DIVERSITY OF THOUGHT IN THE BLOGOSPHERE: IMPLICATIONS FOR INFLUENCING AND MONITORING IMAGE

A Dissertation

by

PAUL DWYER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Marketing

DIVERSITY OF THOUGHT IN THE BLOGOSPHERE: IMPLICATIONS FOR INFLUENCING AND MONITORING IMAGE

A Dissertation

by

PAUL DWYER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Rajan Varadarajan
Committee Members,	Benton Cocanougher
	Venkatesh Shankar
	Richard Woodman
Head of Department,	Jeffrey Conant

August 2008

Major Subject: Marketing

ABSTRACT

Diversity of Thought in the Blogosphere: Implications for Influencing and Monitoring Image.

(August 2008)

Paul Dwyer, B.S.C.E., University of South Florida;

M.B.A., Texas A&M University

Chair of Advisory Committee: Dr. Rajan Varadarajan

A *blog*, a shortened form of *weblog*, is a website where an author shares thoughts in *posts* or *entries*. Most blogs permit readers to add comments to posts and thereby be a conversational mechanism. One way that companies have started to use blogs is to monitor their corporate image (in this dissertation, the term *image* is used in reference to corporate, brand and/or product image). This study focuses on how common socio-psychological processes mediate consumers' revelation of corporate image in the blogosphere. Centering resonance analysis, a means of measuring similarity between two bodies of text, is used in conjunction with multidimensional scaling to locate text as cognitive objects in a space. Clusters are then detected and measured to quantify diversity in the thoughts expressed. Detected patterns are studied from a social process theory perspective, where complex phenomena are hypothesized to be the result of the interaction of simpler processes.

A majority of blog commenters compromise the expression of their thoughts to gain social acceptance. This study identifies the most extreme of such people so companies who monitor blogs can assign less weight to image indications gained from them as they may be merely expressing thoughts that are intended to maintain social acceptance.

It was also found that single-theme blogs attract a readership with similarly narrow interests. The boldest and most diverse thinkers among comment writers have the most impact because of their ability to provoke the thinking of others. However, commenters who repeat the same ideas have little effect, suggesting that introducing shills is unlikely to shift the sentiment of a blog's readership.

People participate in blog communities for reasons (e.g., need for community) that may undermine thought diversity. However, there may be value in serving those needs even though no valuable insights are provided into image or directions for product development. Members of homogeneous-thinking communities were observed to more actively participate, with greater longevity. This may increase loyalty to the company hosting the blog.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Varadarajan, and my committee members, Dr. Cocanougher, Dr. Shankar, and Dr. Woodman, for their guidance and support throughout the course of this research.

Thanks also go to my friends, colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

The term *blog*[1], a shortened form of *weblog*, refers to a website where an author shares thoughts in *posts* or *entries*. Most blogs permit readers to add comments to posts and thereby be a conversational mechanism. Popular blogs have become meeting places for communities of readers. Organizations are trying to better understand the relevance of these online communities to their operations and how they should respond to the phenomenon.

Murray (2007) ranked blog readership as a percentage of population among world nations. He reported that Japan ranked highest with 74% of the population reading blogs, followed by South Korea (43%), China (39%), the United States (27%), and the United Kingdom (23%) in the top five. In each country, the younger the reader, the more time they spend reading blogs. Additionally, blog readers were more likely to self-identify as taking part in political, community or social welfare activities and were thus designated "influencers" (p. 5).

This dissertation follows the style of *Journal of Marketing*.

[1] The first use of a term central to this study is italicized and the word is defined in the glossary in Appendix A.

Technorati (2007), a blog search engine, reports that it tracks almost 100 million blogs. These blogs cover such a variety of subject matter that they resist the creation of the kind of taxonomy that scientists generally seek to create. The blogosphere eschews top-down taxonomies in favor of *folksonomies*, a form of collaborative categorization based on the aggregation of *tags*, key words descriptive of content (Golder and Huberman 2006). Nevertheless, in Table 1.1 a categorization framework is described that delineates types of blogs suggested most of interest to the marketing community.

The framework begins by partitioning marketing-related blogs into those related to business, and those related to public policy and non-profit issues, so as to capture the full breadth of marketing areas of practice. While, it may seem that more importance has been conferred on business blogs, as the category breakdown is more nuanced, the general structure of the business blog hierarchy can be applied to public policy and non-profit blogs as well.

Established information providers have been augmenting their current service offerings with new ones based on the extraction of information from the *blogosphere*, the collection of all blogs. One example is VNU, the corporate parent of ACNielsen and Nielsen Media Research, which announced in January 2006 the acquisition of BuzzMetrics and Intelliseek, two pioneers in the monitoring of online *consumer-generated media*, and merged them into Nielsen BuzzMetrics. This acquisition of new competencies by an industry giant signals that they perceive consumer-generated media to be a new information resource that, rather than being a passing fad, is growing in significance.

One type of information of considerable interest to companies is information about their corporate image[2]. Companies have an image they would like to instill in the minds of customers. They are aware that all methods of communicating that image are imperfect and inefficient to varying degrees. Hence, the need to infer the image in the consumers' minds, compare it to the one they wish to create and then develop and execute ways to bring the two images into congruence. Nielsen BuzzMetrics offers a service called BrandPulse that monitors blogs and message boards for mention of their client customers' *brand*s. Blogs of many types are potential sources for marketing insights even though organizations tend to directly connect with consumers in only a few blog types. As most of the world economies are either consumerist or in the process of becoming so, any blog can be a venue for the discussion of corporate brands, trademarks and other image assets.

[2] Throughout this dissertation, the term *image* refers generally to corporate, brand and/or product image.

TABLE 1.1

A Hierarchy of Marketing-Relevant Blog Types

			Example and Description
Business-Oriented	**Corporate**	**External**	**IBM developerWorks.** A large set of blogs authored by IBM product developers from a wide array of product lines.
		Sanctioned Employee	
		Senior Executive	**GM Fastlane.** GM Vice Chairman Bob Lutz' blog on current and future product offerings.
		Character	**The Family Guy Blog.** Supposedly written by a fictional character, often a TV character, mascot or brand icon.
		Internal	A blog intended for the exchange, accumulation and management of internal corporate knowledge.
	Personal	**Communal**	**Hiptop Nation.** A consumer's blog that became the primary "moblog" (i.e., mobile weblog) for Danger Hiptop and T-Mobile Sidekick enthusiasts to post photos, etc.
		Diary-like	**Wil Wheaton.** An actor's personal blog.
		Firm-Sponsored	**Bob Balfe.** An IBM Lotus developer's blog, linked to by IBM, but hosted apart from developerWorks.
		Review-like	**Paul Stamatiou.** The widely linked to blog of a college student who reviews products and gives technical support.
		Unsanctioned Employee	**The Masked Blogger.** Rumored to be the blog of an Apple employee. Apple authorizes no external blogs.
	Third-Party	**Corporate Watchdog**	**WalMart Watch.** Analysis and commentary of WalMart.
		Consumer Advocacy	**The Consumerist.** A consumer news and retaliation blog.
		Industry-level	**AutoBlog.** News and reviews of cars in general.
Public Policy and Non-Profit	**Organizational and Personal**	**Religion and Culture**	**The Evangelical Outpost.** Commentary on culture and religion from an evangelical perspective.
		Social and Health	**Kalyn's Kitchen.** Recipes for healthy meals.
		Political	**Blog for America.** The official blog of the Democratic Party.

Scoble and Israel (2006), the authors of *Naked Conversations*, based on their informal observations, note that blogs can be a rich source of insight into how customers view a company. However, they caution that these insights may only reflect the views of a vocal minority, a phenomenon they call the *echo chamber*. This expression of caution inspired the general research question that started this study's line of research: Is there a way to measure the extent to which a blog has become an echo chamber?

The echo chamber phenomenon is one of many models or *social forms* that have been used to explain behavior observed within group or community settings. The term "social forms" is drawn from *social process theory* and refers specifically to lesser socio-psychological processes that operate both in tandem and conjunction to create complex social phenomena (Cederman 2005). This study investigates the expression of thought in a diverse selection of corporate, political and individual *weblogs* from a social process theory perspective.

Kawasaki (2004), an entrepreneur and venture capitalist, opines that the most successful products are polarizing: consumers either love them or hate them; only mediocre products are viewed with indifference. Kawasaki's point of view is consistent with Muniz and O'Guinn (2001) and McAlexander et al. (2002) who see passion as what motivates community and results in loyalty. It might be thought then that the echo chamber, and its attendant limited expression of thought, is the most desirable condition for consumer communities as it allows companies to monitor their image among its core constituency without the introduction of noise from those who passionately dislike, or are merely indifferent to, them and their products. This argument has surface merit, but

since an echo chamber is an extreme situation of limited thought expression, it creates a context that by focusing on the status quo engenders few ideas for new products or to solve problems that might expand the customer base without alienating the core. These limited perspectives also provide no indication of how a target market may be evolving in its determinants of satisfaction.

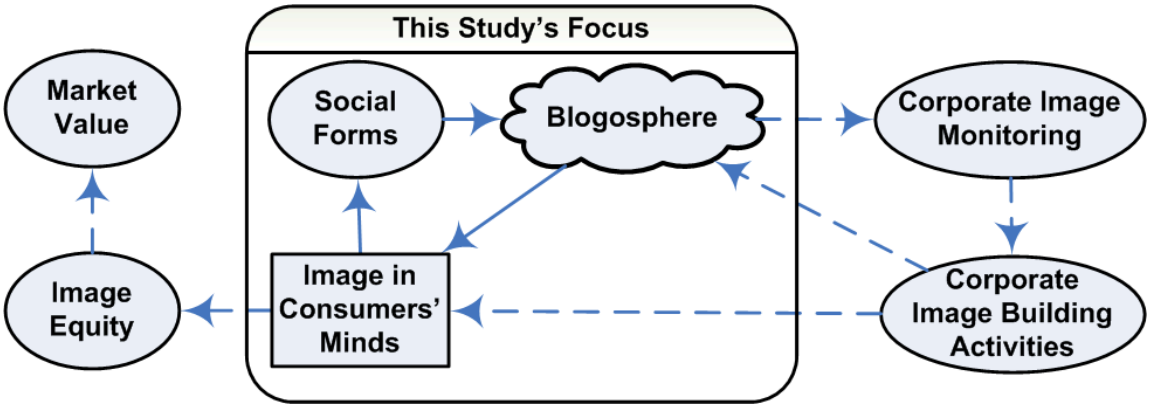
Image monitoring is complementary to satisfaction analysis frameworks like Kano's (1984) model, where product attributes are differentiated on the basis of being basic essentials, determinants of performance and sources of excitement. By maximizing diversity of perspectives companies increase the likelihood of finding product enhancements that are basic essentials to new consumer segments while being at worst, factors of indifference, and at best, performance or excitement increasers to existing customers. Additionally, it is naïve to believe that other companies will not compete for their competitors' loyal and most profitable core. Jones and Sasser (1995) note that for a consumer to be truly loyal they must be completely satisfied. Furthermore, they note that complete satisfaction is a moving target and that communing only with existing customers can miss leading indicators of market shifts and their attendant satisfaction determinants. They recommend that companies use a variety of methods to listen to existing, potential and former customers. Clearly, the goal of complete satisfaction can only be met by maximizing diversity of insight from the market.

Figure 1.1 shows how the blogosphere could fit into a corporate image management process. Kapferer (2004) notes that corporate image can be affected by conventional marketing communications such as public relations, advertising,

sponsorship and personal selling, as a part of a multifaceted image management program. This study investigates the efficacy of adding blogging, as a new and emerging form of interactive marketing communication, and blog monitoring to the array of image management activities. Companies can monitor how they are portrayed in the blogosphere and use those insights to modify their image building activities thus creating a feedback loop. Consumers are assumed to interact with the blogosphere in a manner mediated by social forms. This study broadly focuses on this mediation and its impact on the extent to which the image in the blogosphere presents an accurate representation of brand image in consumers' minds.

FIGURE 1.1

Corporate Image Management and the Blogosphere



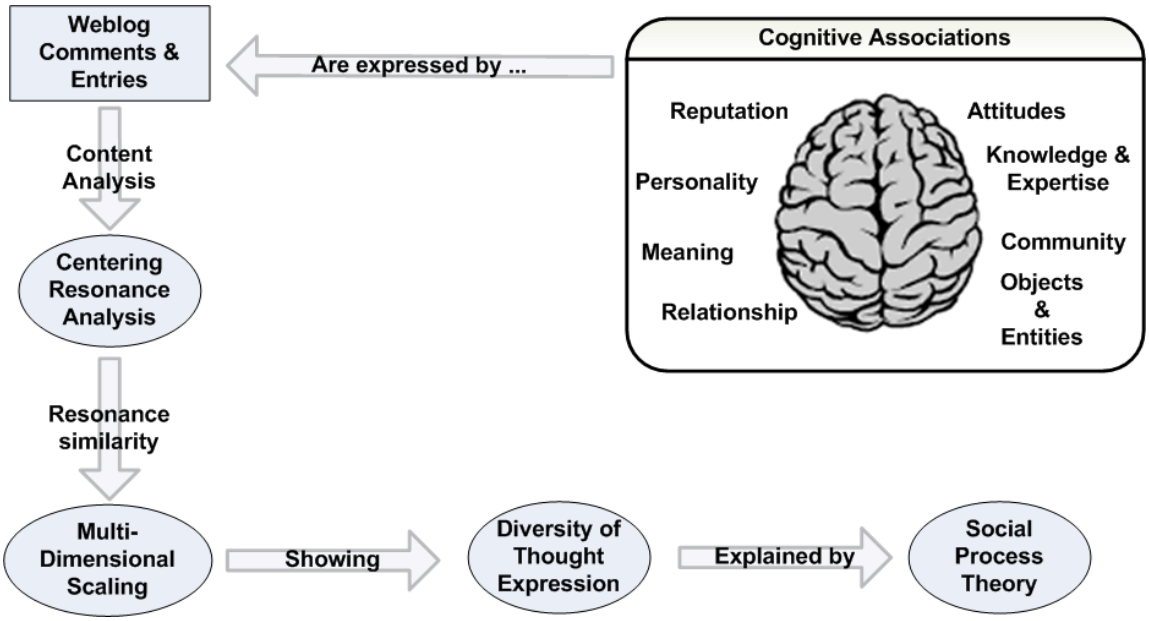
Literature Overview

This dissertation draws on an eclectic body of past research. Figure 1.2 provides an overview of the literature underpinnings of this research. In Chapter II, each entity in

Figure 1.2 is described in the order given and greater detail is provided into how these entities connect together. A more detailed definition of a blog is provided. In addition, the large body of prior research into the nature of brand cognitive associations and how these associations have previously been elicited from consumers in a context similar to that of a blog is summarized. Since blogging uses written communications, advances in textual analysis are also summarized, culminating in a detailed description of *centering resonance analysis* (CRA) which yields a similarity measure for pairs of text bodies. Then, how this similarity measure can be translated with multidimensional scaling into points in a cognitive space and used as a basis for measuring diversity in the thoughts expressed is described. Finally, some alternative models (social process theories) are discussed that explain the diversity of thought expression, and thought expression in general, in a group context.

FIGURE 1.2

Conceptual Map of the Literature Reviewed



Research Questions

As previously stated, companies are interested in knowing how they are perceived by consumers and are turning to online community monitoring to gain insights. However, the question of the extent to which perceptions gained from these communities are a true reflection of their image in the overall market is not known and has not been researched. Blog-based research can be characterized as *quasi-experimental* because it has some of the attributes of experimental scientific research, like the collection of data to which statistical methods can be applied, but it suffers from a lack of control over factors that might confound the identification of testable cause-and-effect relationships. Without an understanding of the processes that are operant within blogs,

not much faith can be placed in findings based on statistical analyses. This research study is aimed at identifying some of the major processes operating within blogs, particularly the ones that affect how thoughts get expressed and thereby perform some basic research for the modeling of cause-and-effect within blogs. Toward this end, the following research questions are investigated:

1. What mechanisms explain the expression of thoughts by individuals in a blog? Since comments are a vehicle for consumers to express their thoughts, there is a need to determine which social processes are most relevant in determining whether or not comments are made, and if so, when they are made.
2. What processes expand *cognitive diversity*, the expression of diverse thought? Following Scoble and Israel's (2006) focus on the echo chamber phenomenon as a factor that could reduce the expression of diversity in thought, research questions 2 and 3 seek to look at both sides of the issue of factors affecting the expression of thought diversity. Research question 2 focuses on which of the processes identified in the first research question tend to increase the diversity of thought expression.
3. What processes limit cognitive diversity? Addressing this question has the potential to shed insights into Scoble and Israel's (2006) concerns about the echo chamber phenomenon and reveal other factors that might limit cognitive diversity.
4. What is the relative importance of these processes? The core concept of social process theory is that social processes act in tandem and in conjunction. It may

be easier to cope with any distortion these processes bring to brand image insights if it is known how these processes are ranked according to influence.

Potential Contributions

This study seeks to make the following substantive and methodological contributions to the field of marketing:

1. Substantive: Provide empirically-based insights into how individuals reveal their cognitive associations in groups, particularly blogs. Demonstrate the extent to which social process theory explains the expression of thought in blog communities, and thereby enhances our understanding of the usefulness (value) of using blogs as a source of image insight.
2. Methodological: Demonstrate a novel integrative application of CRA with a type of multidimensional scaling that uses a spring-based iterative algorithm, whereby similar objects (in this case, bodies of text) are pulled together as dissimilar ones are pushed apart until a stable equilibrium positioning is attained within a multidimensional space. When these bodies of text have settled into stable coordinates, thematic clusters are detected and measured to quantify diversity in expressed thought.

CHAPTER II

OVERVIEW OF RELEVANT LITERATURE

A blog differs from a forum, newsgroup or general online community in that a single author, or discrete set of authors, initiates the discussion. Most blogs permit readers to add comments to posts and facilitate *trackbacks*, permanent linking to posts by other blogs and websites. The collection of all blogs is often called the *blogosphere*. Blogging pundits, such as Scoble and Israel (2006), hail the blog as an ideal forum where organizations can converse with their market and constituents to gain insights into how they are perceived. Since their widespread emergence after the Internet achieved mainstream adoption, a number of marketing researchers have focused on virtual communities in their work (e.g., Shoham 2004; Dholakia, Bagozzi and Pearo 2004); however, there is a dearth of research studies that have used blogs as the research context.

A discussion of the various kinds of image associations that consumers might reveal in a blog follows in the next section.

Cognitive Associations

People tend to have very complex entwined cognitive associations with organizations, brands and products. Govers and Schoormans (2005) make a distinction between people's perceptions of a brand and a product. Similarly, distinctions can be made between an organization and a brand in its portfolio. Organization, product and

brand are levels in a hierarchy of consumer associations. While a consumer may have different associations between these three levels, the nature of these associations is the same in that they are all cognitive structures. While the literature reviewed in this chapter tends to focus on one level of this hierarchy, the thoughts expressed generally apply to all. Since this discussion is focused on cognitive structures in general, the words organization, brand and product are used interchangeably with "brand" receiving the most use.

Reputation

Brown et al. (2006) describe the perspectives from which an organization can be viewed. *Identity* is the way an organization views itself. *Intended image* is the way an organization would like to be perceived by outsiders. *Construed image* is the way an organization thinks it is being perceived and *reputation* is the way it is actually perceived. This taxonomy generally is inline with Gioia, Schultz and Corley (2000); however, Gioia et al. made a distinction between *image*, the transient aspects of how an organization is perceived, and *reputation*, the long-term perception. Brown et al. adopted Brown and Dacin's (1997) definition of reputation as *organizational associations,* a "label for all the information about a company [organization in the broadest sense] that a person holds" (p.104).

Brown et al. (2006) also made a distinction between reputation as an attribute of an organization and reputation as the sum of all mental associations possessed by an individual. This distinction is important as it explains why there is a need to distinguish between the various organizational viewpoints: the way it is viewed by others is only

partially under an organization's control. An organization is constantly communicating through overt means such as advertising and through more subtle means such as its products and its actions within the general community. Theorists have proposed various models for how an organization's message could be perceived differently than intended. Shannon and Weaver (1949) introduced an encoding-decoding model where a message is encoded by a sender, transmitted through a medium and decoded by a receiver into a different message. They attribute a difference in comprehension to *noise*, random interference, in the transmission medium. Later research, such as that reported by Grayson (1998) attributes it to differences in the *interpretive repertoire* of the sender and receiver. The sender encodes the message presuming that the receiver will not decode it differently. Hall (1980) describes a *circuit of communication* whereby, through feedback and clarification, a sender and receiver can converge on a shared meaning. The capacity for reciprocal communication is one of the most attractive attributes of the blog.

