurs a cost of 0.198. In comparison, under the High-Budget the same individual will incur no cost.

respectively. In contrast, the probabilities of work and leisure reading have significant increases from the base level, going from 19% to 85% and 23% to 63%, respectively.

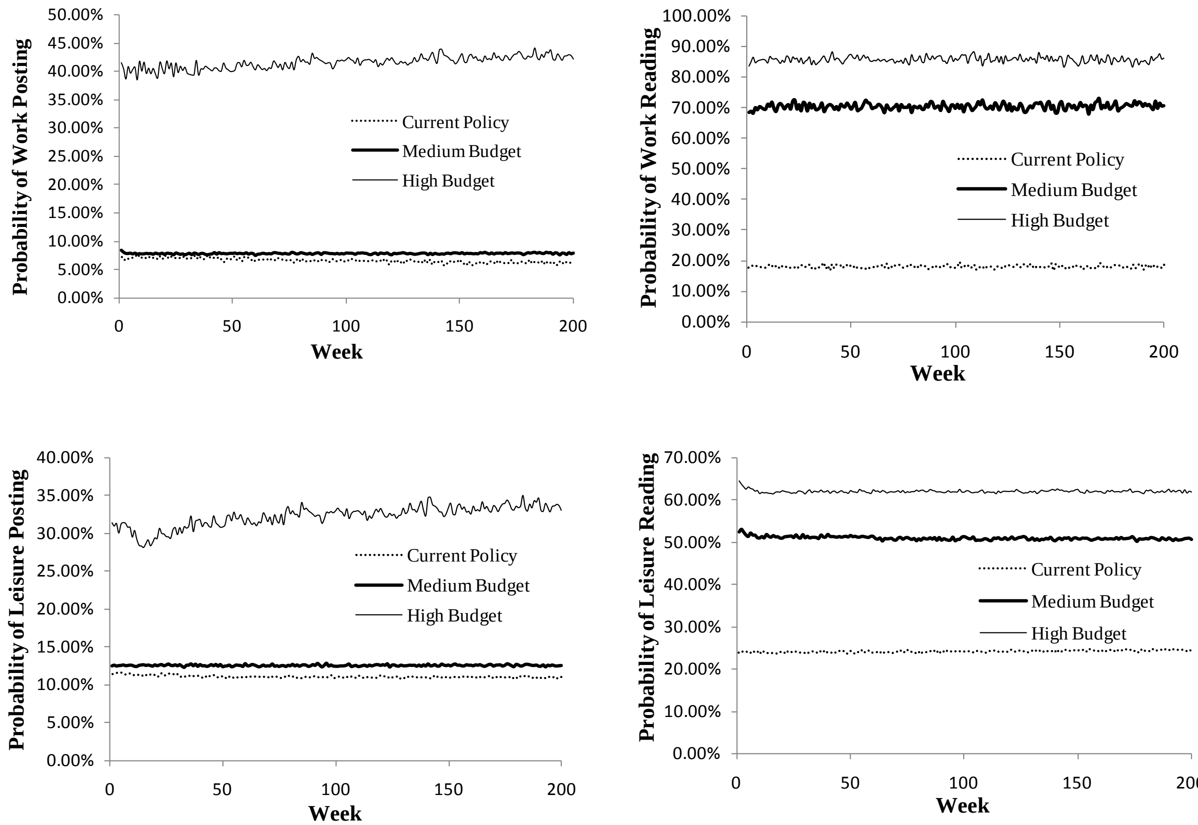

Figure 12: Comparison of Current Policy with Relax Budget Policy

Thus, we see that the probabilities of both work and leisure posting increase by a larger amount than the probabilities of work and leisure reading when the firm switches from "Medium-Budget" to "High-Budget". In other words, the tradeoff between posting and reading is different under different budget constraints. As mentioned earlier, work posting is the main driver of the knowledge sharing within enterprise blogs. Our policy simulation shows that only by providing great amount of additional time for blogging, can a firm significantly enhance knowledge sharing among employees.

6. Discussion and Conclusion

Over the past few years, we have seen that Web 2.0 and social media technologies have become a powerful lure for organizations; their interactivity promises to bring more employees into daily contact at lower cost (Mckinsey 2009). When used effectively, they also may encourage participation in projects and idea sharing. Blogs provide traces of personal expertise and practices. Making it visible helps to get an idea of who knows what, which is a starting point for collaboration and for allowing knowledge to spread more effectively. While the practice of blogging is ubiquitous in many businesses, research that could inform them is rather limited.