Sabate and Puente (2003) concluded that a good corporate reputation results in higher profits and higher profits result in a better reputation. Roberts and Dowling (2002) decomposed reputation into financial (i.e., the company is good investment) and general (the company is perceived as a good social actor) components. On this foundation, Dowling (2006) concluded that companies can achieve superior financial performance by investing in "being more profitable and in being perceived as good" (p. 135). Dowling observed that a good reputation aids corporate growth by making sales easier and by making investors more willing to fund a company's attempts at testing new ventures. Obloj and Obloj (2006) noted that investments in increasing reputation are

done at the cost of investing in greater profitability. They concluded that superior profits only accrue when there is a large difference between the reputations of competing firms. Therefore, firms should strive only to be a close follower in reputation-building, not the leader in their market.

Personality

In the *brand identity prism* proposed by Kapferer (1992), *personality* was denoted as merely one facet of a body of interdependent associations that make up image. Aaker (1997) took issue with this portrayal and proposed that it be regarded as a separate construct: "the set of human characteristics associated to a brand." Azoulay and Kapferer (2003) argued that Aaker's (1997) definition was too broad and not in keeping with psychology's definition of personality. They supported Kapferer's (1992) model that personality was merely an aspect of image and should only include those human attributes that are appropriate to brands. As a result, they discarded cognition, skills and abilities, gender, social class and ethnicity from the array of human characteristics that can be associated with a brand.

Even though Azoulay and Kapferer (2003) advanced theoretical support for their views, Freling and Forbes' (2005) qualitative study of brand personality found that consumers associate the full spectrum of human characteristics to brands. They ascribe this to a natural human tendency to *anthropomorphize* objects with which they have a relationship.

Belk (1988) noted that consumers make possessions a part of themselves and a reflection of their identity. According to Belk, possessions serve a variety of self-

expressive purposes: they are a tangible link to one's past as well as expressions of the multiple levels of self such as family and groups one belongs to. It would not be surprising to discover that the identity of a consumer may be so entwined with a perceived brand image that they are unable to separate the two in their discourse.

Meaning

Brown, Kozinets and Sherry (2003) added the idea of brand *meaning* to the pantheon of cognitive associations attached to brands. They described four classes of meaning: a*llegory* (a brand story), *aura* (core values), *arcadia* (idealized community) and *antinomy* (paradox: old style, new technology).

Attitude

Petty, Unnava and Strathman (1991) noted that attitude is a relatively global and enduring evaluation of a product. Ajzen (2001) differed by noting that attitude toward an object may not be exclusively a global impression, but may be expressed on a series of independent scales: bad-good, useful-useless, reliable-unreliable, etc. Attitude is relevant to consumption because, as Ajzen (2001) notes "attitude predicts behavior," people purchase products they have a positive attitude toward. Howard and Sheth (1969) position attitude as the mediator between choices and intention.

Relationship

Fournier (1998) distinguished a *relationship* from a one-time transaction and mere habitual purchase from loyalty. Her definition of relationship also included the idea that consumers and brands form partnerships of true interdependence where a brand is typically attributed a human-like personality, much like the findings of Aaker (1997), and the brand is perceived as an active partner in goal fulfillment. These goals can be tangible/utilitarian and/or emotional as brands can facilitate the expression of self. Fournier (1998) also points out that brand relationships exist in the context of other relationships and may be entwined with them. She also stated that, much like human-to-human relationships, brand relationships change over time. Arnould and Thompson (2005) include the brand relationship in their taxonomy of consumer culture theory because assigning human-like status to an *intangible* set of perceptions (i.e. a brand) is an aspect of a constructed reality. John (1999) points out that children involve brands in their socialization process and thereby create the foundations for brand relationships early in life.

Knowledge and Expertise

Keller (2003) defined brand knowledge very broadly, including the intangible aspects discussed here with the purely utilitarian attributes that compose most researcher definitions. Mitchell and Dacin (1996) conceptualize knowledge as an awareness of attributes and how they contribute to performance. Sujan (1985) includes the presence of product categories in her definition, while Ratneshwar and Shocker (1991) unite categorization with an appreciation of appropriate usage context. Expertise is

consistently presented as a high degree of knowledge combined with an ability to properly apply it.

Community

Kleine, Kleine and Kernan (1993) observed that people use brands and products in their identity-crafting processes. They noted that social connections are an important part of that process because they allow a person to get feedback from knowledgeable others about how well they are doing in their identity-crafting. As a result, the knowledge of brand community is an important aspect of the mental associations that a consumer has with brands. McAlexander, Schouten and Koening (2002) supplemented that view by showing that association with a brand community increases a consumer's enjoyment of a product and provides a motivation for having a knowledge of available communities.

Now that the primary image associations someone might reveal in a blog have been described, in the next section the issue of corporate image management is discussed.

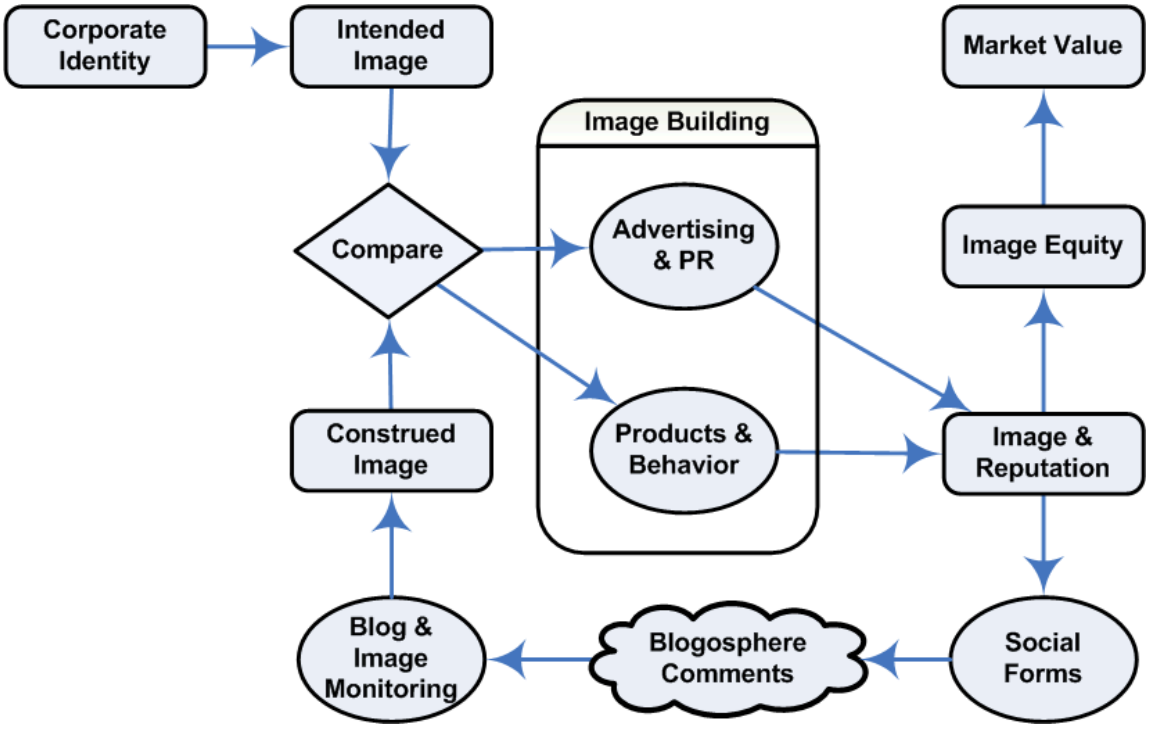
Image Management

Figure 2.1 portrays a possible corporate image management process incorporating blog insights. Companies have a self-image, the way they view themselves, denoted their *corporate identity*. That identity is assumed to be the basis of their *intended image*, the impression they would like to create in the minds of others. Companies compare their intended image with a *construed image*, how they think others view them. If the intended and construed images are too dissimilar, companies can try to

alter the way they are perceived by either tangible (i.e., through products and behavior) or intangible (e.g., advertising and public relations) means. Consumers develop an impression of a company from these image-building activities. That impression is expressed in blog comments and subsequently read by corporate image monitoring. This construed image is again compared with the image the company is trying to convey and image building activities are once again adjusted to compensate for any mismatch. Blog monitoring is thus an integral part of the image-building feedback loop.

FIGURE 2.1

Corporate Image Management with Blogs



The next section provides an overview of literature that supports examining blog communications to detect consumer image associations. It then describes how these associations can be detected and extracted.

Detecting and Extracting Cognitive Associations

As previously stated, this research project is concerned with investigating how individual brand image cognitive associations are revealed in a group context. In order to find patterns in how cognitive associations are revealed it is necessary to measure them and the relationships between them.

Blogging, where an author posts an entry and the world is free to comment, is similar to Olson and Muderrisoglu's (1979) *free elicitation* procedure where "respondents are free to say anything and everything that comes to mind when presented with a stimulus probe cue." They tested the stability of responses between two points in time and found that 50-60% of the associations revealed in the first experiment also appeared in the second. They also found that the order of mention of these associations had a correspondence of about 0.35. This led them to conclude that free responses are a reasonably consistent and reliable measure of thought structure.

Hutchinson (1983) demonstrated the use of a *network-based model* of cognitive associations. He analyzed the transcripts of *unconstrained free response* from consumers and product-category experts to construct a network of brand attributes connected by directional ties (i.e. arrows) showing the order of recall (a temporal network). He built his research on ideas introduced by a number of researchers who found that cognitive associations are retrieved from memory in an order that reflects the way they are stored

(Bousfield and Cohen 1955; Bousfield, Sedgewick and Cohen 1954; Friendly 1979). At the time when he proposed his network mapping of cognitive associations, two and three dimensional visualizations based on data processed with *multidimensional scaling* (Green, Wind and Claycamp 1975) was the state-of-the-art in modeling how "brand-features" were represented in the mind of the consumer.

Ward and Reingen (1990) introduced another network-based model explaining how a shared group cognitive structure emerges from individual structures and then, in a kind of feedback mechanism, homogenized these individual cognitive structures. This means that the knowledge or understanding of individuals becomes shared group knowledge as individuals express their thoughts. Hearing the thoughts of others cause individuals to modify their personal knowledge, making it more like the knowledge shared by the whole group. They used this model to explain the process of how a group arrives at a consensus-based consumption decision. Their work was heavily based on theories of *group polarization*, such as those advanced by Burnstein and Sentis (1981). Groups tend to be attracted to the "initially favored alternative", that is, the first alternative that is well explained and supported by logical arguments. Chandrashekaran, Walker, Ward and Reingen (1996) expanded this model by demonstrating how decisions could be predicted and not merely explained.

It should be noted that all this prior research confined cognitive association maps to those involving the purely utilitarian attributes of products, believed at the time to be most relevant to the purchase decision. As described, consumers associate a much richer set of intangible and non-*fungible* attributes to organizations, brands and products. As

Keller (2003) points out, the mapping of all these cognitive associations has not been explored:

> An important future research challenge … is to develop holistic perspectives toward brand knowledge that would encompass the full range of all the different kinds of information involved, for example, approaches to create and apply detailed mental maps for brands. An ideal mental map representation would be a blueprint of brand knowledge, as comprehensive while also as parsimonious as possible, that would provide the necessary breath and depth of understanding of consumer behavior and marketing activity. (p. 596)

Even though early research embarked on a very promising trajectory it seems to have stalled, possibly waiting for more advanced methodologies to be developed. The paragraphs to follow set the stage for introducing one such new methodology.

Content Analysis

Among the many definitions of content analysis summarized in Kassarjian (1977), the one provided by Kerlinger (1964) most closely fits the objectives of this study:

> Content analysis, while certainly a method of analysis, is more than that. It is … a method of observation. Instead of observing people's behavior directly, or asking them to respond to scales, or interviewing them, the investigator takes the communications that people have produced and asks questions of the communications. (p. 544)

Most of the current methods of content analysis are unchanged from the ones described by Kassarjian (1977). Methodologies tend to be of two types: *manifest* and *latent*. Manifest methods tend to be word use frequency counting systems that cannot capture the way words are used in context, only that some words are used more often than others. Latent methods generally employ a group of human interpreters to overcome

the deficiencies of manifest methods. However, latent methods are often characterized by inter-rater reliability problems.

Thompson (1997) describes a "*hermeneutical* framework for deriving marketing insights" from consumer-generated text that is particularly germane to this study. He uses the word *hermeneutic* to denote a situation where the consumer reflects on the role products play in their present and past circumstances in the context of the broader culture in which they live. The nature of a blog encourages commenting consumers to write at length about an issue that interests them. This permits them sufficient scope for reflection and elaboration. The ongoing communal nature of the blog encourages repeat contribution, allowing them to further elaborate on issues they want to discuss as time and reflection allows them to refine their thinking.

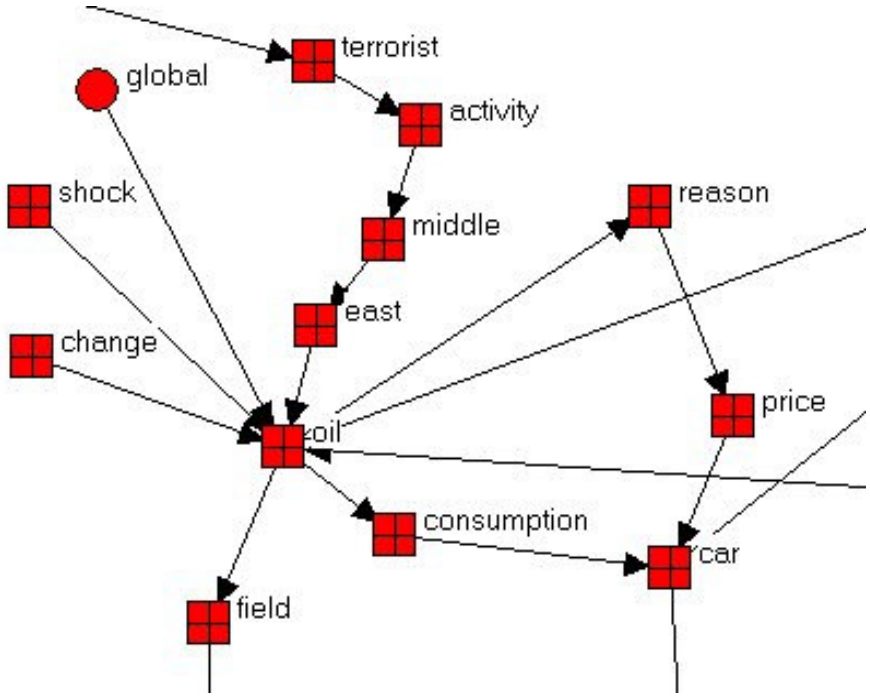
Centering Resonance Analysis

Centering Resonance Analysis (CRA), a method of content analysis proposed by Corman et al. (2002), turns a body of text into a network of nouns and adjectives. The words that have the most effect in giving the text coherence are then found by finding the words with the highest *betweenness centrality*, or simply *betweenness*. These are words that are important because they connect other words together. Figure 2.2 shows part of the word network composed from a comment posted to GM's *Fastlane* blog. It is apparent that the word "oil" has high betweenness because it connects many other words together. Corman et al. (2002) postulate that these word networks provide insights into the way knowledge is structured in the mind of the writer. The words with high betweenness represent the most salient cognitive associations. A quantitative measure of

a word's betweenness is called its *influence*. Two separate bodies of text can be compared on the basis of commonalities in the influence of the same words used. A quantitative measure of this similarity is called *resonance*.

FIGURE 2.2

Betweenness Centrality in Text



CRA operates in a manner consistent with Thompson's (1997), framework because it models a whole body of text as one coherent network and seeks to detect the role specific words play in that network of meaning. The ability to measure the resonance between separate texts allows the researcher to construct an integrative interpretation. This study proceeds on the premise that the most influential words in the

comments are also the most salient organizational, brand and product associations elicited by the author's stimulus posting.

Having described the means of extracting cognitive associations from a body of text, the next section discusses how the many cognitive associations that consumers reveal can be organized to allow the detection of patterns among them.

Mapping Cognitive Associations

Multidimensional Scaling

Multidimensional scaling (MDS) (Green, Wind and Claycamp 1975) has become an important tool of marketing research as it is typically used to create two or three dimensional visualizations of data that leverage a natural human facility for identifying patterns. Even though many datasets reflect more latent dimensions than can be reduced effectively to two or three dimensional visual models without too great a loss of information, there is no need to abandon MDS in such situations. Strictly speaking, MDS is not a technique for data visualization; it is a means of converting pair-wise similarity measures expressed as distances (like one minus the resonance between two bodies of text) into coordinates in a space of any dimensionality. Matrix manipulation and spring models (described below) are the two most common ways of performing MDS. Spring models are usually favored when a large number of distances are used because the manipulation of large matrices is time consuming and resource intensive when performed on personal computers. Since the use of spring model MDS is unusual in marketing, the underlying principles are described here. However, first an even more

unusual means of performing MDS using neural networks is described because the founders of that methodology have used it in situations similar to that of this study.

Woelfel and Fink (1980) introduced *Galileo*, a computer program that extracts "objects of cognition" from a body of text and plots them in a multidimensional space "whose properties are determined by the patterns of interrelations among the objects" (p. 8). Galileo embodies the research findings of Barnett and Woelfel (1979) who documented critical problems with reducing psychological measurements to the minimal set of *eigenvalue*-based dimensions commonly used to define a multidimensional scaling space. Woelfel and Fink (1980) still claim (Woelfel 2007) to be the only researchers who have successfully mapped the cognitive content described in a body of text to a multi-dimensional space. In order to do so, they employed a neural network to translate such textual content to coordinates. Typically, users of two and three dimensional scaling have endeavored to assign labels to the axes of the space that denote attributes on which objects (whether products, customers or cognitive objects) vary. Barnett and Woelfel (1979) point out that in high dimensional spaces, the axes cannot be labeled and *psychological constructs* are points within the space.

Spring Models in Multidimensional Scaling

Kamada and Kawai (1989) devised the algorithm used to perform MDS with spring models. Their algorithm models each data point as connected to every other data point by a spring tensioned according to some measure of similarity (such as CRA resonance). Through an iterative mechanism, similar data points are pulled together as dissimilar ones are pushed apart until an equilibrium configuration is attained. If that methodology were used in this study, representing blog entries and comments as spring-connected objects, once the objects self-organized into stable positions within multi-dimensional space, *thematic clusters* could be detected using methods of *inherent classification* such as those discussed below. Thematic clusters can then be measured in the following ways to indicate diversity in expressed thought:

1. The mean number of clusters across blog entries can be a blog-wide measure of cognitive diversity, while the number of clusters within a blog entry can indicate the degree of cognitive diversity within a single blog entry conversation.
2. Radius and density: Each cluster has a center of mass (also called a *centroid* or barycenter). The mean and standard deviation of the distance between every cognitive object and the centroid are measures of a cluster's radius and density, respectively.
3. Proportion of outliers: The mean and standard deviation of the distance between every cognitive object and its cluster centroid can also be used to identify *outliers*, cognitive objects located far from a cluster's center.

It should be apparent from the above description that the CRA and spring model multidimensional scaling algorithms are particularly appropriate for automation as they make no qualitative determinations that imitate human judgment. There is, therefore, no need to employ multiple human referees to assess the reliability of CRA or spring model multidimensional scaling as the mapping of the same body of text to multidimensional space is certain to have identical results regardless of the number of trials.

Inherent Classification

Most clustering algorithms, such as K-Means and Expectation Maximization (EM), require that the researcher specify the number of clusters the algorithm is to find. However, in this research study, it is necessary to know what clusters are naturally present. One popular clustering algorithm that does not require the number of clusters to be specified beforehand was developed by Chiu et al. (2001). Chiu et al. described an *entropy-based hierarchical clustering* algorithm that implements *agglomerative hierarchical clustering* in two stages:

1. Pre-cluster the data points into a large number of sub-clusters (Chui et al. calls these “dense regions”).
2. Cluster the resulting sub-clusters into a smaller “optimal” number of clusters.

In the first phase, Chiu et al. (2001) use Zhang et al’s (1996) BIRCH method of constructing a cluster feature (CF) tree. Each CF consists of a finite number of data points that are within a certain radius of each other. The maximum number of data points per CF and the radius must be selected prior to building the tree. As the dataset is

traversed, data points are added to CFs, new CFs are created and existing CFs are split until every data point in the dataset is accommodated.

Prior to the second phase, the "optimal" number of clusters is determined in its own two-step procedure:

1. Calculate a coarse estimate of the optimal number of clusters using Bayesian Information Criterion (BIC). BIC is a likelihood-based metric that decreases as the number of clusters increases to a certain point and then increases beyond that point. The aim of this first step is to find the number of clusters at the BIC minimum. Chiu et al. (2001) note that since the BIC estimate almost always overshoots the true optimum number of clusters, another step is needed to refine the estimated optimum.
2. Simulate the merging of all the sub-clusters and calculate the ratio of inter-cluster distance and the number of clusters after each merge. A big jump in the change in this ratio from one merge to another usually indicates that merge should not have occurred. This process can also be done by plotting the number of clusters versus classification error and then locating the "knee" of this curve. This method is sometimes criticized for finding the knee locally by looking at pairs of adjacent points rather than at the whole curve. It is however, a simple method to use.

After the optimal number of clusters has been determined, the CF tree is collapsed into that number of clusters based on the combining of sub-clusters in close proximity to each other. Proximity is measured as a log-likelihood distance related to the

decrease in log-likelihood as the two sub-clusters are combined into one cluster. The distance between sub-clusters w and z is calculated as defined in equation 2.1.

(2.1) $$d(w,z) = \xi_w + \xi_z - \xi_{<w,z>}$$

In equation 2.1, <w, z> denotes the index of the combined cluster and ξ is the entropy, or information homogeneity, of a cluster. The more homogeneous a cluster with respect to the distribution of values within it, the lower its entropy. This clustering method therefore tries to join sub-clusters where the difference between their separate entropies and their joined entropy is small.

Now that a means of organizing and analyzing the relationships between cognitive associations has been described, the literature that offers explanations for these relationships and any patterns detected is reviewed next.

Social Process Theory

Simmel introduced the idea that all social phenomena are the collective result of transactions between individuals (Simmel and Levine 1972). He observed that there is no universal law or force compelling any social act from the top down, it is solely created from the bottom up by individuals acting under their own volition. Cederman (2005) sees Simmel as the progenitor of the *science of complexity*, the search for simple rules (also called *generative rules*) that underlie complex phenomena. Phelan (2001) differentiates between traditional and complexity science:

> Traditional science seeks direct causal relations between elements in the universe whereas complexity theory drops down a level to explain the

> rules that govern the interactions between lower-order elements that in aggregate create emergent properties in higher-level systems. (p. 128)

Cederman (2005) builds on this concept in defining *social process theory*:

> The sociological process approach starts with an observed social phenomenon, whether unique or ubiquitous, and then postulates a process constituted by the operation of mechanisms that together generate the phenomenon in question. (p. 867)

The processes that operate together to generate a phenomenon are often called *social forms*.

The blog is a prime example of an arena dominated by the dynamic of complexity: the blog author can post content but cannot compel response. Readers are moved to respond in a manner reflecting the diversity of ways they interpret the blog entry, possibly inspiring a cascade of responses from more readers in a cycle that repeats until readers no longer have anything to write. In the remainder of the chapter, six social forms that previous research has found to govern collective action are discussed. These six complex processes do not constitute an exhaustive list, but are merely the processes that previous research suggests might be most operant and important to communities in the blogosphere.

The Echo Chamber

Scoble and Israel's (2006) "echo chamber" effect refers to the illusion of a vibrant community that frequent communication between a few parties can create:

> Blogging can fool you. You may think you are conversing with the world, when it's just a few people talking frequently, back and forth to each other, creating the illusion of amplification. The echo chamber can deceive a business into thinking it is either more widely successful or further off the mark than it is in reality, because a few people are making a lot of noise. (p. 134)

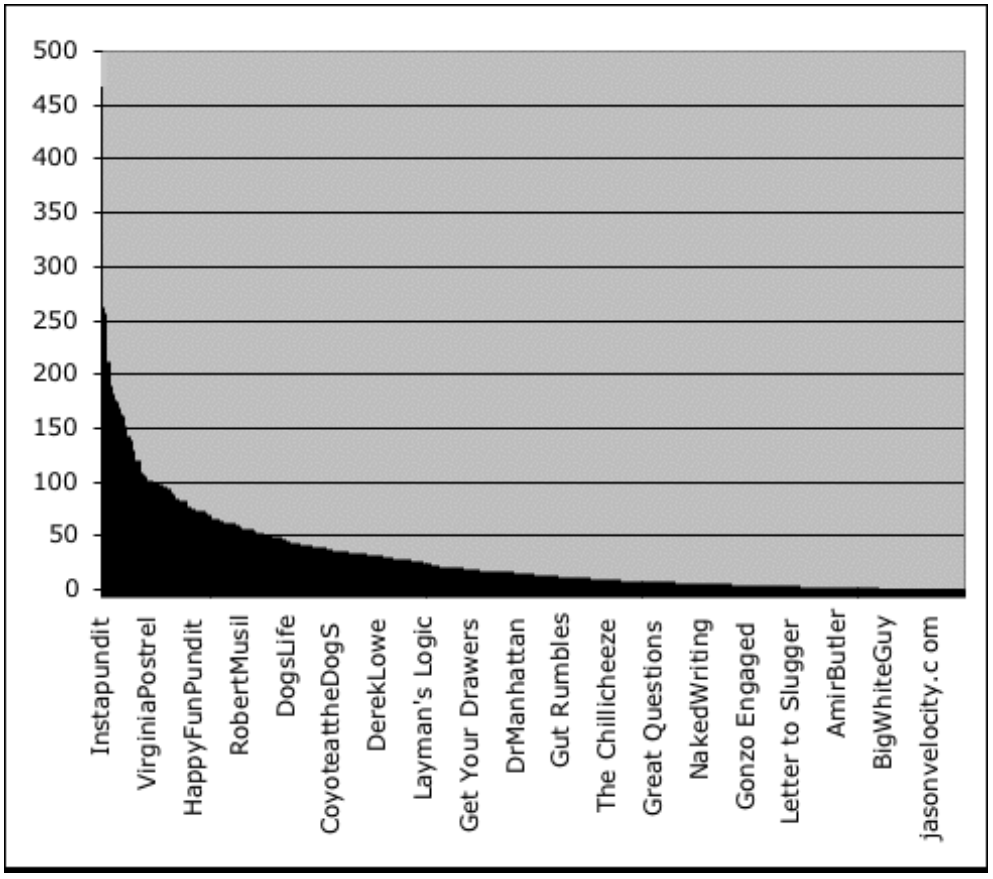
Three similar phenomena have been addressed in academic research under the terms: *cultural tribalism*, *groupthink* and *group polarization*. Kitchin (1998) described cultural tribalism in this way:

> … communities based upon interests and not localities might well reduce diversity and narrow spheres of influence, as like will only be communicating with like. As such, rather than providing a better alternative to real-world communities cyberspace leads to dysfunctional on-line communities … (p. 90)

Cultural tribalism is thus portrayed as the ultimate equilibrium condition of all online communities. Since the cost of trial and switching are low, people will sample a large number of communities and migrate to the ones wherein they feel most at home, those where they hear enough of what they want to hear to feel cognitively at ease. Such groups are likely to be small, due to an intolerance of dissent, albeit close-knit. Conversation may be lively, but always among the same people, sharing a limited range of thoughts. Therefore, this migration to comfortable cognitive spaces causes cognitive diversity to be sacrificed. If cultural tribalism is as strong an influence as Kitchin (1998) opines, then communities characterized by diverse thought are likely to diminish over time with high cognitive diversity being only temporary.

FIGURE 2.3

The Power-Law Distribution of Attention in the Blogosphere



This argument is supported by Shirky (2003), whose meta-analysis cites the findings of several researchers who observed from different perspectives that attention metrics (e.g., numbers of readers, comments and trackbacks) in the blogosphere tend to follow a *power-law* distribution similar to that depicted in Figure 2.3. Only a few very popular blogs attract the attention of a large number of people while most blogs, occupying the *long tail* of the distribution, attract a niche of people whose unique preferences closely match the blog's content.

The discussion thus far might well lead one to wonder whether the mechanism that underlies cultural tribalism is solely one of random drift to centers of thought sameness. Wallace (1999) offers another perspective as he describes a different situation that will lead to low diversity in ideas expressed:

> Each person can share what he or she knows with the others, making the whole at least equal to the sum of the parts. Unfortunately, this is often not what happens…. As polarization gets underway, the group members become more reluctant to bring up items of information they have about the subject that might contradict the emerging group consensus. The result is a biased discussion in which the group has no opportunity to consider all the facts, because the members are not bringing them up… Each item they contribute would thus reinforce the march toward group consensus rather than add complications and fuel debate. (pp. 81-82)

Wallace (1999) uses the term "polarization," but is really describing a similar phenomenon called groupthink. The mechanisms of the two phenomena are similar but the outcomes differ. In groupthink, a group's desire for unanimity causes people to withhold information that might contradict what they perceive to be a growing group consensus (Janus 1972). In group polarization, people gradually come to advocate extreme versions of their personal beliefs causing the group to divide into more homogeneous sub-groups along well-demarcated ideological lines (Moscovici and Zavalloni 1969). The two processes differ in their details, but both involve individuals compromising their personal knowledge or beliefs to achieve a form of collective unity. It is possible therefore that cultural tribalism is a superficial condition that does not truly indicate the diversity of thought among its members, it may be the result of a deliberate seeking of collective unity among otherwise diverse thinkers.

However, Matz and Wood (2005) found that heterogeneous attitudes create dissonance or tension and discomfort between members of a group. They found that the level of such discomfort was proportional to the numerical minority status of those with atypical views. The discomfort is partly relieved if the minority view-holders felt free to affirm their attitudes without pressure to conform. But, true relief of discomfort was only achieved if the minority could persuade the majority of their error, the majority could present more convincing support for their position, or if the minority could leave and join a more compatible group. It should be concluded then that compromising personal values to achieve collective unity is not a sustainable situation in the blog context as unity is not a forced necessity. The migratory aspect of cultural tribalism, given its ease among Internet blogs, seems to be the result of a natural, unavoidable coping mechanism.

Flocking Theory

Reynolds (1987) proposed *flocking theory* as a computational model that explains how the coordinated movement of a group can emerge from individuals making decisions based on personal information. He discovered that by using three simple "steering behaviors," coordination would emerge without any explicit management activity: *separation* ("steer to avoid crowding local flockmates"), *alignment* ("steer toward the average heading of local flockmates"), and *cohesion* ("steer toward the average position of local flockmates"). These steering behaviors are really just heuristic components of the more complicated calculation of which direction an individual should move to be in its desired location, relative to the group, one unit of time in the future.

Although flocking theory was developed as a solution to modeling the behavior of flocking birds and animals in computer graphics, it has been extensively investigated in a variety of academic disciplines and mathematically modeled by physicists, Toner and Tu (1998). Rosen (2002) proposed that flocking theory was a good explanation for self-organization in human social systems. He proposed that communication was the mechanism of cohesion in human society where a social network of individuals shares access to a collective body of knowledge that acts as a "roadmap" for coordinated action with little centralized control.

Rosen (2002) based his model on multiple literature streams. Simmel and Levine (1972) state that for social relationships to occur "the personalities must not emphasize themselves too individually … with too much abandon and aggressiveness." Eisenberg and Phillips (1990) proposed that community cohesion is always a balance between autonomy and interdependence, corresponding to Reynold's (1987) separation and cohesion steering behaviors. Rosen (2002) concludes that to some extent uniformity and common interest is essential to flock maintenance and that individuals must sacrifice some autonomy to keep group acceptance.

From the discussion thus far, flocking and the unity-seeking compromises of groupthink and group polarization may seem not to be distinct concepts. The important difference is that people who engage in groupthink and group polarization are motivated by a perceived need for collective unity, while people who engage in flocking ("flockers") are motivated by a personal need for social belonging. Additionally, in the blogosphere, nothing constrains a flocker from engaging in a cultural tribalism-style

migration to another blog if the discomfort felt from compromise exceeds the value gained from social acceptance. So the flocker's compromise probably causes less dissonance than that felt by the migrating tribalist.

Threshold Models of Collective Behavior

Granovetter (1978) criticized previous research on crowd behavior for being too accepting of a universal model of individual compliance with social norms. He used as an example the situation of a riot where individuals who would never start a riot would participate nonetheless if sufficient other individuals joined in. He proposed that everyone had a threshold for such behavior and that if the diversity of individual thresholds were uniformly present in a crowd, a riot, once started, would propagate throughout the group getting everyone to participate. This model can be readily applied to a blog, where a few readers might have something they would like to say but are too shy to be the first to comment. Once some people who are less shy express their thoughts, the shyer will feel comfortable enough to post their comments.

Memetics

Marsden (1998) called memetics the study of *mind viruses*, ideas that are highly contagious and quickly spread through a population. The term "memetics" is derived from *meme*, a word introduced by Dawkins (1976) in his classic book *The Selfish Gene*. Dawkins (1976) was primarily concerned with arguing that most human behavior is directed toward genetic propagation. However, as Blackmore (2001) observed, he wanted to show that the core processes of *evolution* (mutation, propagation and natural selection) are not confined to biological genetics but also to other *replicators* such as

ideas. Through memes or systems of related memes (what Blackmore called *memesets*), humanity passes certain cultural practices from generation to generation because these practices increase survivability. Marsden (1998) described a three stage process where acquired ideas ("you get an idea") become ideology ("the idea gets you") resulting in specific behavior. It is readily apparent that this mechanism is superior to forcing people to behave in a certain way because when the force stops the behavior will probably stop. However, if a person can be persuaded as to the value of a certain behavior, then that behavior is much more likely to be continued indefinitely. Just as not everyone is equally susceptible to catching a biological virus, there are various degrees of resistance and immunity to memes. Dawkins (2006) described two strategies that have evolved to maximize the success of meme propagation: *intense indoctrination* and *mass dissemination*. During intense indoctrination, the target person or population is subjected to the high repetition of a meme, or memeset, in an attempt to overwhelm thresholds of resistance. Indoctrination, in its attempt to suppress the opposing influence of *critical thinking*, often involves the exclusion or ridicule of competing memes. In mass dissemination, the targets are the people with the lowest barriers to adoption, so the idea is spread as widely as possible to increase the probability of reaching people with a low threshold of resistance.

It may seem from the discussion thus far that there is little difference between contexts influenced by intense indoctrination and those influenced by cultural tribalism and flocking. Cultural tribalism and flocking are purely grassroots phenomena, while indoctrination is driven by top-down intention to persuade others. Ultimately, to be

successful, efforts at indoctrination must inspire grassroots support. People most susceptible to, and accepting of, the indoctrinated meme will cluster (cultural tribalism) while others, motivated to be accepted by the core, flock around the periphery.

Need-for-Cognition

Thus far in the discussion, two innate motivations have been discussed along with their resultant behaviors: the desire for uncompromised self-expression, revealed in cultural tribalism, and the desire for acceptance, revealed in flocking. It is apparent that neither of these two motivations contributes to the creation of diverse thought within a blog as both are associated with situations where diversity of thought will be either reduced or constrained. For blogs to be centers of diverse thought, influences must be present that tend to increase the expression of diverse perspectives. As already stated, the focus of this study is detecting diversity in attitudes toward organizations, brands and products. It is proposed that the best place to see what attitudes are present is one where attitudes are formed and modified. The blog context is likely to be such a place as blog authors, even without exercising such extreme methods as indoctrination, are likely to use persuasive language intended to affect attitudes. Petty and Cacioppo (1986) proposed one of the most accepted models of attitude change: the Elaboration Likelihood Model (ELM). The concept of an *elaboration continuum*, where thought among targets of persuasion varies from low to high, is central to the ELM. Part of what determines the level of thought exercised by targets of persuasion is their innate *need-for-cognition*, an enjoyment of cognitively demanding tasks. A community of blog commenters that reflects the full diversity of the market should exhibit the full spectrum

of the elaboration continuum in their commenting activity. Commenters at higher need-for-cognition levels should be a grassroots influence that increases the level of thought diversity expressed in a blog.

Thus far in this discussion, need-for-cognition is portrayed as an internal motivation that inspires thinking. Neither the psychological nor the marketing literature connects need-for-cognition with social behavior. However, past research has described the phenomenon of *sensemaking* as an internal activity that inspires collaborative social activity. In the next section, sensemaking is described and need-for-cognition is suggested to be its underlying motivation.

Need-for-Cognition and Sensemaking

Sensemaking, as the word implies, is an act of making sense of circumstances. It has already been noted that communicated messages are often perceived differently than intended by a sender and differently again among receivers because of a noisy transmission medium or because of differences in the interpretive repertoire of senders and receivers. Thus, it should be no wonder that blog entries and comments might be a source of reader confusion. The conceptualization of sensemaking in the blogosphere used in this dissertation is based on a framework developed by Weick, Sutcliffe and Obstfeld (2005) to explain how sensemaking occurs, and is a formative process, in organizations. The association of organizing with blogging is particularly apt because both are contexts where individuals react to their understanding of current conditions and thereby shape the evolution of a community. Weick et al. describe sensemaking as an ongoing individual process of retrospection where a person asks “what is going on

here?" and then uses the explanation they consider most plausible as a basis for action. It seems reasonable to propose that the depth of such questioning and subsequent inquiry should be related to an individual's need-for-cognition. The action resulting from an individual's sensemaking changes the context, inspiring others to enter the sensemaking process to update their understanding of what's going on before they perform their own actions. Hence, individuals enter and leave the sensemaking process in an ongoing cycle driven by the actions of community members. Weick et al. see the sensemaking process as a response to surprising or unexpected actions; however this study sees the element of surprise as only affecting the degree of sensemaking an individual undergoes in response to change. Since there is always a mismatch between how an act is intended and perceived, all action contains an element of the unexpected.

Weick et al. see sensemaking as starting with a noticing of something "at variance with the normal" (also described as "disruptive ambiguity") that is then investigated and "labeled" so as to become part of the common currency for communicational exchange. The investigation begins with a personal retrospective search of past experience in the person who first notices the novel event. The result of that personal search is the initial labeling. Weick et al. see sensemaking as always a social process where the initial labeling facilitates discussion with others where the sensemaking continues and propagates, possibly resulting in a change of the labeling: "thinking is acted out conversationally." Behind the sensemaking is always a consideration of response: "what should I do?" Action and talk thus become a cycle. Even though every element of conversation in a blog is preserved, not every thought

expressed becomes the impetus for new thought expression. The more unexpected the thought, the more it is likely to become part of a chain of collective learning.

In the marketing literature, Rosa et al. (1999) show how product markets are knowledge structures that are socially constructed through a collaborative sensemaking process among producers and consumers. In this process, consumers and producers tell each other stories in published media: the producers convey what the products can do; the consumers reveal what they need and their experiences with products. Through the process, consumers and producers converge on a common understanding of what a product category can and should do. The blog should be an ideal platform for the sensemaking process because it is publicly visible and designed to be an interactive conversational tool.

Cognitive Diversity

In the preceding sections on cultural tribalism, flocking, memetics and need-for-cognition, their relevance to cognitive diversity (the diversity of thought expressed in a blog) has been the focus of discussion. For this reason, the focus of this study is on cognitive diversity as a construct. This is not the first study to use cognitive diversity as a focal construct. Recent works by Page (2007) and Surowiecki (2004) extol the virtues of cognitive diversity in groups of all kinds: diverse thinking leads to better collaborative outcomes. Page (2007) recognizes a dual nature to diversity in cognition: diversity in preferences and diversity in thinking styles, or as he calls them, *cognitive toolboxes*. Page also notes there is some degree of interdependence between preferences and

toolboxes since people develop the thinking skills needed to satisfy their preferences and then change their preferences based on their new ways of thinking (pp. 285-296).