In this paper, we present a dynamic structural model in which employees in an enterprise compete with each other in the process of reading and writing blog posts. There are two kinds of blogs in our context: work-related and leisure-related posts. Users make choices about reading and writing based on their preferences for either type of content. We explicitly model the two distinct kinds of tradeoffs that employees have to make in this context. First, how much time they would like to spend on posting blogs as opposed to reading blogs. Secondly, how much time should they dedicate to (reading and writing) work-related blogs vs. leisure related blogs. Our paper aims to understand employees' motivations for reading and writing work-related and leisure-related posts within an enterprise setting and then draw policy implications based on employees' incentive structure. The model is estimated on a dataset consisting of employees from a large Fortune 1000 IT services and consulting firm.

Our estimates suggest that employees are indeed forward-looking in their behavior. Bloggers have to deal with the effects of visibility that comes because of blogging. While visibility might be a driving factor for blogging, it also comes with challenges of dealing with changes in power distribution when crossing hierarchical boundaries, increased expectations from individuals (especially those with higher reputations) and making errors that are publicly available and thus costly. Our paper shows that it is only in the long term that the benefits of blogging outweigh the costs. Thus any incentive structure that an organization puts forth to encourage employees to blog needs to consider such nuances in user behavior.

There is also evidence of strong competition among employees with regard to attracting readership for their posts. While readership of leisure posts provides little direct utility, employees still post a significant amount of these posts as there is a significant spillover effect on the readership of work posts from the creation of leisure posts. Indeed, we find that there are two countervailing effects of leisure posting. While on one hand, there is market expansion effect wherein people who come to read leisure blogs also are exposed to work blogs and this expands the market of readers. On the other hand, a high reputation for leisure posts can also have a negative signaling effect since it suggests that the blogger may not be sufficiently committed to work-related blogging. Whether one effect dominates the other depends on the level of the employee's work reputation.

Our policy simulations suggest that prohibiting leisure-related posting will be counter-productive for organizations since it leads to a reduction in work-related posting too. There is a positive spillover effect from leisure posting to work-posting that enterprises should account for before implementing such policies. Overall, these results shed light on how enterprise adoption of social media tools is associated with employee behavior and choices.

Our paper has several limitations. First, we model content consumption and creation as binary choices. Future research can model them as quantity or count variables. Second, we tested for individual-specific cross-sectional unobserved heterogeneity in only the intercept term. Future research can potentially account for cross-sectional unobserved heterogeneity in the slopes too and use methods similar to Misra and Nair (2009) to have a nonparametric accommodation of unobserved heterogeneity. Third, in term of reading, we only model whether an individual reads work related or leisure related posts. We do not model whose blog an individual reads. It is possible that an individual may follow a blogger rather than a specific work or leisure posts. Individual reading dynamics may be affected by the dynamics of content created by the blogger he/she follows. Future research can explicitly model this relationship. Finally, we accounted for blog reading behavior only within the enterprise. Some of the blog reading and creating dynamics may be affected by events outside of our data context. For example, employees can also access outside blogs. However, such data are not available to us. Notwithstanding these limitations, we hope our work can pave the way for future research in this emerging area of enterprise social media.

7. References

Aggarwal, R., Gopal, R., Sankaranarayanan, R., and Singh, P. V. 2010. Blog, Blogger and the Firm: Can Negative Posts by Employees Lead to Positive Outcomes? *Information Systems Research* forthcoming.

Aguirregabiria, V., and Ho, C. 2009. A Dynamic Oligopoly Game of the US Airline Industry: Estimation and Policy Experiments. Working Paper.

Aguirregabiria, V., and Ho, C. 2010. A Dynamic Game of Airline Network Competition: Hub-and-Spoke Networks and Entry Deterrence. *International Journal of Industrial Organization (28:4)*, pp 377-382.

Aguirregabiria, V., and Mira, P. 2007. Sequential Estimation of Dynamic Discrete Games. *Econometrica (75:1)*, pp 1-53.

Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K., Jamal, Z., and Koh, K. 2010. Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content. Working Paper, University of Rochester.

Arcidiacono, P., and Miller, R. 2006. CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity. Manuscript, Duke University.

Argote, L., McEvily, B., and Reagans, R. 2003. Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes, *Management Science (49:4)*, pp. 571-582.

Bajari, P., Benkard, C., and Levin, J. 2007. Estimating Dynamic Models of Imperfect Competition. *Econometrica (75:5)*, pp 1331-1370.

Bajari, P., Chernozhukov, V., Hong, H., and Nekipelov, D. 2009. Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game. Unpublished manuscript, Dep. Econ., Univ. Minn., Minneapolis.