Reciprocity and Game Theory

A blog is a *public good*, depending on the participation of a community for value, but not diminished in value by the number of readers. In a blog, where the author posts (a costly effort) in anticipation of response (a payoff), there is always uncertainty about the amount of response the author will receive from any post in return for his cost of posting. This is an example of a trust-reciprocity scenario, shown by Pereira, Silva and Andrade e Silva (2006) to be robust in certain situations:

> … *reciprocity* is a pattern of behavior that may be considered 'robust' because it emerges in specific environments where the marginal costs of such behavior are reasonable, but not prohibitive. (p. 420)

Since the unique feature of a blog is that a single author controls the discussion, each blog entry is like a sequential game where the author moves first. As already stated, this first move is in expectation that the value received from all the resultant community contributions will exceed the cost of posting. If Pereira et al's conclusion is applied to blogs, then it can be supposed that if the value of the author's post exceeds the cost of reciprocal response, community members will reply under the pressure of a sense of obligation to reciprocate equal to the difference between value received from the post, and the effort expended to contribute a comment. The author's original post plus the cascade of member contributions will continue to raise the value received by the community, increasing the level of obligation felt by those that have not contributed, inspiring more contribution as thresholds of tolerance for withheld reciprocation (i.e.

cost of reciprocation) are exceeded. Some community members may have very high levels of tolerance or costs of reciprocation and be *free riders* in any specific game. While their receipt of benefits does not diminish the value of existing content to the rest of the community, it does cause the maximum value of content to peak at less than what could result from full participation. While the author's post might be seen as an act of altruism, it is rational to assume the author will eventually stop trying to initiate a response if repeated attempts generated less return than the cost of posting. This model's "thresholds of tolerance" is similar to the type of model proposed by Granovetter (1978).

Summary

This chapter provides a review of the principal streams of literature relevant to this research study. Brand image is encoded in the minds of consumers in a variety of ways. Stimuli, similar to that provided by a blog author's entry, evoke free-form responses from consumers that reveal their brand associations. These responses can be mapped to a multidimensional space where similarity between cognitive objects can be assessed by looking at their spatial proximity. Various social processes have been described that purport to explain patterns in the spatial proximity of cognitive objects, including social process theory, which ascribes these patterns to the interaction of many social processes. This review suggests that social process theory offers a credible explanation for the way people contribute to blogs and the diversity of thought they express. The next chapter builds on this conclusion and proposes a conceptual model and hypotheses to test it.

CHAPTER III

CONCEPTUAL MODEL AND HYPOTHESES

As previously stated, this research study investigates the extent to which postings and comments in blogs reflect a company's image in the overall market. This research study seeks to lay a foundation for answering the above by testing for the effect of six social processes (or forms), identified in prior research and detailed in Chapter II, that are known to have an effect on the diversity of thought expressed in a group forum.

First, this chapter presents a discussion on this study's dependent variable, a latent variable construct representing cognitive diversity, that is, diversity of thought expression. Next, six latent independent variable constructs are described. Each represents one of the social processes described in Chapter II. Two conceptual models, hypotheses and conceptual support for the hypotheses are presented.

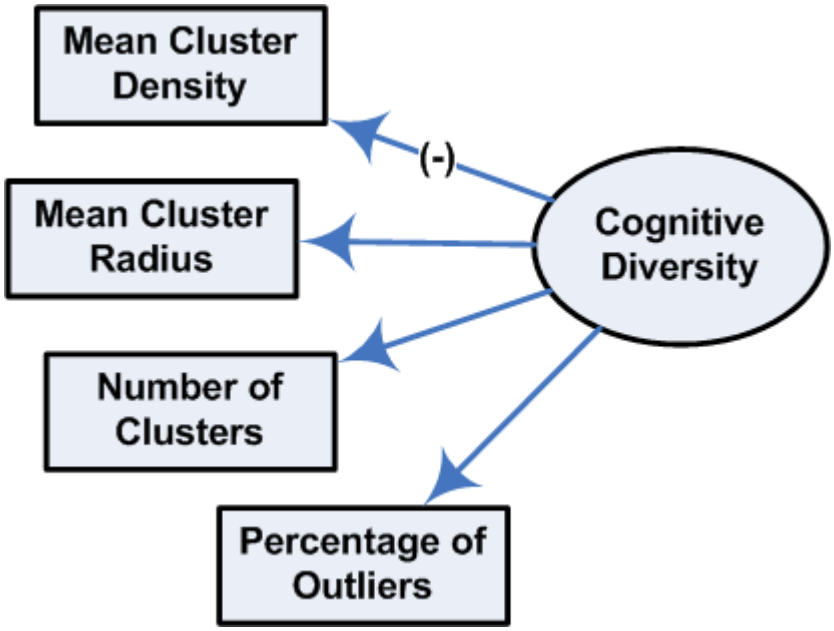
Cognitive Diversity

Once multidimensional scaling has located blog entries and their associated comments into stable positions within cognitive space, thematic clusters can be detected. In Chapter II, four cognitive diversity metrics were introduced from measures of thematic clusters: the number of clusters associated with a blog entry, their mean radius and density, and the percentage of outliers. As depicted in Figure 3.1, these metrics can be used as indicator variables for the diversity of thought (i.e., cognitive diversity) latent variable. The influences on diversity of thought are all formative effects and are defined as independent variables in the sections to follow. As stated in Chapter I, blog authors

write posts or entries that prompt commenters to write replies. Since it is the relative locations of comments to each other and their associated blog entry that are the basis for this study's measures of cognitive diversity and the constructs hypothesized to affect it, the blog entry is the unit of analysis, unless otherwise noted.

FIGURE 3.1

Cognitive Diversity Latent Variable

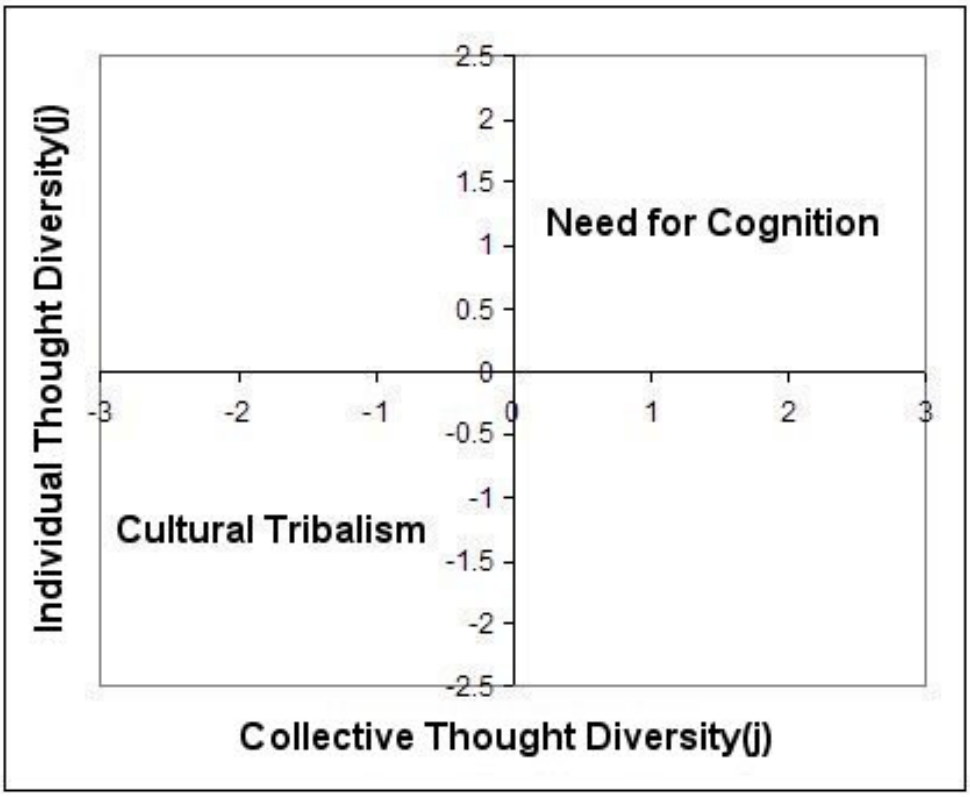


Cultural Tribalism, Need-for-Cognition and Cognitive Diversity

In Chapter II, cultural tribalism was described as a process whereby like-minded people are brought together by a desire to feel validated by the exercise of uncompromised self-expression. Since this is generally only possible among people who agree, it results in the creation of contexts were members only hear their own limited ideas and perspectives echoed by each other. Need-for-cognition was described as a competing motivational influence where people who like intense thinking are attracted into the blog conversation to engage in high intensity cognitive tasks (like sensemaking) and thereby expand the diversity of thought expressed. As explained more fully next, this study conceives of these two competing motivations as occupying opposite quadrants of a conceptual space defined by variations in *individual thought diversity* (i.e., the variety across an individual's past and current idea expressions) and *collective thought diversity* (i.e., current variety across different individuals) as shown in Figure 3.2.

FIGURE 3.2

Cultural Tribalism and Need-for-Cognition as Opposing Motivations



In Chapter I, it was mentioned that blog monitoring can be part of a feedback loop, allowing companies to monitor and adjust their self presentation. As such, blog monitoring facilitates organizational learning. March (1991) argues that cognitive diversity is the key to such knowledge development because it results in new alternatives being applied to problem solving rather than a constant recycling of old solutions. March opines that cognitive diversity is achieved by the constant influx of new voices; these new people will not have developed enough attachments with or within the group to be overly concerned with achieving cohesion or forming alliances. Since cultural tribalism

causes people to form long term associations with communities wherein they feel most comfortable, it is reflected in the mean longevity, or mean time between the first comment and the current comment, of commenters to the current blog entry ("Mean Commenter Longevity") and the number of commenters who commented on the last blog entry ("% Repeat Contributors (T-1)"). The latter indicator captures a recency effect; that is, it is assumed that if a person commented on the last blog entry, then they are more likely to comment on the current blog entry and thus lower the diversity in thought expression for the whole community. This assumption is based on Holme's (2003) finding that ties to virtual communities decay over time with decreasing contact. A recent comment is evidence of current membership and therefore a strong predictor of further comments in the immediate future.

However, long term associations are not sufficient to distinguish cultural tribalism from the other influences that affect cognitive diversity. Commenters that receive value from satisfying their need-for-cognition can also be expected to be regular commenters. As described in Chapter II, cultural tribalism is predicted to cause closer-knit relationships between people who express similar views. As a result, Figure 3.3 also depicts cultural tribalism as reflected by a negative relationship to: (1) the mean entry-to-comment and comment-to-comment cognitive distance (denoted "Mean Collective Thought Separation" in Figure 3.3), (2) the standard deviation of entry-to-comment and comment-to-comment cognitive distance across blog entries (denoted "Stdev Collective Thought Separation"), (3) the mean cognitive distance between an individual's comments across blog entries (denoted "Mean Individual Thought Separation"), and (4)

the standard deviation of the pair wise cognitive distances between an individual's comments across blog entries (denoted "Stdev Individual Thought Separation"). Need-for-Cognition is depicted in Figure 3.3 as reflected by positive relationships to the same indicators. Thus,

H_1: *Cultural tribalism* is negatively related to *cognitive diversity*.

H_2: *Need-for-Cognition* is positively related to *cognitive diversity.*

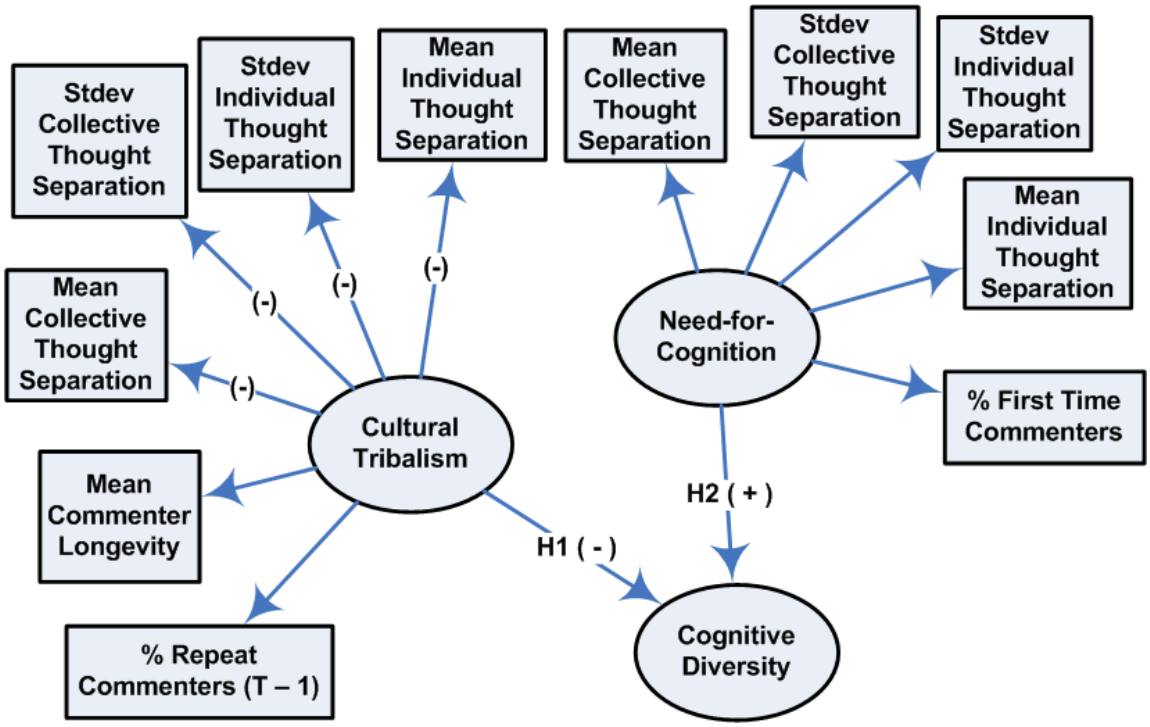
Hypotheses 1 and 2 are reflectively modeled in Figure 3.3. In Figure 3.3, the unit of analysis is the blog entry.

Blog entry comments primarily influenced by cultural tribalism can be distinguished from those primarily influenced by need-for-cognition by a discriminant model (equation 3.1) that differentiates on the basis of diversity of thought expressed using similar metrics to those depicted in Figure 3.3.

(3.1) $$Z_j = \beta_0 + \beta_1 \overline{d_C} + \beta_2 \sigma_C + \beta_3 \overline{d_I} + \beta_4 \sigma_I$$

FIGURE 3.3

Cultural Tribalism / Need-for-Cognition Reflective Sub-Models



In equation 3.1, the unit of analysis is the blog entry comment, Zj is the discriminant Z score of the discriminant function j, $\overline{d_C}$ is the mean entry-to-comment and comment-to-comment cognitive distance across comments to the current blog entry (a limited form of the metric denoted "Mean Collective Thought Separation" in Figure 3.3), σ_C is the standard deviation of entry-to-comment and comment-to-comment cognitive distance across comments to the current blog entry (a limited form of the metric denoted "Stdev Collective Thought Separation" in Figure 3.3), $\overline{d_I}$ is the mean comment-to-comment cognitive distance for the current commenter across blog entries

(a limited form of the metric denoted "Mean Individual Thought Separation" in Figure 3.3) and σ_I is the standard deviation of comment-to-comment cognitive distance for the current commenter across blog entries (a limited form of the metric denoted "Stdev Individual Thought Separation" in Figure 3.3).

The parameters of equation 3.1 can be used to distinguish between contexts primarily motivated by cultural tribalism and those motivated by need-for-cognition. To conceptually distinguish between cultural tribalism and need-for-cognition, the parameters of equation 3.1 can be divided into two groups to create sub-factors that represent *individual thought diversity* (mean and standard deviation in individual thought separation) and *collective thought diversity* (mean and standard deviation in collective thought separation). These sub-factors are used as axes of the perceptual map shown in Figure 3.2. If blog entry comments are plotted on this perceptual map, their type should be readily distinguishable by their presence in one of the marked quadrants. Since the unit of analysis of equation 3.1 is the comment, comments must be added to Figure 3.2 one-by-one in the order they were created. Individual and collective thought diversity coordinates would be calculated at each step using the comment being added, and its thematic relationship to the comments and blog entry already present.

Threshold Models of Collective Behavior and Cognitive Diversity

In Chapter II, Granovetter's (1978) model of a generative process was described as a process that builds on itself. Granovetter used his model to elaborate an example scenario that assumes everyone has a threshold for violent behavior. If the full variety of individual thresholds were uniformly distributed within a crowd, a riot, once started,

would propagate throughout the group getting everyone to participate. The idea behind this process is used in this study to propose a mechanism for the emergence of divergent themes within the comments to a blog entry. In this situation, the existence of a class of commenters, denoted *idea seeders* is predicted. Idea seeders are the riot-starters of a blog. They are among the most uninhibited, and therefore the most prolific, commenters. They are willing to express ideas on the most diverse array of themes. Just as riot participation spreads from the originators to the rest of a crowd in a cascading fashion, idea seeders are consistent starters of new thematic clusters; their willingness to express themselves emboldens others with similar ideas to write comments, cascading through the readership until all like thoughts are expressed. As the number of commenters increase, the probability of having an idea seeder among them likewise increases. Thus,

H_3: *Idea seeding* is positively related to *cognitive diversity.*

Flocking Theory and Cognitive Diversity

In Chapter II it was noted that Rosen's (2002) proposal that flocking theory, a model explaining how the coordinated movement of a group can emerge from individuals making decisions based on personal information, was a good explanation for self-organization in human social systems. Rosen concluded that, to some extent, uniformity and common interest is essential to flock maintenance and that individuals must sacrifice some autonomy to keep group acceptance. In a blog conversation, consistency of theme is created and preserved by commenters controlling how far their comments diverge from the overall group theme. Commenters balance their desire to express unique, individualistic ideas with the need to conform to the conversational

direction of the group. They thus take the conversation in the direction they perceive the flock is already going. Thus,

H_4: *Flocking* behavior is negatively related to *cognitive diversity.*

It is proposed that the flocking process is closely related to the threshold behavior discussed earlier, in that idea seeders are the emergent flock leaders. Their comments initiate a cascade of other comments whose thematic consistency exhibits the flocking phenomenon. There is therefore some probability that there will be two classes of commenter in a blog entry conversation: idea seeders who often lead the flock and those who primarily follow.

An estimate of the effect of flocking and idea seeding on a blog entry is determined in a three step process: separate idea seeders from flock followers (hereafter denoted *flockers*), associate the classified individuals with the blog entries they commented on and then estimate the value of the flocking and idea seeding latent factors from the participation history of the associated commenters.

Idea seeders can be separated from flockers by a discriminant model (equation 3.2) that differentiates between commenters on the basis of diversity of thought expressed. Commenter type is assessed on the basis of how prolific their comment activity is (C), the mean cognitive distance ($\overline{S}$) between their comments across blog entries (widest for idea seeders), the standard deviation of the cognitive distances (σ_S) between their comments across blog entries (largest for idea seeders), the number of new thematic clusters they started (K) and the standard deviation of the cognitive distances

(σ_K) between their comments and the centroids of the closest thematic clusters (widest for idea seeders) across blog entries.

(3.2) $$Zj = \beta_0 + \beta_1 C + \beta_2 \overline{S} + \beta_3 \sigma_S + \beta_4 K + \beta_5 \sigma_K$$

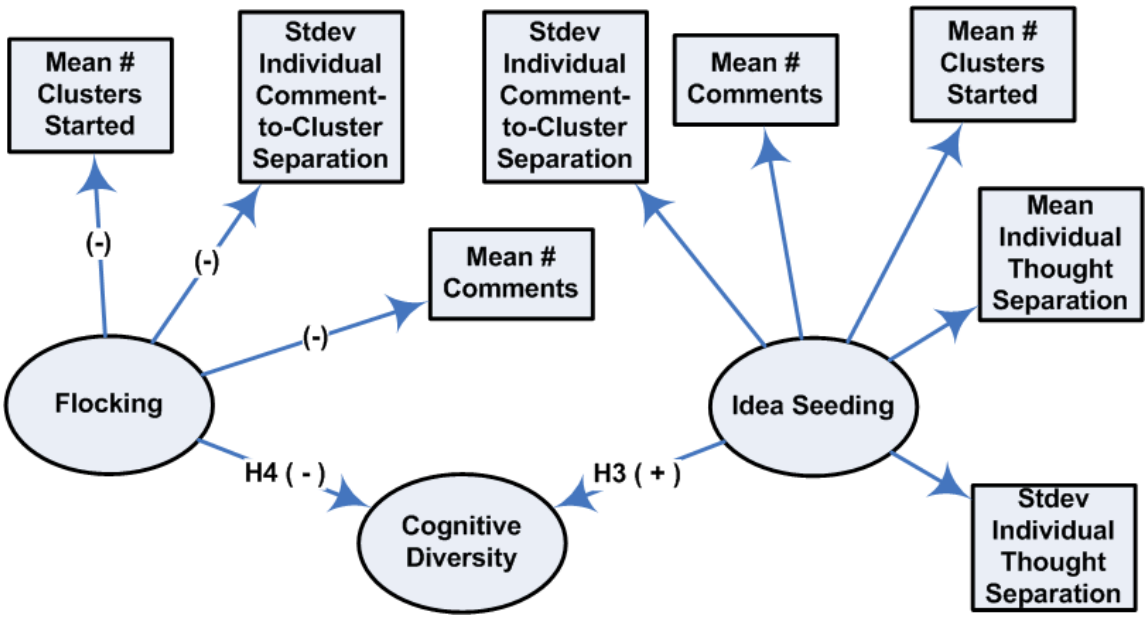
In equation 3.2 the unit of analysis is the individual commenter, *Zj* is the discriminant Z score of the discriminant function *j*, and the rest of the parameters are as above. It should be noted that the mean cognitive distance between comments and the centroids of the closest thematic clusters across blog entries is not expected to differentiate between idea seeders and flockers because, as described in Chapter II, flockers can have wide variety in their preferred distance from the rest of the flock. The distinguishing factor is that there will be a high level of consistency in this distance. That consistency is fully captured in the standard deviation of the cognitive distances between their comments and the centroids of the closest thematic clusters across blog entries.

Once commenters are separated into idea seeders and flockers, then the influence that each type exerts on the cognitive diversity of a blog entry can be modeled as shown in Figure 3.4. In Figure 3.4, all the parameters of equation 3.2 are useful in assessing the relationship between idea seeding and cognitive diversity. The relationship between flocking and cognitive diversity is fully captured by using three indicators: the mean number of clusters started, the standard deviation of individual comment-to-cluster centroid distances and the mean number of comments written by commenters to the current blog entry. This is because flockers are only characterized by being very consistent in the degree their thought expression over time differs from that of the group

and by being more inhibited in their thought expression as a result of being more careful. Therefore, flockers will comment less and start fewer new thematic clusters. Variation in the cognitive distance between their comments across blog entries is not important because that is completely dependent on the thematic perspective of the flock they belonged to when they make comments to the current blog entry.

FIGURE 3.4

Flocking and Idea Seeding Reflective Sub-Models

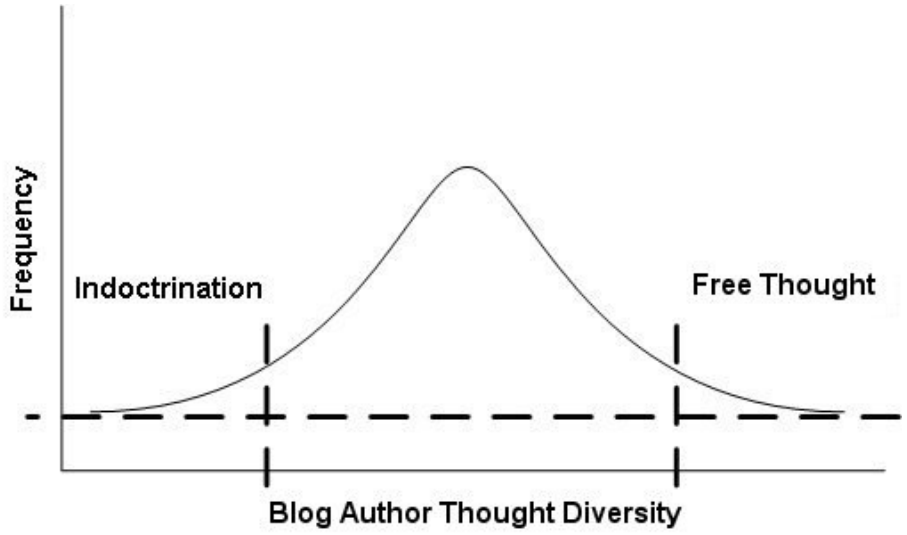


Memetics and Cognitive Diversity

In Chapter II, memetics was described as the study of mind viruses (i.e., memes). It was mentioned that Dawkins (2006) described two strategies that evolved to maximize the success of meme propagation: intense indoctrination and mass dissemination. Indoctrination uses repetition of the same meme or system of related memes (a *memeset*), often while simultaneously excluding or ridiculing competing memes, in a closed cognitive context (in this case, a blog) to overcome resistance, increase acceptance and spread of the meme ideas. Mass dissemination spreads a meme or memeset in as many cognitive contexts as possible, looking for a sympathetic audience of potential converts with little resistance to the meme ideas. Note that, in the blogosphere, it is not suggested that mass dissemination occurs under the organizing influence of some central control. The personality and enthusiasm of individuals drive them to spread the memes they have embraced in a wide variety of cognitive contexts. This is proposed, in addition to flocking, to be another example of emergent behavior, individuals behaving in the same manner, under their own volition, who inadvertently create the appearance of collective order.

FIGURE 3.5

Spectrum of Blog Author Thought Diversity

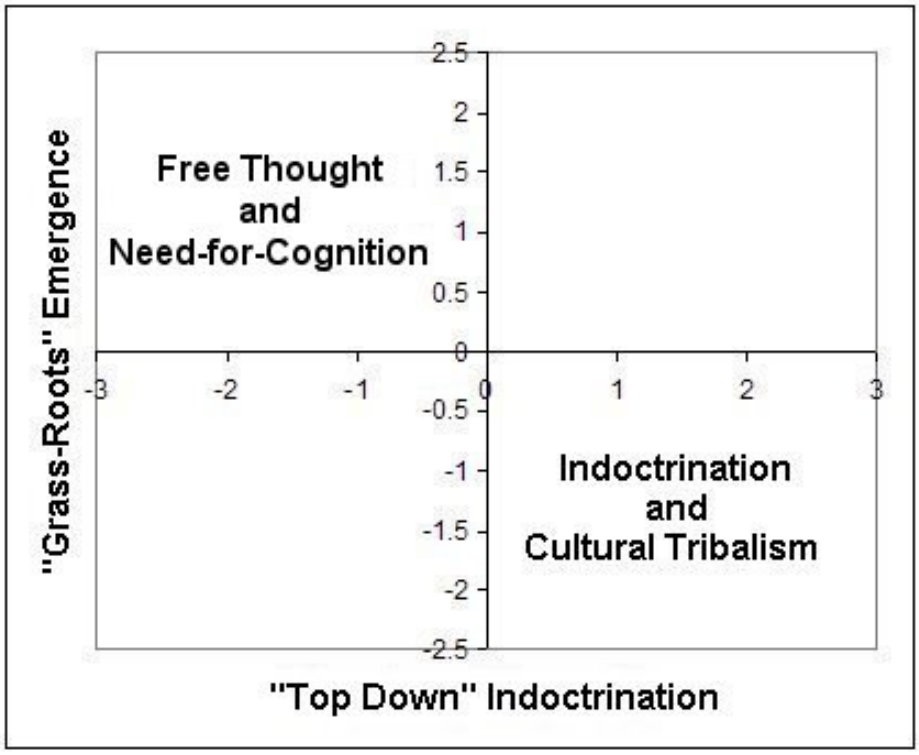


Alternatively, as noted in Chapter II, indoctrination, as a top-down influence, can only occur under the influence of a central control. In the blogosphere, that control can only be a blog author, whose cognitive distance and, probably, time interval between blog entries are less than average across the blogosphere. Since indoctrination is, by definition, an extreme behavior it is proposed that blog authors occupy a spectrum of thought diversity, and by implication, tolerance of diverse thought, as depicted in Figure 3.5. Throughout this discussion of indoctrination, it was assumed that there was some intent on the part of the blog author to propound a limited set of views. It is possible though that the appearance of indoctrination might be created by a blog author who watches the blog to see what topics garner the most discussion and then repeatedly introduces those topics with no intent but to keep the conversation lively.

In Chapter II, indoctrination and mass dissemination were both connected to Granovetter's (1978) thresholding dynamic. It was reasoned that indoctrination seeks to overcome thresholds of resistance to meme ideas through repetition in a closed context, while mass dissemination seeks to find people with low thresholds of resistance in a variety of contexts (e.g., those characterized by both cultural tribalism and need-for-cognition). Just as rioting behavior can spread through a crowd as differing individual thresholds of riot resistance are overcome by observing others' rioting behavior, so meme acceptance can be expected to spread through any group as the unconverted see others in their vicinity become converted. This cascading propagation of meme acceptance is expected to be aided by flocking behavior. People with flocking propensity will sense if their local group's thematic direction has been affected by the acceptance of a meme or memeset. In order to preserve group acceptance, they will control the extent to which their expressed ideas diverge from their perception of the group norm, thus aiding in memetic propagation.

FIGURE 3.6

Locus of Influence Conceptual Map

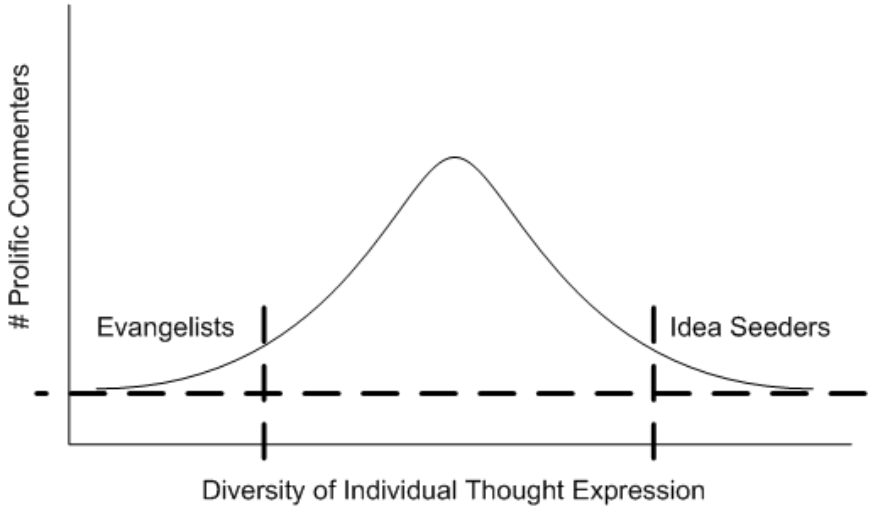


On the basis of the foregoing, it is further argued that indoctrination and mass dissemination can be opposing means of memetic propagation. An individual attempting to disseminate a competing meme in an indoctrination context will face the opposition of an entrenched set of memes, and a process of constant reinforcement of those existing memes, that will work to suppress the propagation of a new meme. If the new meme does manage to take hold and propagate, the commenters who embrace the new meme will probably migrate, as a cultural tribe, to a different blog context that may indeed be an indoctrination arena for the new meme. Recall in Chapter II, that indoctrination was

denoted as a top-down influence while cultural tribalism was denoted a "grassroots" phenomenon. It was noted that for indoctrination to have any effect it must garner grassroots support. It is argued that indoctrination will more often gain its grassroots support from cultural tribalism, while blog author free thinking will most often gain its support from need-for-cognition. These proposed relationships are depicted as a conceptual map in Figure 3.6.

FIGURE 3.7

Evangelists and Idea Seeders

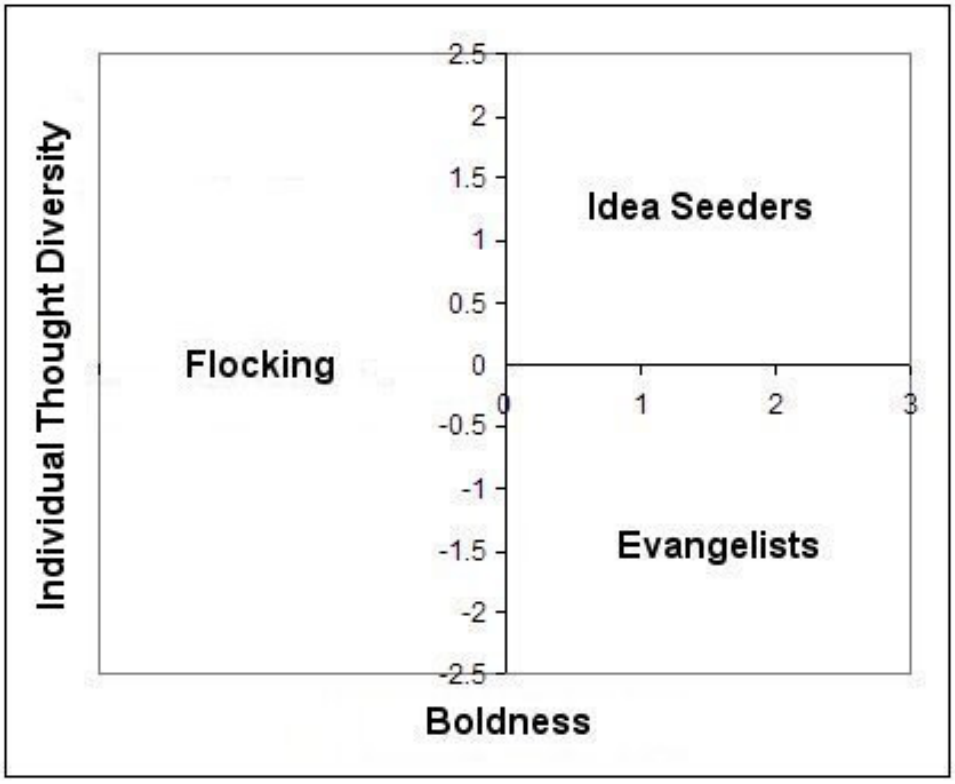


As already stated, mass dissemination is manifested as an individual, rather than a systemic, activity where certain prolific commenters have low diversity in thought expression as they repeat the same ideas. These commenters are denoted as *evangelists*. This description might make evangelists appear similar to idea seeders; however, idea seeders are not only prolific but diverse in the expression of their thoughts. The

distinction is made more explicit in Figure 3.7 which shows an assumed normal distribution in the diversity of expressed thought among prolific commenters; commenters in the low diversity tail are evangelists while those in the high diversity tail are idea seeders.

FIGURE 3.8

Perceptual Map of Commenter Types



The parameters of equation 3.2 can be re-estimated to distinguish between evangelists, idea seeders and flockers. To conceptually distinguish between flocking, idea seeding and evangelism, the parameters of equation 3.2 can be divided into two

groups to create sub-factors that represent *individual thought diversity* (mean and standard deviation in individual thought separation) and *boldness* (number of thematic clusters started, standard deviation in comment-to-closest-thematic-cluster separation and the number of comments contributed). These sub-factors can be used as axes of the perceptual map shown in Figure 3.8. If commenters, based on their participation history, are plotted on this perceptual map, their type should be readily distinguishable by their presence in one of the marked quadrants.

Merriam-Webster (2007) defines *boldness* as "showing a fearless, daring spirit." Roget (1962) defines it as "willingness to take risks," (p. 891) with synonyms such as "adventuresomeness" and "venturesomeness." These terms and definitions are very much in keeping with the concept of *entrepreneurship*, defined by Merriam-Webster as "assuming the risks of a business." Another similar term is *innovativeness*, defined by Roget as: "being characterized by, or productive of, new things or ideas." (p. 139) Clearly, *boldness*, *innovativeness* and *entrepreneurial* are related terms and concepts. However, *entrepreneurial* has an explicit business connotation, where *boldness* is a more generic term. This difference is important to this study, as a word denoting a *willingness to take risks in expressing ideas* is called for, despite the marketing context.

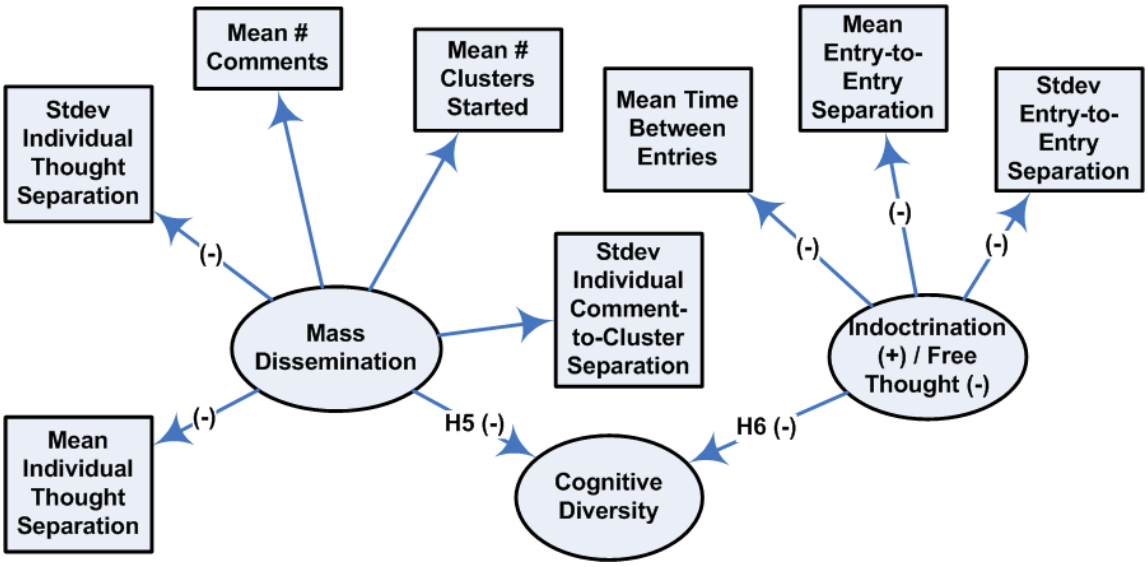
Innovativeness also does not fully capture the need as it involves producing new ideas, while idea seeders (those high in the quality for which a label is sought) need not be idea originators; they may also be conduits for ideas acquired elsewhere that they think are applicable to a particular blog conversation. Idea seeders thus enable what Pink (2005) calls *symphony*: taking little pieces of information, and weaving them together so

a broader pattern can be discerned (p. 125). Idea seeders, and later the bolder and more intellectually diverse flockers, individually bring little pieces of information, both original and from diverse sources, while the collective conversation weaves them together to find the broader pattern. This makes idea seeders a new type of influential (Katz and Lazarsfeld 1955) distinct from the concept of mavens and opinion leaders (as presented by Vernette 2004). The latter were seen as static in their information dissemination role (Watts and Dodds 2007), while these are information conduits on a more happenstance basis, dependent on what the conversation calls to mind and what their *boldness* allows them to express to the group.

Now that the relevant constructs are defined, the influence of indoctrination and mass dissemination on cognitive diversity can then be reflectively modeled as shown in Figure 3.9.

FIGURE 3.9

Memetic Propagation Reflective Sub-Models



While it should be apparent from the preceding discussion that cognitive diversity will be lower in blogs highly influenced by indoctrination, it might be thought that mass dissemination is just an isolated act of individual fanatics that can have no effect on overall thought diversity. However, it has already been noted that mass dissemination can cause memes to enjoy a threshold-overwhelming cascading propagation of acceptance, aided by flocking behavior. Therefore, both mass dissemination and indoctrination can be major influences on cognitive diversity. Thus,

H_5: *Mass dissemination* is negatively related to *cognitive diversity.*

H_6: *Indoctrination* is negatively related to *cognitive diversity.*

Reciprocity and Cognitive Diversity

In Chapter II, it was mentioned that in a blog, the author posts (a costly effort) in anticipation of response (a payoff). Equation 3.3, a variation of a model proposed by Gintis (2004), describes the value received by the author (v_A), the difference between the author's valuation of the community's contribution (B_{TA}) and his own cost (c_A).