Breusch T. S. and Pagan A. R. 1980. The Lagrange Multiplier Test and its Applications to Model Specification in Econometric, *The Review of Economic Studies(47:1)* , pp. 239-253.

Ching, A. 2010. A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration. *International Economic Review*, *(51:4)*, pp.1175-1207.

Chung, D., Steenburgh, T., and K. Sudhir 2009. Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus Based Compensation Plans, Working Paper.

Constant, D., Kiesler, S., and Sproull, L. 1994. What's Mine Is Ours, or Is It? A Study of Attitudes About Information Sharing. *Information Systems Research (5:4)*, pp. 400-421.

Dhar, V., and Chang, E. 2009. Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing (23:4)*, pp. 300-320.

Dubé, J.P., Hitsch, G. and P. Manchanda. 2005. An Empirical Model of Advertising Dynamics. *Quantitative Marketing and Economics, (3:2)*, pp. 107-144, 2005.

Dubé, J. P., Sudhir, K., Ching, A., Crawford, G., Draganska, M., Fox, J., Hartmann, W., Hitsch, G., Viard, B., Villas-Boas, M., Vilcassim, N. 2005. Recent Advances in Structural Econometric Modeling: Dynamics, Product Positioning and Entry. *Marketing Letters*, *(16:3-4)*, pp. 209-224.

Erdem, T., Keane, M., and Sun, B. 2008. The Impact of Advertising on Consumer Price Sensitivity in Experience Goods Markets. *Quantitative Marketing and Economics (6:2)*, pp. 139-176.

Ericson, R., and Pakes, A. 1995. Markov-Perfect Industry Dynamics: A Framework for Empirical Work. *The Review of Economic Studies (62:1)*, pp. 53-82.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Information in Electronic Markets, *Information Systems Research (19:3)*, 291-313

French, J. R. P., and Raven, B. H. 1959. The bases of social power. In D. Cartwright (Ed.), Studies in social power, Ann Arbor, MI: Institute for Social Research, pp. 150-167.

Ghose, A., and Han, S. 2009. An Empirical Analysis of User Content Generation and Usage Behavior in Mobile Digital Media. Working Paper.

Ghose, A., and Han, S. 2010. A Dynamic Structural Model of User Learning in Mobile Media Content. Working Paper.

Grossman, M., and McCarthy, R. 2007. Web 2.0: is the Enterprise ready for the adventure, *Issues in Information Systems (8:2)*, pp.180-185.

Gurram, R., Mo, M., and Gueldemeister, R. A. 2008. Web Based Mashup Platform for Enterprise 2.0, *Web Information Systems Engineering 2008 Workshops*, 5176, pp. 144-151.

Hofstetter, R., Shriver, S., and Nair, H. 2009. Social Ties and User Generated Content: Evidence from an Online Social Network. Working Paper, Stanford University.

Hotz, V., and Miller, R. 1993. Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies (60:3)*, pp 497-529.

Hausman, J. A. 1978. Specification Tests in Econometrics, *Econometrica(46:6), pp.* 1251-1271.

Huh, J., Jones, L., Erickson, T., Kellogg, W., Bellamy, R., and Thomas, J. 2007. Blogcentral: The Role of Internal Blogs at Work. *Conference on Human Factors in Computing Systems*, pp. 2447-2452.

Kavanaugh, A., Zin, T., Carroll, J., Schmitz, J., Pérez-Quiñones, M., and Isenhour, P. 2006. When Opinion Leaders Blog: New Forms of Citizen Interaction. *Proceedings of the 2006 international conference on Digital government research*, pp. 79-88.

Kumar. V., and Sun, B. 2010. Why do Consumers Contribute to Connected Goods? A Dynamic Game of Competition and Cooperation in Social Networks. Working Paper.

Lee, S., Hwang, T., and Lee, H. 2006. Corporate Blogging Strategies of the Fortune 500 Companies. *Management Decision (44:3)*, pp 316-334.

Levy, M. 2009. WEB 2.0 implications on knowledge management, *Journal of Knowledge Management (13:1)*, pp. 120-134.

Lu, Y., Singh, P. V, and Sun, B. 2010. Blind Men Can Judge No Color: A Dynamic Structural Analysis of Enterprise Knowledge sharing. Working Paper.

McAfee, A. P. 2006. Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Sloan Management Review, (47:3)*, pp. 21–28.

McAfee, A. P. 2009. Enterprise 2.0 is vital for business, *Financial Times (9 December)*.