(3.3) $$v_A = \left|B_{TA}\right| - c_A = \sum_{j=1}^{K} \left(\frac{b_j(1-\varepsilon_j)}{(1-\delta_j)}\right) - c_A$$

Where: K is the number of responders, b is the benefit the responder intended to contribute ($0 < b < 1$), ε is the responder's ineffectiveness in contributing the intended benefit ($0 < \varepsilon < 1$), δ is the author's premium (or discount) for the value of each responder's contribution ($-1 < \delta < 1$), and c_A is the author's cost of posting.

Both active (those contributing to the discussion) and passive community members receive the value from the author's post and from the contributions of active members. Equation 3.4 shows the value received by each contributing member (v_C) and equation 3.5 shows the value received by passive members (v_P).

(3.4) $$v_C = \left|B_{OC}\right| + b_{AC} - c_C = \sum_{\substack{j=1 \\ j \neq c}}^{K} \left(\frac{b_j(1-\varepsilon_j)}{(1-\delta_j)}\right) + \frac{b_A(1-\varepsilon_A)}{(1-\delta_A)} - c_C$$

(3.5) $$v_P = \left|B_{TP}\right| + b_{AP} = \sum_{j=1}^{K} \left(\frac{b_j(1-\varepsilon_j)}{(1-\delta_j)}\right) + \frac{b_A(1-\varepsilon_A)}{(1-\delta_A)}$$

Where: $|B_{OC}|$ or $|B_{TP}|$ is the prospective contributor's valuation of the contributions of other active members, b_{AC} or b_{AP} is the prospective contributor's valuation of the author's post and c_C is the prospective contributor's cost of participation.

Equations 3.4 and 3.5 can be applied to the community as a whole. Equation 3.6 is the total value received by a community with at least one participating member (v_{TC}). Equation 3.7 is the value received by a passive community (v_{TP}).

(3.6) $$v_{TC} = |B_{TC} - C_{TC}| + b_{AT} = \sum_{j=1}^{K} \left(\frac{b_j(1-\varepsilon_j)}{(1-\delta_j)} - c_j\right) + \frac{b_A(1-\varepsilon_A)}{(1-\delta_A)}$$

(3.7) $$v_{TP} = b_{AT} = \frac{b_A(1-\varepsilon_A)}{(1-\delta_A)}$$

Where: $|B_{TC} - C_{TC}|$ is the net benefit of community contribution and b_{AT} is the community's valuation of the author's post.

As a game, this scenario can be described by the following matrix:

		Author	
		Post	**No Post**
Community	**Reply**	v_{TC}, v_A	
	No Reply	$v_{TP}, (-c_A)$	0,0

If the author posts and the community reciprocates, the community receives certain payoff v_{TC} while the author receives contingent payoff v_A. If the author posts and the community fails to respond, the community receives certain payoff v_{TP}, its valuation of

the author's post. The community always knows the current value of the discussion and will only post when the value of existing content exceeds the lowest reciprocation cost of any member under the terms described below. The author never knows his payout prior to his post; however, for the author to keep posting in repeated games, he must expect his payoff to at least exceed his cost ($v_A > c_A$).

The community maximizes its self-interest differently based on its expectations of whether the author will keep on posting content of at least minimal quality from its own perspective (repeated game) or abandon the blog if no or too little response is made in the first game (non-repeated game). Since the community can always count on the v_{TD} payout, the v_{TC} payouts are easily calculated. Equilibrium values for the community payoffs are as given in equations 3.8 and 3.9:

(3.8) $$v_{TP} \leq \begin{cases} |C_{TC}|, & repeated \\ 2\times|C_{TC}|, & non-repeated \end{cases}$$

(3.9) $$v_{TC} > \begin{cases} |C_{TC}|, & repeated \\ 2\times|C_{TC}|, & non-repeated \end{cases}$$

The return to the reciprocating community is greater than that of the passive community if the game is repeated more than once ($3\times|C_{TC}|$ versus $2\times|C_{TC}|$). So, this model supports the prediction that the community will select repeated game play from the beginning.

Since blogs can be dedicated to a variety of subjects, there can be no fixed standard by which to assess value, content is valuable if the blog community sees it as being valuable. This study proposes to objectively measure value under these terms by adopting a variation of Google's PageRank algorithm similar to that specified in Dwyer (2007). Dwyer (2007) observed that Google faced a similar problem as the one faced here when they wanted to find the most valued web content associated with any search term. Since webpages can contain content on any topic, and they needed a way of assessing value without analyzing the content, they adopted a populist criterion for assessing value whereby the most link-referenced webpages were considered most valued. According to Bianchini, Gori, and Scarselli (2005), the PageRank (x_p) of page *p* is computed by taking into account the set of pages (*pa*[*p*]) pointing to *p*:

(3.10) $$x_p = d \sum_{q \in pa[p]} \frac{x_q}{h_q} + (1-d)$$

Where $d \in (0,1)$ d is a *proportioning factor* and h_q is the *outdegree* of *q*, that is, the number of links coming out from page *q*. The proportioning factor determines the amount of importance added to *p* by the pages linking to it. Page *p* has an inherent importance of $1-d$.

In this study, a blog entry and its comments are modeled as being connected based on their resonance or similarity of theme. It is proposed that equation 3.10 be modified to assess the value of a comment as described by equation 3.11:

$$v_p = d \sum_{q \in cm[p]} R \bullet v_q + (1-d) \tag{3.11}$$

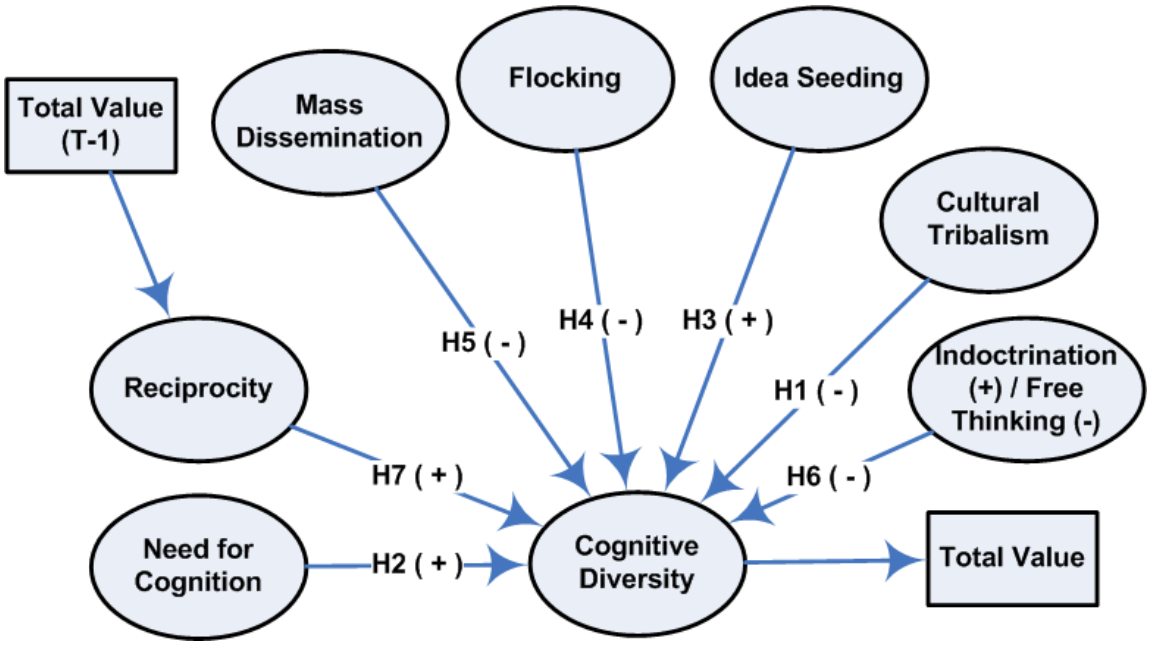
In equation 3.11, v_p is the value assigned comment p, q denotes all the other comments with a resonance (R) to p greater than 0. The total value of an author's entry can be calculated the same way, as the *recursive* sum of the values of all resonant comments.

In Chapter II, it was argued that some proportion of readers become commenters under the influence of a sense of obligation over prior value received from the blog. It is argued that the total value calculated in equation 3.11 reflects the degree of future reciprocity that will result from an author's entry. It is also argued that since some proportion of commenters will be influenced by reciprocity, the amount of reciprocity will be proportional to the total number of commenters. These two indicators are depicted in their relationship to reciprocity in Figure 3.10. Since reciprocity should bring in commenters from the full spectrum of commenter types, it is argued that:

H_7: *Reciprocity* is positively related to *cognitive diversity.*

FIGURE 3.10

Naïve Model: Hypothesized Relationships



All eight latent variables and hypotheses are integrated into a single conceptual model presented in Figure 3.10. Definitions of the latent variable indicators are aggregated in Appendix B.

Alternative Model

In this chapter, the relationships between six social forms and cognitive diversity have been combined into the naïve model of Figure 3.10. As a naïve model, the social forms are seen exclusively as acting directly on cognitive diversity. However, since this research study takes a social process theory perspective, it must be recognized that the six social forms may be interlinked in a complex manner. While it may seem that these interlinked social forms are interaction effects that should be captured by the

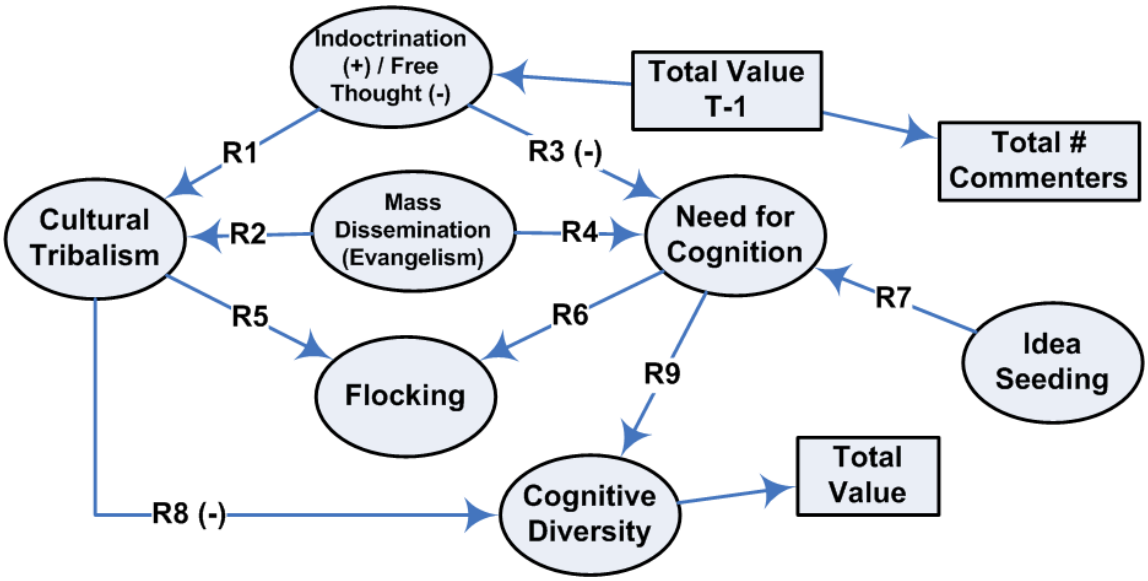
multiplicative interaction terms commonly found in regression models, it is not certain that the nature of the interlinking will be detected in this manner. The social forms in this study are often highly dependent on initial conditions, are path dependent, nonlinear and only loosely coupled to each other. They often also involve random delays and different timescales. Quantification of the relationships between such processes is seldom attempted in marketing research and is an issue this study lays a foundation for future research to address.

Nevertheless, certain interrelationships that can be captured in a conventional structural equation model have been implied in this discussion. Those interrelationships have been combined into the model shown in Figure 3.11. In addition to including hypotheses 1 and 2 from the naïve model, as R_8 and R_9 respectively, this model depicts three exogenous influences:

a) *The spectrum of blog author thought diversity*, a top-down influence acting from indoctrination to free thinking.
b) *Mass dissemination*, a grass-roots influence acting on the following individually motivated or need-based social processes that subsequently influence cognitive diversity (R_8 and R_9 in Figure 3.11):
    i. Cultural tribalism, inspired by a need for self-expression.
    ii. Flocking, motivated by a need for acceptance.
    iii. Need-for-cognition.
c) *Idea seeding*, also a grass-roots influence, acting on the same social processes as mass dissemination.

FIGURE 3.11

Alternative Model: Relationships



It was mentioned in the discussion of Figure 3.5 that the appearance of indoctrination might be created by a blog author who watches the blog to see what topics garner the most discussion and then repeatedly introduces those topics with no intent but to keep the conversation lively. This scenario is included in the model of Figure 3.11 with the inclusion of prior collaborative value ("Total Value T – 1") as a formative influence on indoctrination.

In Chapter II, indoctrination and mass dissemination were both connected to Granovetter's (1978) thresholding dynamic. It was reasoned that indoctrination seeks to overcome thresholds of resistance to meme ideas through repetition in a closed context, while mass dissemination seeks to find people with low thresholds of resistance in a variety of contexts (e.g., those characterized by both cultural tribalism and need-for-

cognition, relationships R_2 and R_4 in Figure 3.11). Just as rioting behavior can spread through a crowd as differing individual thresholds of riot resistance are overcome by observing others' rioting behavior, so meme acceptance can be expected to spread through any group as the unconverted see others in their vicinity become converted. This cascading propagation of meme acceptance is expected to be aided by flocking behavior. People with flocking propensity will sense if their local group's thematic direction has been affected by the acceptance of a meme or memeset. In order to preserve group acceptance, they will control the extent to which their expressed ideas diverge from their perception of the group norm, thus aiding in memetic propagation (R_5 and R_6 in Figure 3.11).

The logic underlying most of the relationships in this alternative model should now be clear. However, the manner in which flocking and mass dissemination are depicted may raise some questions. In this narrative, flocking commenters have been differentiated from evangelists and idea seeders by their lack of boldness. As flockers are thus characterizable as followers rather than leaders, the presence of flocking is portrayed in Figure 3.11 as inspired by the more aggressive influences of cultural tribalism (R_5) and need-for-cognition (R_6). It is argued that these more aggressive influences result in the formation of core groups of interaction around whom flockers gather depending on which group they desire to be affiliated with.

Mass dissemination, as implemented by evangelists, has been depicted in Figure 3.11 as related to both cultural tribalism (R_2) and need-for-cognition (R_4), while idea seeding is exclusively related to the latter (R_7). To the blog reader, evangelists and idea

seeders initially look the same. They are both commenters that often introduce radical ideas into the blog conversation; only with close study does it become apparent that evangelists repeatedly propound the same ideas. As a result, both will activate need-for-cognition in the people to whom the ideas are new. The people who like the evangelized ideas will naturally form a tribe around them.

Relevance of Sociological Theory

Hulberg (2006) examined the standard practices of corporate image management using Burrell and Morgan's (1979) sociological framework. Hulberg notes that Burrell and Morgan's framework describes four paradigms, or perspectives, that purports to classify all sociological theories: (1) Functionalist: Consumers' image perceptions can be influenced by corporate messages and behavior; (2) Interpretive: Consumers' image perceptions are entirely formed through social interactions among them and cannot be influenced by corporate messages or behavior; (3) Radical humanism: Consumers view corporate attempts to convey image as manipulative; and (4) Radical structuralism: Consumers view all corporate action as motivated toward gaining power and profit that will be used to suppress others.

Clearly, the image management process of Figure 2.1 is predominantly based on the functionalist paradigm: companies are seen as able to influence image perceptions. However, the inclusion of the social forms (socio-psychological processes) hypothesized to be the primary mediators in the expression of consumers' thoughts to blogs, allows the use of not only functionalist, but interpretive and radical humanist perspectives:

a) *Functionalist.* It is readily apparent that the memetic processes of indoctrination and mass dissemination explicitly assume that one party can influence the image perceptions of another. Even though idea seeding and reciprocity do not result in perceptions of image being influenced, they do initiate cascades of behavior where individuals influence one another by causing boldness and sense of obligation thresholds to be exceeded. Idea seeding and reciprocity can thus be viewed from a functionalist perspective.

b) *Interpretive*. Flocking and cultural tribalism are both processes where outcomes depend on group norms established without the intervention of the blog author or company hosting the blog. The degree of cognitive diversity expressed by flockers will be dependent on individual perceptions of how much divergent expression the group will tolerate. Since cultural tribalists are motivated by a need for self expression, they will be attracted to blog communities already focusing on their preferred themes. The tribe can be expected to migrate to any blog whose theme more closely matches their preference, abandoning any blog whose theme changes sufficiently. Therefore, even though cultural tribalism is supported by indoctrination that matches tribal preference, the phenomenon is caused by like individuals coming together and thereafter acting in concert.

c) *Radical humanist*. As discussed in the section on memetic processes, indoctrination and mass dissemination are specifically intended to be tools of persuasion. Thus, any blog readers who are motivated primarily by need-for-cognition are expected to recognize these influences for what they are and will,

depending on the intensity of their use, either disregard them or abandon the blog for being too narrow in subject matter. Blog readers motivated by the need for self expression underlying cultural tribalism are also expected to have a similar response. Note that blog readers whose preferences match ideas spread through these memetic influences will not see them as manipulative.

While the association of sociological processes with image management is not new, this study increases awareness of the variety of ways that sociological processes influence image diffusion and the expression of image.

Summary

Earlier in the chapter, the focus of this study was stated to be investigating the effects of six social forms, identified by prior research to strongly affect cognitive diversity, the diversity of thought expressed in a group forum. If the operation of these social forms cause the expression of consumer brand image to be distorted or incomplete, then companies that mine blogs for brand image insights will not gain an accurate picture of brand image in the market. To that end, measurable constructs for the six social forms have been proposed as eight latent variables with suitable indicators. These latent variable constructs and their indicators are summarized in Table 3.1. These latent variables have been configured into two models, a naïve model proposing direct effects on cognitive diversity and an interaction model proposing interrelationships between seven of the latent variables. Full depictions of these models, with all reflective and formative relationships, are included in Appendix C as Figures C-2 (Naïve) and C-3

(Alternative). The next chapter details the methodology used to test the hypothesized relationships.

TABLE 3.1

Hypotheses and Construct Definition Summary

	Constructs	Indicators
	Cognitive Diversity. The many cognitive styles or ways individuals think, perceive and remember information and then use that information to solve problems or, in general, enact their behavior.	*Mean cluster density.* After the comment clusters for each blog entry are found, the distance from each comment's location in cognitive space to the centroid of its cluster can be calculated. *Mean cluster radius.* After the comment clusters for each blog entry are found, the distance from each comment's location in cognitive space to the centroid of its cluster can be calculated. *Number of clusters.* After the comment clusters for each blog entry are found, the number of clusters is proposed to be indicative of the number of conversational themes. *Percentage of outliers.* After the comment clusters for each blog entry are found, the distance from each comment's location in cognitive space to the centroid of its cluster can be calculated. The mean and standard deviation of each cluster's comment-to-centroid distances is calculated and a z-score calculated for each comment. Any comment with a z-score greater than 2.0 is considered an outlier.
H_1	**Cultural Tribalism.** A state that a virtual community can be in where the members share a strong ideological identity, creating a kind of intellectual segregation where like only talks to like.	*Mean collective thought separation.* The mean of entry-to-comment and comment-to-comment cognitive distances for the current blog entry. *Stdev collective thought separation.* The standard deviation of entry-to-comment and comment-to-comment cognitive distances for the current blog entry. *Mean indiv. thought separation.* The mean comment-to-comment cognitive distance, across blog entries, between comments made by the same individual, a commenter on the current blog entry.

TABLE 3.1 (CONT)

	Constructs	Indicators
H_1 (cont)		*Stdev indiv. thought separation.* The standard deviation of entry-to-comment and comment-to-comment cognitive distances for the current blog entry. *Mean commenter longevity.* The mean time between the first comment and the current comment of all commenters to the current blog entry. *% Repeat commenters (T - 1).* The number of commenters contributing to the current blog entry who contributed to the previous blog entry.
H_2	**Need-for-Cognition.** Part of what determines the level of thought exercised is innate *need-for-cognition*, an enjoyment of cognitively demanding tasks (From Petty and Cacioppo's (1986) Elaboration Likelihood Model).	*Mean collective thought separation.* The mean entry-to-comment and comment-to-comment cognitive distances for the current blog entry. *Stdev collective thought separation.* See above. *Mean indiv. thought separation.* The mean entry-to-comment and comment-to-comment cognitive distances for the current blog entry. *Stdev indiv. thought separation.* See above. *% First time commenters.* The number of commenters contributing to the current blog entry who have never contributed to previous entries.
H_3	**Idea Seeding.** The most uninhibited commenters expressing ideas on the most diverse array of themes.	*Mean clusters started.* The mean number of thematic clusters started over the participation lifetime of a blog commenter. *Stdev indiv. thought separation from cluster centroid.* The standard deviation of the cognitive distance between a commenter's comments and the centroids of the closest thematic cluster across blog entries over the participation lifetime of a blog commenter. *Mean number of comments.* The mean number of comments (across blog entries) created by individuals commenting on the current blog entry. *Mean indiv. thought separation.* See above. *Stdev indiv. thought separation.* See above.