McKinsey. 2009. How companies are benefiting from Web 2.0: McKinsey Global Survey Results, *McKinsey Quarterly (18 September).*

Misra, S. and H. Nair. 2009. A Structural Model of Sales-Force Compensation Dynamics: Estimation and Field Implementation, *Quantitative Marketing and Economics* forthcoming.

Moe, W., and Schweidel, D. 2010. Online Product Opinion: Incidence, Evaluation and Evolution. Working Paper.

Nardi, B., Schiano, D., Gumbrecht, M., and Swartz, L. 2004. Why We Blog. *Communications of the ACM (47:12)*, pp 41-46.

Pakes, A. And Mcguire, P. 1994. Computing Markov-Perfect Nash Equilibrium: Empirical Implicaitons Of A Dynamic Model. *Rand Journal of Economics*, pp. 555-589.

Pesendorfer, M., Schmidt-Dengler, P., and Street, H. 2003. Identification and Estimation of Dynamic Games. NBER working paper.

Rabiner, L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE (77)*, pp. 257-286.

Ryan, S., Tucker, C. 2008. Heterogeneity and the Dynamics of Technology Adoption, Working Paper.

Roberts, J. A, Hann, I. H, Slaughter, S. A. 2006. Understanding the Motivations, Participation, and Performance of OSS Developers: A Longitudinal Study of the Apache Projects, *Management Science (52:7)*, pp. 984-999.

Singh, P., Sahoo, N., and Mukhopadhyay, T. 2010a. Seeking Variety: A Dynamic Model of Employee Blog Reading Behavior. Working Paper.

Singh, P.V., Tan, Y., and Youn, N. 2010b. A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects. *Information Systems Research,* forthcoming.

Sweeting, A. T. 2007. Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry, Working Paper.

Thomas-Hunt, M. C., Ogden, T. Y., and Neale, M. A. 2003. Who's really sharing? Effects of social and expert status on knowledge exchange within groups. *Management Science (49:4)*, pp. 464-477.

Trusov, M., A. Bodapati and R. E. Bucklin. 2010. Determining Influential Users in Internet Social Networks. *Journal of Marketing Research, (47:4)*, pp. 643-658.

Venkataraman. S. , Chintagunta, P. K., Gopinath, S. 2009. Do Online Blogs Influence Offline Movie Box-office Performance, Working Paper, WIMI.

Yardi, S., Golder, S., and Brzozowski, M. 2009. Blogging at Work and the Corporate Attention Economy. *International Conference on Human factors in Computing Systems*, pp. 2071-2080.

Online Appendix

The HMM in our model is comprised of three elements: 1) the initial knowledge state distribution, π; 2) the knowledge state transition probability matrix Q (t, t+1); and 3) the observed outcome probability vector Λ. We assume that the evolution of K_{itw} only depends on d_{itwr} and an unobserved error and the evolution of K_{itl} only depends on d_{itlr} and an unobserved error. Nevertheless, the observed outcome probability vector, which is CCP defined above, is jointly determined by s_i and s_{-i} which include both K_{itw} and K_{itl}, as well as all other state variables. Note that for each period we observe the work and leisure related reputations of all employees. Hence, the state transition matrix in our HMM corresponds to only knowledge state transitions.

Let $\lambda = (\pi, Q, \Lambda)$ denote the complete parameter set of the HMM model. Let $s_i = (s_{i1}, s_{i2}, \ldots, s_{iT})$ denote employee i's sequences and $a_i = (a_{i1}, a_{i2}, \ldots, a_{iT})$ denote an observed outcome sequence. The probability of observing the sequence actions a_i conditional on the current states are defined as

$$P(a_i|s_i, s_{-i}) = \prod_{t=1}^{T} P(a_{it}|\lambda, s_{it}, s_{-it})$$

Given that any two adjacent observed outcomes are linked only through the hidden states, we have

$$P(a_i|\lambda, s_i, s_{-i}) = P(a_{i1}|\lambda, s_{i1}, s_{-i1})P(a_{i2}|\lambda, s_{i2}, s_{-i2}) \ldots \ldots P(a_{iT}|\lambda, s_{iT}, s_{-iT})$$

Where $P(a_{it}|\lambda, s_{it}, s_{-it})$ is probability of observing action a_{it} given states s_{it} and s_{-it}. States other than K_{itw} and K_{itr} are all observed in the data, so they are deterministic. For notational simplicity we suppress all states other than K_{itw} and K_{itr}. Since the evolutions of $K_{iw} = (K_{i1w}, K_{i2w}, \ldots, K_{iTw})$ and $K_{il} = (K_{i1l}, K_{i2l}, \ldots, K_{iTl})$ are independent conditional on d_{itwr} and d_{itlr}, according to Markov property, the probability of such an unobserved state sequence, K_{ij} (j= w or l), is given by

$$P(K_{ij}|\lambda) = \pi(ij)P(K_{i2j}|K_{i1j})P(K_{i3j}|K_{i2j},) \ldots \ldots P(K_{iTj}|K_{iT-1j})$$

Then the joint probability of a_i and K_i=(K_{ip}, K_{ir}) is given by

$$P(a_i, K_i, K_{-i}|\lambda) = P(a_i|\lambda, K_i, K_{-i})P(K_i, K_{-i}|\lambda)$$

Applying rule of total probability, we get the likelihood of observing a sequence of outcome is

$$L(a_i) = P(a_i|\lambda) = \sum_{\forall K_i, K_{-i}} P(a_i|\lambda, K_i, K_{-i}, s_{-K})P(K_i, K_{-i}|\lambda)$$

To simplify calculation of the likelihood, following MacDonald and Zucchini (1997), we rewrite the equation above as

$$L(a_i) = \pi(i)\Lambda(i,1)Q(i,1,2)\Lambda(i,2)Q(i,2,3)\ldots Q(i,T-2,T-1)\Lambda(i,T)1' \tag{10}$$

For consistency, we assume that K_{iw} and K_{il} have kw and kl possible levels. We try multiple values of kw and kl and implement maximum likelihood estimation under each value. We choose the value of kw and kl that jointly minimize Bayesian Information Criterion (BIC). In the Equation (10), $\pi(i)$ is the initial state distribution K_i. Then for each j in w or l, define $Q_j(i,t,t+1) = P(K_{it+1j}|K_{itj}, d_{itjr})$. Assume that employees' knowledge states can only transit between adjacent states, we have

$$P(K_{it+1j} = m+1 \middle| K_{itj} = m, d_{itjr}) = 1 - \frac{\exp(\mu(h,m) - \eta_{jm} d_{itjr})}{1+\exp(\mu(h,m) - \eta_{jm} d_{itjr})}$$

$$P(K_{it+1j} = m-1 \middle| K_{itj} = m, d_{itjr}) = \frac{\exp(\mu(l,m) - \eta_{jm} d_{itjr})}{1+\exp(\mu(l,m) - \eta_{jm} d_{itjr})},$$

$$P(K_{it+1j} = m \middle| K_{itj} = m, d_{itjr}) = \frac{\exp(\mu(h,m) - \eta_{jm} d_{itjr})}{1+\exp(\mu(h,m) - \eta_{jm} d_{itjr})} - \frac{\exp(\mu(l,m) - \eta_{jm} d_{itjr})}{1+\exp(\mu(l,m) - \eta_{jm} d_{itjr})}$$

Here $\mu(h,m) = \infty$ and $\mu(l,m) = -\infty$. $Q(i,t,t+1)$, the joint state transition matrix considering both work-knowledge and leisure knowledge state, will be the Kronecker product of $Q_w(i,t,t+1)$ and $Q_l(i,t,t+1)$.

$\Lambda(i,t) =$
$\mathrm{diag}\big(p(a_{it}|\lambda, K_{itw} = 1, K_{itl} = 1), p(a_{it}|\lambda, K_{itw} = 1, K_{itl} = 2), \ldots, p(a_{it}|\lambda, K_{itw} = k, K_{itl} = k)\big)$ is a kw.kl$\times$ kw.klmatrix and 1' is a kw. kl $\times$ 1 vector. Finally, we need to know $P(a_{it}|\lambda, K_{it})$, or more generally, $p(a_{it}|\lambda, s_t)$. This is exactly CCP defined by Equation (9). Following Bajari et al. (2009), we use sieve logit method to estimate $\zeta^B_{a_i}$ with a second degree orthogonal polynomial, defined by $b^B(s_i, s_{-i}) = (b_1(s_i, s_{-i}), b_1(s_i, s_{-i}), \ldots b_B(s_i, s_{-i}))'$, as bases. The CCP is expressed as

$$\Pr(a_i|s_i) = \frac{\exp(b^B(s_i, s_{-i})' \zeta^B_{a_i})}{\sum_{a_i=0}^{15} \exp(b^B(s_i, s_{-i})' \zeta^B_{a_i})}$$

We apply maximum likelihood method to estimate λ. The model with kw=2 and kl=2 has the smallest BIC, indicating that each type of knowledge state has 2 levels. We define the two levels as high and low.