TABLE 3.1 (CONT)

	Constructs	Indicators
H_4	**Flocking.** People control the expression of their thoughts to balance individuality with group cohesion.	*Mean clusters started.* See above. *Stdev indiv. thought separation from cluster centroid.* The mean cognitive distance between a commenter's comments and the centroids of the closest thematic cluster across blog entries over the participation lifetime of a blog commenter. *Mean number of comments.* See above.
H_5	**Mass Dissemination.** One of the two primary memetic propagation mechanisms where the targets are the people with the lowest barriers to adoption, so the idea is spread as widely as possible to increase the probability of reaching people with a low threshold of resistance.	*Mean clusters started.* See above. *Stdev indiv. thought separation from cluster centroid.* See above. *Mean number of comments.* See above. *Mean indiv. thought separation.* See above. *Stdev indiv. thought separation.* See above.
H_6	**Indoctrination.** One of the two primary memetic propagation mechanisms where the target person or population is subjected to a high repetition of the meme idea in an attempt to overwhelm thresholds of resistance.	*Mean time between entries.* The mean interval between successive entries posted to a blog. *Mean entry-to-entry separation.* The mean pair-wise cognitive distances between entries to the same blog. *Stdev entry-to-entry separation.* The standard deviation in the pair-wise cognitive distances between entries to the same blog.
H_7	**Reciprocity.** As people read the blog and get value from it, a sense of obligation to contribute builds up.	*Total number of commenters.* The total number of unique commenters to a blog entry.

CHAPTER IV

METHODOLOGY

Chapter III introduced an integrated naïve model of how six social processes interact to affect cognitive diversity and an alternative model portraying certain interrelationships between the six social processes. In this chapter, a means of testing the explanatory sufficiency of these models is described. It will be described how the model in Figure 3.10 was instantiated as a multi-group model. It will be shown how blog entry comments were separated into those influenced primarily by cultural tribalism and those characterized by need-for-cognition in so much as they embody the attributes of the presence of those grassroots influences. Within each blog, commenters were segmented into idea seeders, evangelists and flockers as they embodied the influences of threshold behavior, mass dissemination and flocking, respectively. All commenters, regardless of type were predicted to be capable of engaging in behavior motivated by reciprocation, need-for-cognition and cultural tribalism. The aggregate model simultaneously captured the influence of the full spectrum of blog author thought diversity levels and the three groups of commenters on overall cognitive diversity.

This chapter begins by describing the data set used in this study and the considerations made in its selection. Then, a step-by-step description is given of how the raw data was processed to make it suitable for testing the hypotheses presented in Chapter III. Next, the criterion used in determining whether or not hypotheses are

supported is described. Finally, the limitations associated with the methodology are described.

Data

Kozinets (2002) introduced *netnography* as a methodology where the principles of ethnography, or unobtrusive observation, were applied to the study of a virtual community of coffee aficionados. He specified a selection criterion for subject communities that differed from the practice of standard ethnography: select communities that are focused on a research question-relevant topic, receive above average posting traffic, have a large number of contributing members, contain descriptively rich content and thus, in general, enjoy a high level of member-to-member interaction. Having a high level of member-to-member interactions is important to this study because it increases the number of cognitive objects (i.e., bodies of text) and, as a result, the probability of detecting the effects of the social processes described in Chapter III.

The study used author posts and community response data from the 15 diverse blogs listed in Table 4.1. Consistent with the netnography methodology proposed by Kozinets (2002), these blogs were purposively selected because they are:

a) *Highly active*. Blogs whose authors post frequently and each post tends to attract more than 20 comments from readers (Table 4.2).

b) *Representative of the wide diversity in blog topics relevant to marketing*. In light of the broad scope of marketing as a field of study, as summarized in Table 4.1, in addition to blogs related to companies and products, blogs related to political

ideology, popular culture and people who are business thought-leaders are included in the study.

c) *Relevant to the research questions*. Indoctrination is a social process whose effects are not likely to be strong in all blogs. Therefore, three blogs (Blog for America, The Evangelical Outpost and Townhall) were purposely selected as indoctrination blogs because an informal examination of their posts and comments indicated a high level of narrowly-scoped ideological content. Three additional blogs (EnGadget, Gizmodo and Joystiq) were selected as need-for-cognition blogs because they frequently introduce new products to consumers. The overview of need-for-cognition and sensemaking theory presented in Chapters II and III suggests that such blogs might prompt a higher level of cognitive activity motivated by need-for-cognition as consumers collaboratively converge on an understanding of the benefits that these new products might offer. The remaining blogs were selected because their high level of activity should allow most of the processes studied to be manifested.

d) *Allow comments from readers*. In Chapters I and II, it was emphasized that the purpose of blogging was for consumers and producers to converse. Unfortunately, many corporate blogs do not allow for reader comments. Since this study focuses on the expression of diverse thought, only blogs that allow comments are suitable.

e) *Retain an archive of posts and comments*. Some blogs only show the most recent activity. Hypotheses 1, 2 and 4 require that the long term activity of commenters be studied to detect regularities in the commenting activity of individuals.

TABLE 4.1

Data Sources

ID	Blog Name	Description
1	AutoBlog (71)[3]	A blog covering the auto industry with test drives and commentary on articles from other sites.
2	Blog for America (21,406)	The Democratic Party's blog.
3	Blog Maverick (269)*	Entrepreneur Mark Cuban's blog.
4	The Consumerist (24)	A consumer retaliation blog.
5	EnGadget (1)	A product review blog for high-tech consumer goods.
6	The Evangelical Outpost (1668)	Reflections on culture, politics and religion from an evangelical worldview.
7	Fastlane (8890)*	GM Vice Chairman Bob Lutz's blog.
8	Freakonomics (137)*	The blog for the book *Freakonomics*
9	Gizmodo (3)	A product review and news blog for high-tech consumer goods.
10	Google Blogoscoped (100)	A consumer's blog covering Google and its services.
11	Joystiq (39)	A product review blog for the gaming industry.
12	PaulStamatiou (99)*	A college student reviewing technology products and offering technical support
13	Townhall (73)*	Hugh Hewitt's, a radio talk show host, blog on politics, religion and conservative social commentary.
14	TV Squad (104)	Commentary and review of TV shows.
15	The Unofficial Apple Weblog (52)	Since Apple has no corporate blog, their fans have started blogs about them.

[3] The number within parentheses is the weblog's position on the Technorati, a weblog search engine, ranking of the most authoritative (most widely cited by other websites) of the almost 100 million weblogs tracked (as of Aug 07). An asterisk denotes situations where the whole blog archive was used because there were fewer than 2000 blog entries.

TABLE 4.2

Data Source Quantities

Blog	Dates	Posts	Comments	Commenters
AutoBlog	Mar 07 – Jul 07	2,036	27,826	14,657
Blog for America	Mar 03 – Dec 04	2,232	252,395	15,388
Blog Maverick	Nov 99 – Sep 07	435	16,681	7,546
The Consumerist	Mar 07 – Aug 07	2,599	73,052	9,141
EnGadget	May 07 – Jul 07	2,030	51,738	14,203
The Evangelical Outpost	Oct 03 – Sep 07	1,731	52,027	4,813
Fastlane	Jan 05- Aug 07	266	15,914	5,572
Freakonomics	Mar 05 – Aug 07	1,142	23,445	8,375
Gizmodo	Jul 07 – Aug 07	2,243	39,091	7,078
Google Blogoscoped	Jan 06 – Aug 07	2,460	20,227	4,015
Joystiq	Jun 07 – Sep 07	2,058	52,130	7,218
Paul Stamatiou	Sep 05 – Sep 07	757	10,195	3,102
Townhall	Apr 06 – Sep 07	985	15,455	2,689
TV Squad	Jan 07 – Jul 07	2,034	24,994	6,605
The Unofficial Apple Weblog	Apr 07 – Sep 07	2,165	23,414	7,465
Total		25,173	698,584	117,867

Two of the blogs (Blog for America and GM Fastlane) were selected as pretest blogs because they are expected to demonstrate activity at the extremities of the various conceptual maps (i.e., Figures 3.2 and 3.6) used to demarcate the diverse influences affecting blogs.

Method

Data Collection and Preprocessing

The first phase of this study was to collect a sample of approximately 2000 authors' posts and associated comments from each blog using a custom software program that automated the collection. Automated processes, used throughout this study and noted where appropriate, are particularly relevant to this study because of the large volume of data. In cases where the blog had less than 2000 author posts, the entire blog archive was collected. Then, each entry and comment was decomposed into centering resonance analysis (CRA) word networks, the influence of each node of the word network was calculated, and then the resonance between posts and comments was calculated on the following dimensions:

a) *Post-to-post.* The resonance of a post with all posts that follow it. This was designed to show the extent to which all posts follow the same theme.

b) *Post-to-comment.* The resonance of a post with the comments that respond to it and the resonance between the responding comments. This was designed to indicate the extent to which the comments thematically diverge from that of the post.

c) *Prior post and its comments to subsequent posts.* The resonance between past posts and their comments with future posts. This was designed to reveal the extent to which authors are responsive to issues important to the commenters.

d) *Posts by the same author.* Since some blogs have more than one author this dimension was designed to indicate whether individual authors keep posting on the same theme.

e) *Comments by the same commenter.* Capturing the diversity of thought expressed by individual commenters was an important part of assessing the degree of autonomy tolerated by the group.

f) *Comment-to-comment.* The resonance between all comments, regardless of associated post, was designed to allow clusters to be detected that indicate major themes or the content of the community's common knowledge structure.

Corman and Dooley (2008), two of the authors of Corman et al. (2002), and inventors of Centering Resonance Analysis (CRA), founded Crawdad Technologies, LLC to market *Crawdad*, a software program that automates the process of turning text into CRA word networks, and calculating influence and resonance. They permit the download of an evaluation copy of their software that processes a body of example text into a CRA word network and performs the relevant calculations. After reading Corman et al. (2002) and Wasserman and Faust (1994) it was possible to write a program in Microsoft's C-Sharp (C#) programming language that duplicated the output of the Crawdad software using its example text. The C# program was then used to process the data for this study.

MDS and Cluster Detection

As described in Chapter II, since CRA's resonance metric allows a pair-wise measure of similarity to be calculated for every document (i.e., post and comment);

Kamada and Kawai's (1989) spring-like methodology was adapted to self-assemble clusters of cognitive objects in a multidimensional space. These clusters were then demarcated using Chiu et al's (2001) two-step methodology and cluster membership confirmed with McQueen's (1967) k-means.

Spring model multidimensional scaling was implemented with a custom program written in C# that utilized a modified version of Dwyer's (2004) *WilmaScope* 3D graph visualization library. WilmaScope was modified to use four dimensions instead of three so as to allow the spring models sufficient freedom to assume their desired unstressed lengths as explained in Chapter II in the sections entitled *Multidimensional Scaling* and *Spring Models in Multidimensional Scaling*.

Negative and Positive Valance

CRA focuses on the detection of thematic consistency between documents; however, it does not capture differences in positive and negative valence in its determinacy of theme consistency. To address this shortcoming, a separate accounting of relative frequency in the use of negative words between two compared documents was maintained. In the previous section, a means of locating a document in a multidimensional thematic space is described where relative differences are turned into relative coordinates that can then be used in the detection of thematic clusters. The accounting of negative valance is used as an additional dimension in the detection of thematic clusters.

Multigroup Analysis

After the comment text was transformed into cognitive objects located in clusters in multidimensional space, the task of identifying blog entry comment (cultural tribalism and need-for-cognition) and commenter (evangelist, idea seeder or flocker) types began. This study regards the commenters in the blogs listed in Table 4.1 as *cluster samples* from the same larger population, people who comment in blogs. Cluster samples differ from random samples in that subjects in the same cluster are more similar to each other than those in a truly random sample. Recognizing a level of similarity seems appropriate because blogs tend to be topical and thus attract people who, at the least, have a common interest in the same topic. As a result, the parameters for equations 3.1 and 3.2, the discriminant models identifying primary influences and commenter types, were estimated separately using commenter data from each of the 15 blogs. Then, the parameter estimations were compared to check for consistency. Additionally, it was necessary to demonstrate both the factorial invariance of the measuring instrument (i.e., the 8 factors proposed in Chapter III) and the invariance of the causal structure (the models in Figures 3.10 and 3.11) across the individual blogs. EQS 6.1 was used in conjunction with the procedures prescribed by Byrne (2006), and summarized below, to demonstrate both types of invariance.

In order to use equations 3.1 and 3.2 for segmentation, the parameter coefficients must have been estimated using data already classified. Discriminant models are then estimated using a portion (typically 70%) of the pre-classified dataset and the accuracy of the determinant model's ability to classify data items is then tested using the

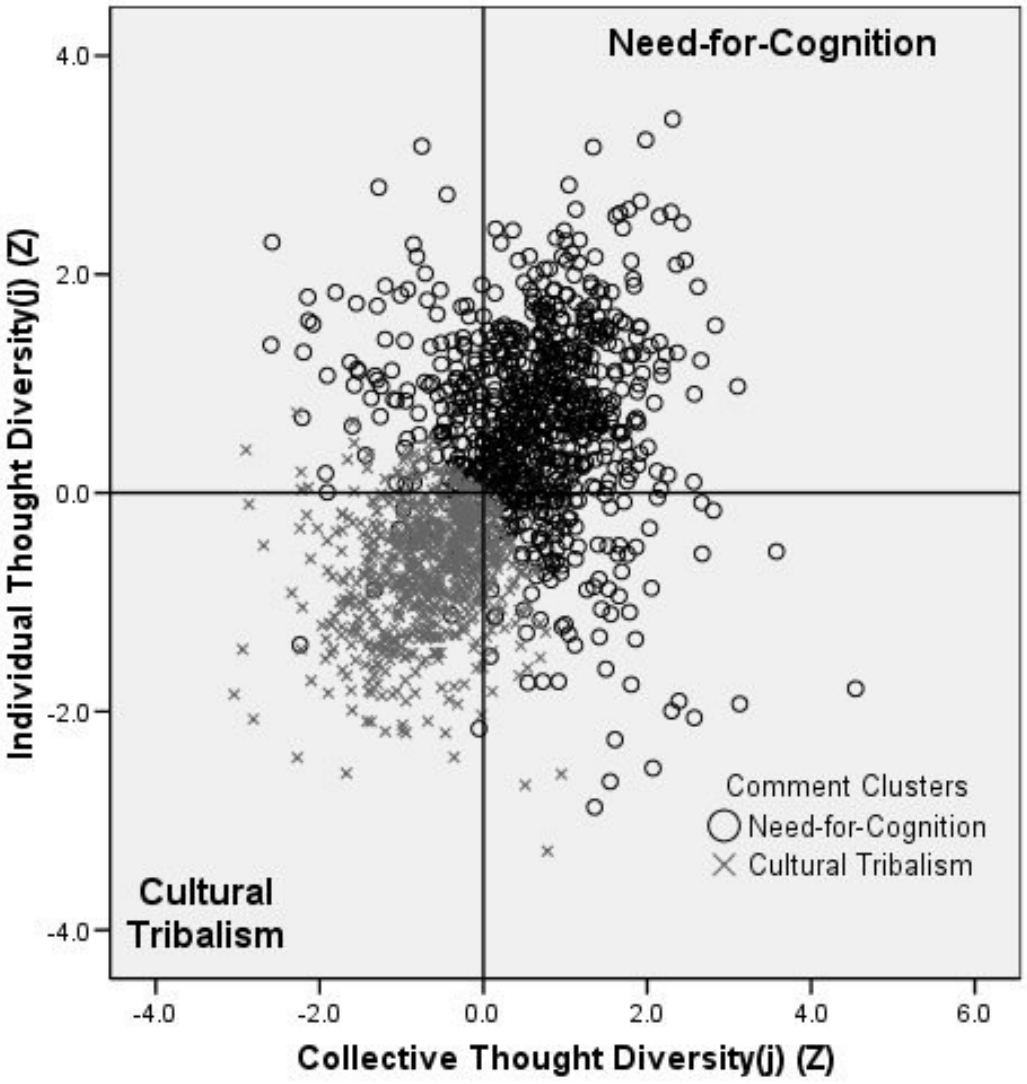
remaining or *holdout* sample. Since it is desired to determine whether the number of clusters naturally present in the data matches the conceptual map topologies given in Figures 3.2 (cultural tribalism and need-for-cognition) and 3.8 (flockers, idea seeders and evangelists), Chui et al's (2001) methodology was used to get an initial cluster demarcation. Then, the assignment of objects to clusters was confirmed with McQueen's (1967) widely used k-means clustering algorithm. When the data was thus pre-classified, the parameters of the discriminant models (equations 3.1 and 3.2) described in this study were estimated.

Classification of Blog Entry Comments

Chapter III described how blog entry comments can be classified as either influenced primarily by cultural tribalism or need-for-cognition using the discriminant model equation 3.1 or by plotting the positions of comments in the conceptual map shown in Figure 3.2, derived from equation 3.1. The two comment types were distinguished, one from the other, using Chui et al's (2001) methodology using all the comments from a random sample of 25 blog entries selected from each of the two pretest blogs. Two clusters emerged and cluster membership was confirmed using McQueen's (1967) k-means. When this pretest sample was plotted on the conceptual map of Figure 3.2, the result is as displayed in Figure 4.1.

FIGURE 4.1

Pretest Blog Entry Comment Mapping



While the results are not perfectly conforming to the proposed perceptual map regions, the higher density of data points within the so-called cultural tribalism and need-for-cognition quadrants suggests support for the proposed topology for blog entry

comment segmentation. A notable aspect of these results is that a near equal number of the comments within the cultural tribalism quadrant belong to the GM Fastlane blog, not the blog predicted to be an indoctrination blog. A positive aspect of this observation is that the topology-based segmentation seems to distinguish one comment from another; however, the unexpected inability to distinguish between blogs suggests that the logic underlying the initial prediction of blog types may have been too superficial.

The parameters of equation 3.1 were estimated from the pretest dataset and the accuracy of the resulting model tested with a holdout sample. One discriminant function emerged from the estimation with a characteristic root eigenvalue greater than 1.0. This function correctly classified 88.3% of the holdout sample.

Classification of Commenters

Chapter III described how commenters can be characterized as evangelists, idea seeders and flockers by plotting their positions on a conceptual map (Figure 3.8, derived from equation 3.2). Idea seeders can be separated from evangelists by high diversity in thought expression: idea seeders express thoughts about a wide variety of things while evangelists always express similar thoughts. Idea seeders and evangelists can be separated from flockers on the basis of differences in boldness: idea seeders and evangelists are bolder when expressing their thoughts.

When the commenting activity of 1,000 commenters, 500 randomly selected from the Blog for America (BFA) and the GM Fastlane blog (GM) as a pretest sample, was demarcated by Chiu et al's (2001) methodology, three clusters emerged (Figures 4.2

and 4.3). Cluster assignments were confirmed using McQueen's (1967) k-means methodology.

FIGURE 4.2

Pretest BFA Commenter Clusters

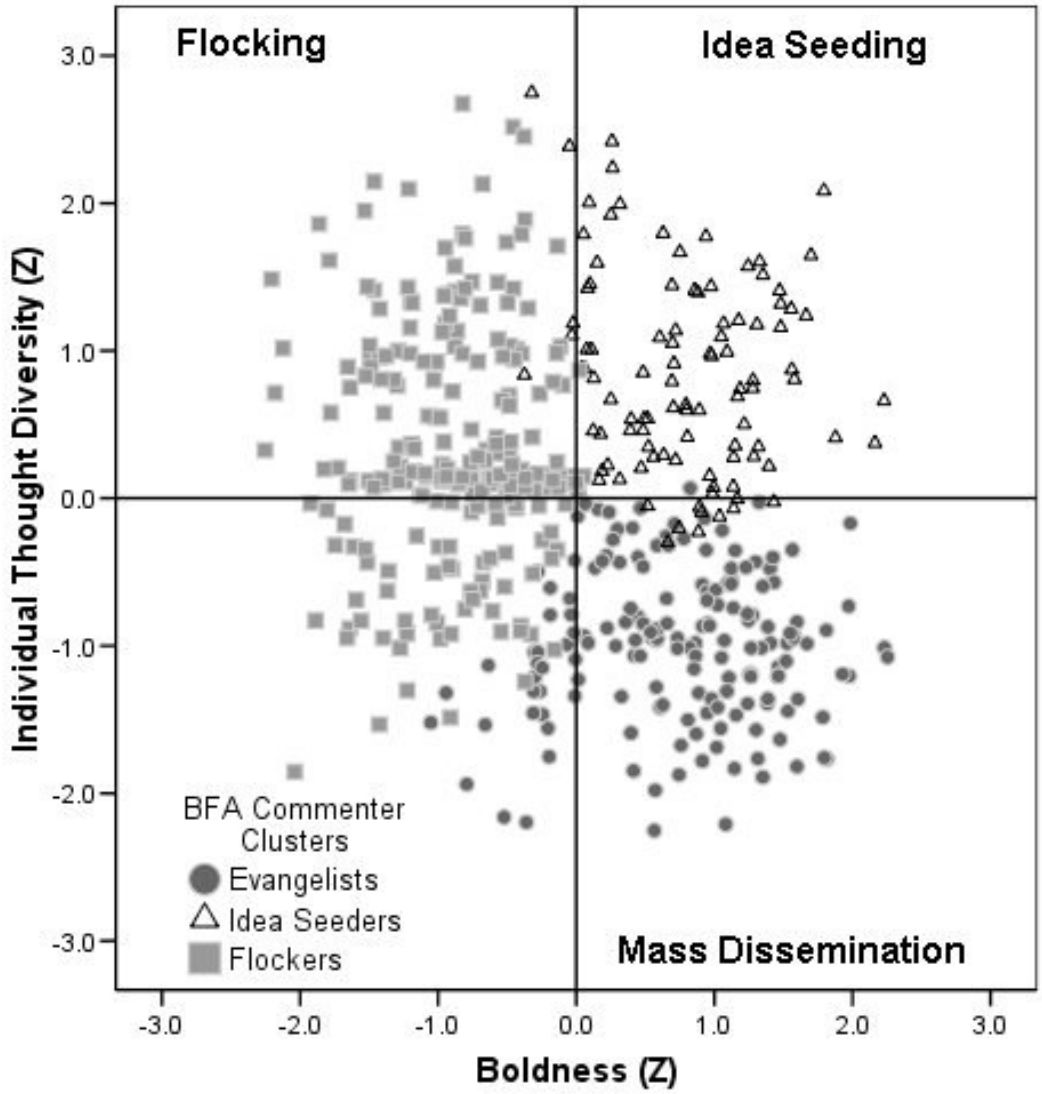
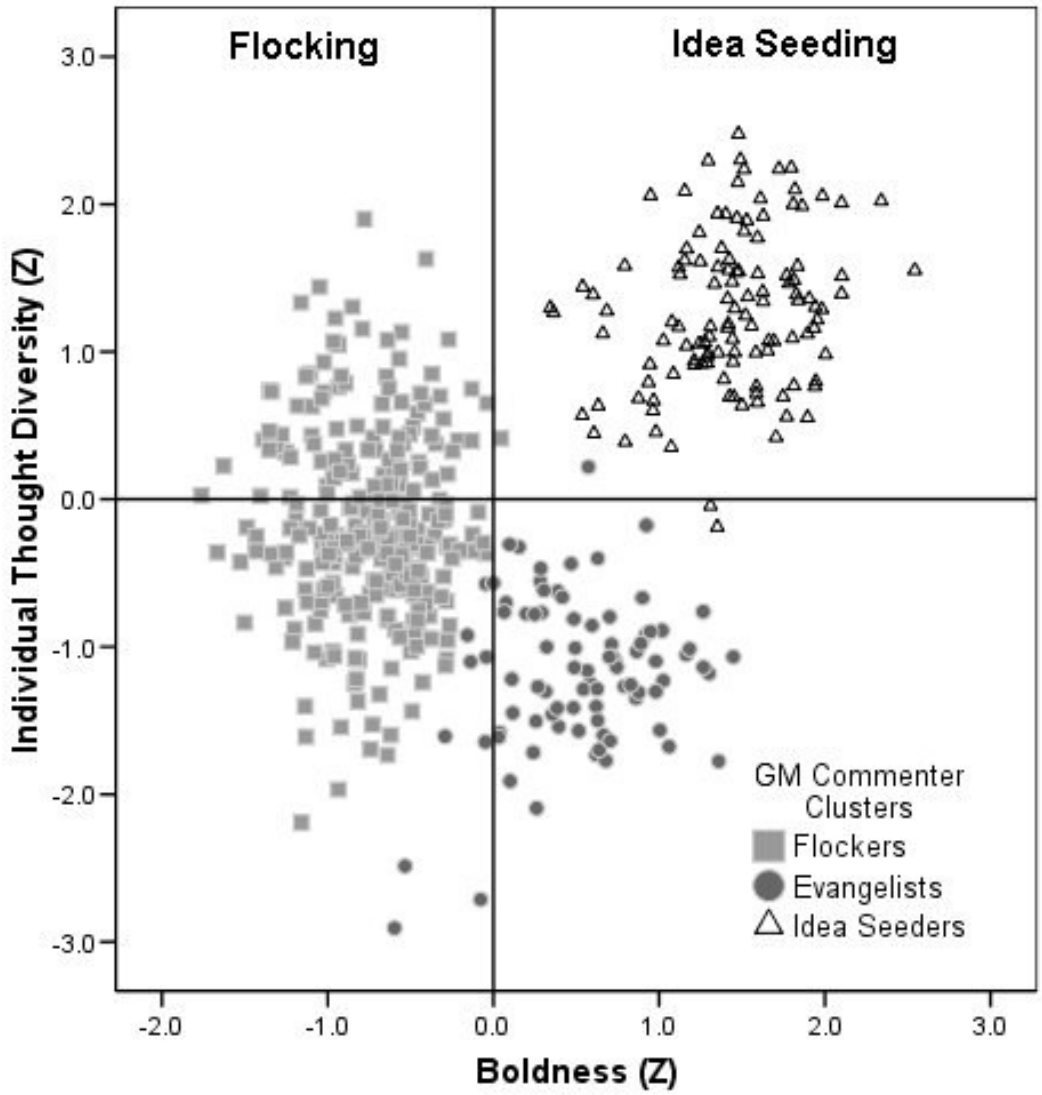


FIGURE 4.3

Pretest GM Fastlane Commenter Clusters



The placement of these clusters was considered conclusive and conforming to expected patterns. It is interesting to note that idea seeding seems to be more pronounced in the GM Fastlane blog.

In preparation for estimating the parameters of equation 3.2 the commenters were forcibly classified into the segments of Figure 3.8. Using these topology-based segment membership assignments, the parameters of equation 3.2 were estimated and the accuracy of the resulting model tested with a holdout sample. Two discriminant functions emerged from estimation with each of the two pretest datasets, correctly classifying 86.5% (GM) and 79.3% (BFA) of the holdout sample.

Construct Validity

Bagozzi, Yi and Phillips (1991) broadly define construct validity "as the extent to which an operationalization [(i.e., a measuring device)] measures the concept it is supposed to measure." (p. 421) Construct validity focuses on the extent to which data exhibits (1) convergent validity, the extent to which different assessment methods concur in their measurement of the same construct (also called factorial or measurement invariance), and (2) discriminant validity, the extent to which the measures of different constructs are distinct (Campbell and Fiske 1959). The demonstration of factorial invariance began by establishing baseline confirmatory factor analysis (CFA) results for each blog dataset. The results were then compared to see if the same indicators loaded on the same factors. After the factorial structure was found to be the same across blogs, hypotheses that each *free* factor loading (each factor's "first" indicator is constrained to unity and is thus not free) was equal across blogs was tested by forcing the factor loadings to be equal and then observing overall fit statistics as well as the LaGrange Multiplier (LM) Test univariate probability for each factor loading. This probability is a measure of significance for the LM Test, a test of the hypothesis that a factor loading

varies across groups (in this case, blogs). LM Test probabilities less than 0.05 are considered statistically insignificant. That is, when the LM Test probability is less than 0.05, the parameter does not have the same value across the blogs. This procedure does not require equal sample sizes, nor does the test of causal structure invariance described below.

Invariance of causal structure was tested in a manner similar to the above. The same structural model was simultaneously fitted to the 15 blog datasets while constraining all factor loadings, structural path loadings and error covariances equal across the blogs. Then multigroup model fit statistics were observed as well as the LM Test univariate probability for each constrained parameter.

Underlying the use of structural equation modeling to perform CFA is a strong assumption of multivariate normality. If evidence shows that this assumption is violated, it has important implications for the interpretability of the findings. None of the indicators in the pretest data set demonstrated significant nonzero univariate skewness (one tail of the distribution is longer than the other) or kurtosis (variance is due to infrequent extreme deviations from the mean). Normalized estimates of Mardia's (1970) multivariate coefficient, an aggregate indicator of kurtosis, were found to be 2.15 (BFA) and 3.98 (GM). Bentler (2005) suggests that only values greater than 5.0 indicate non-normally distributed data. Another test of multivariate normality is discussed later.

The individual pretest blog baseline CFA results showed the same indicators loading on the same factors as given in Table 4.3. Additionally, as Table 4.3 shows, all but eight of the 32 core factor construct indicators met Hair et al's (2005) criterion for

statistical significance (factor loading > 0.4 with a sample size > 200) and 14 met the criterion for practical significance (loading > 0.5). It must be noted that Hair et al.'s criteria, as well as most of the reference values quoted in this study, applies to evaluating instruments used to gather experimental data. The circumstances of this study are much different as Hair et al.'s criteria are being used to evaluate a means of gaining insights from non-experimental, albeit primary, empirical data. Therefore, it is argued that Hair et al's criteria must be referenced with caution as the level of control over extraneous variation that characterizes experiments is absent here. That being understood, the general conformance of this study's factor loadings and other metrics to those recommended by Hair et al. (2005) and Byrne (2006) is worthy of note and indicative of a reasonable level of performance. The multigroup CFA with factor loadings constrained equal across blogs had good fit (CFI = 0.954, RMSEA = 0.019, χ^2 (896) = 1,225) indicating that both blogs have similar, if not the same, factor structures.

TABLE 4.3

Pretest Confirmatory Factor Analysis, Reliability, Variance Explained and Factorial Invariance

Main Factor and Reliability	Indicator	Loading	R^2	LM Test Probability
Cognitive Diversity	Mean Cluster Density	-0.562	0.317	0.000
r = 0.765	Mean Cluster Radius	0.668	0.448	0.000
	Number of Clusters	0.612	0.377	0.000
	Percentage of Outliers	0.537	0.290	N/A
Cultural Tribalism	Mean Collective Thought Separation	-0.496	0.246	0.000
r = 0.683	Stdev Collective Thought Separation	-0.571	0.326	0.000
	Mean Indiv. Thought Separation	-0.493	0.243	0.000
	Stdev Indiv. Thought Separation	-0.450	0.211	N/A
	Mean Commenter Longevity	0.349	0.122	0.072
	% Repeat Commenters (T - 1)	0.493	0.244	0.895
Flocking	Mean Clusters Started	-0.562	0.316	0.040
r = 0.625	Stdev Indiv. Thought Separation From Cluster Centroid	-0.639	0.408	N/A
	Mean Number of Comments	-0.618	0.381	0.000
Idea Seeding	Mean Clusters Started	0.411	0.169	0.092
r = 0.519	Stdev Indiv. Thought Separation From Cluster Centroid	0.419	0.177	0.137
	Mean Number of Comments	0.470	0.221	0.039
	Mean Indiv. Thought Separation	0.365	0.134	0.746
	Stdev Indiv. Thought Separation	0.413	0.172	N/A

TABLE 4.3 (CONT)

Main Factor and Reliability	Indicator	Loading	R^2	LM Test Probability
Indoctrination / Free Thought	Mean Time Between Entries	-0.623	0.406	N/A
r = 0.485	Mean Entry-to-Entry Separation	-0.580	0.337	0.008
	Stdev Entry-to-Entry Separation	-0.474	0.225	0.002
Mass Dissemination	Mean Clusters Started	0.435	0.216	0.000
r = 0.663	Stdev Indiv. Thought Separation From Cluster Centroid	0.501	0.256	0.006
	Mean Number of Comments	0.506	0.259	0.023
	Mean Indiv. Thought Separation	-0.515	0.278	0.000
	Stdev Indiv. Thought Separation	-0.562	0.316	N/A
Need for Cognition	Mean Collective Thought Separation	0.478	0.230	0.001
r = 0.640	Stdev Collective Thought Separation	0.601	0.361	0.195
	Mean Indiv. Thought Separation	0.551	0.304	0.001
	Stdev Indiv. Thought Separation	0.490	0.251	N/A
	% First Time Commenters	0.391	0.153	0.056
Reciprocity				
r = 1.0	Total Number of Commenters	1.0	1.0	N/A

Note: $\rho < .05$. Reliabilities are calculated from the indicators to each individual factor in Table 4.3. Pearson "r" is reported instead of Cronbach's α. N/A in the right hand column denotes the arbitrary first factor.

Table 4.3 also shows that 13 of the 32 factor loadings exhibited a less than 5% LM Test univariate probability, indicating that their factor loadings are not the same across the two blogs even though the structure appears to be the same.

TABLE 4.4

Pretest Correlation and Φ-matrix of Primary Constructs

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.033	-					
3	Flocking	0.004	0.001	-				
4	IdeaSeed	0.265	0.084	-0.014	-			
5	Indoc	0.079	0.475	0.078	0.166	-		
6	MassDiss	0.120	0.020	0.206	0.528	0.574	-	
7	NFC	0.117	0.183	0.109	0.162	0.942	0.586	-
8	Recip	0.205	0.108	-0.062	0.117	0.981	0.209	0.439

*Note: Correlations are corrected for attenuation. All values are significant to $\rho < .05$. CogDiv = Cognitive Diversity, CultTrib = Cultural Tribalism, IdeaSeed = Idea Seeding, Indoc = Indoctrination, MassDiss = Mass Dissemination, NFC = Need-for-Cognition, Recip = Reciprocity.

Discriminant validity was assessed in two stages: (1) correlations between constructs, corrected for attenuation, were confirmed significantly less than unity (Table 4.4 shows the correlations for the GM Fastlane blog), and (2) a χ^2 difference test was performed between the baseline multigroup CFA model (CFI = 0.998, χ^2 (872) = 888) and an alternative multigroup model (CFI = 0.106, χ^2 (928) = 7,338) where the correlation between each latent construct was constrained at unity. According to Cheung and Rensvold (2002), a large CFI or χ^2 difference demonstrates discriminant validity. While there is a widely accepted understanding of what constitutes a significant χ^2

difference for a given degrees of freedom, Cheung and Rensvold's recommendation that a $\Delta CFI > 0.01$ be considered significant is only beginning to gain acceptance (Byrne 2006, p. 341). In this study, an attempt is made to disclose both metrics where possible. It is readily apparent that a ΔCFI of 0.892 and a $\Delta\chi^2$ of 6,450 with degrees of freedom equal to 56 argue strongly in favor of discriminant validity.

As has already been intimated, prior to conducting a confirmatory factor analysis, and the model estimation, it is important to test for multivariate normality in the factor indicators. This is particularly important in this study as many measures of collective phenomena have an underlying power-law distribution, as discussed in Chapter II, rather than a normal distribution. Hair et al. (2005) point out the difficulty in testing multivariate normality (the joint normality of more than one variable) and suggest that testing univariate normality is an acceptable substitute, especially if sample sizes are large. Hair et al. recommend graphical analysis using the normal probability plot as well as statistical techniques such as skewness and kurtosis. Figures 4.4 and 4.5 show normal probability plots for estimated values of the BFA and GM Fastlane blog's cognitive diversity latent variable. These plots show no normality problems.

FIGURE 4.4

Pretest Normal Probability Plot of BFA Cognitive Diversity

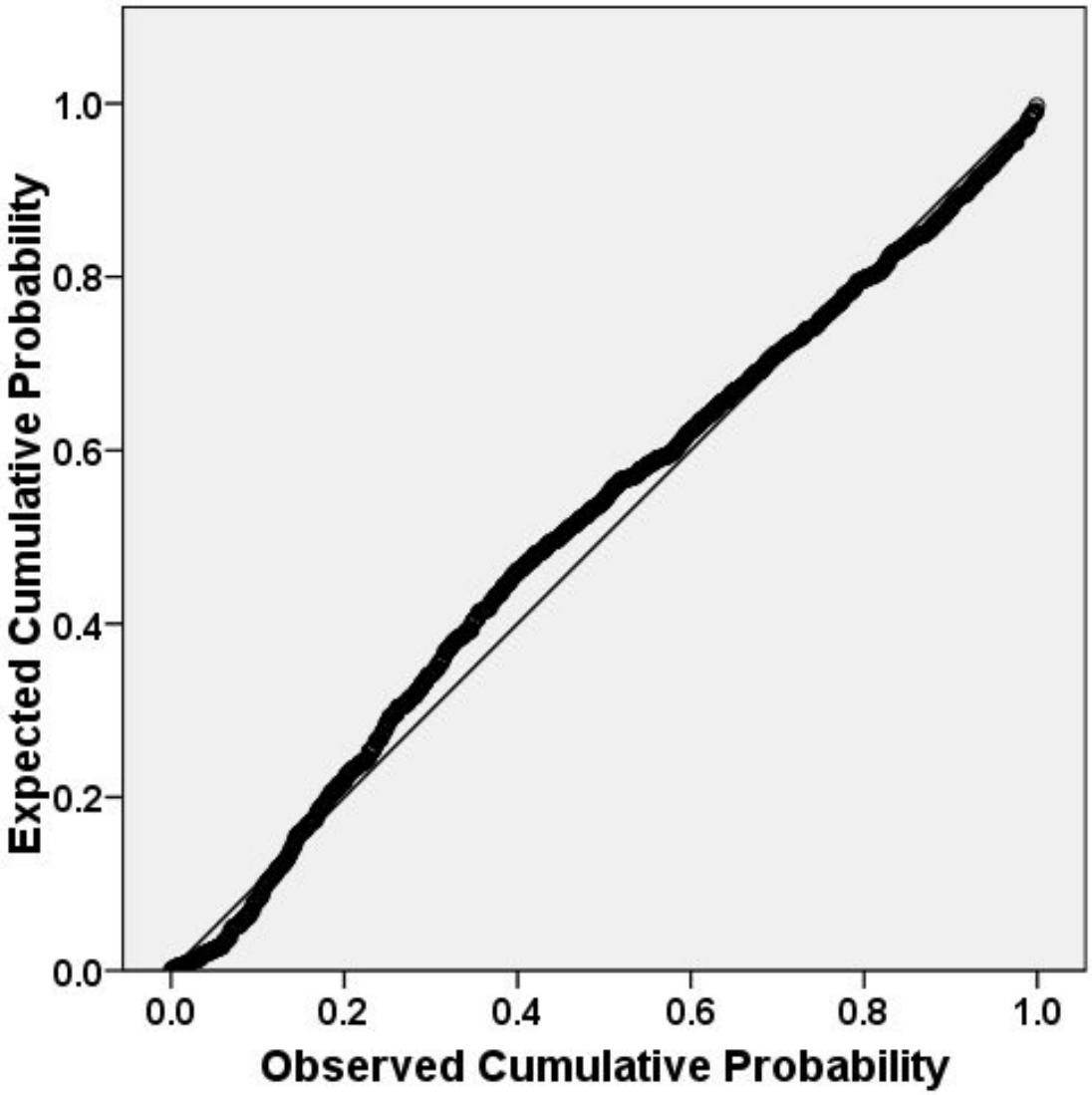
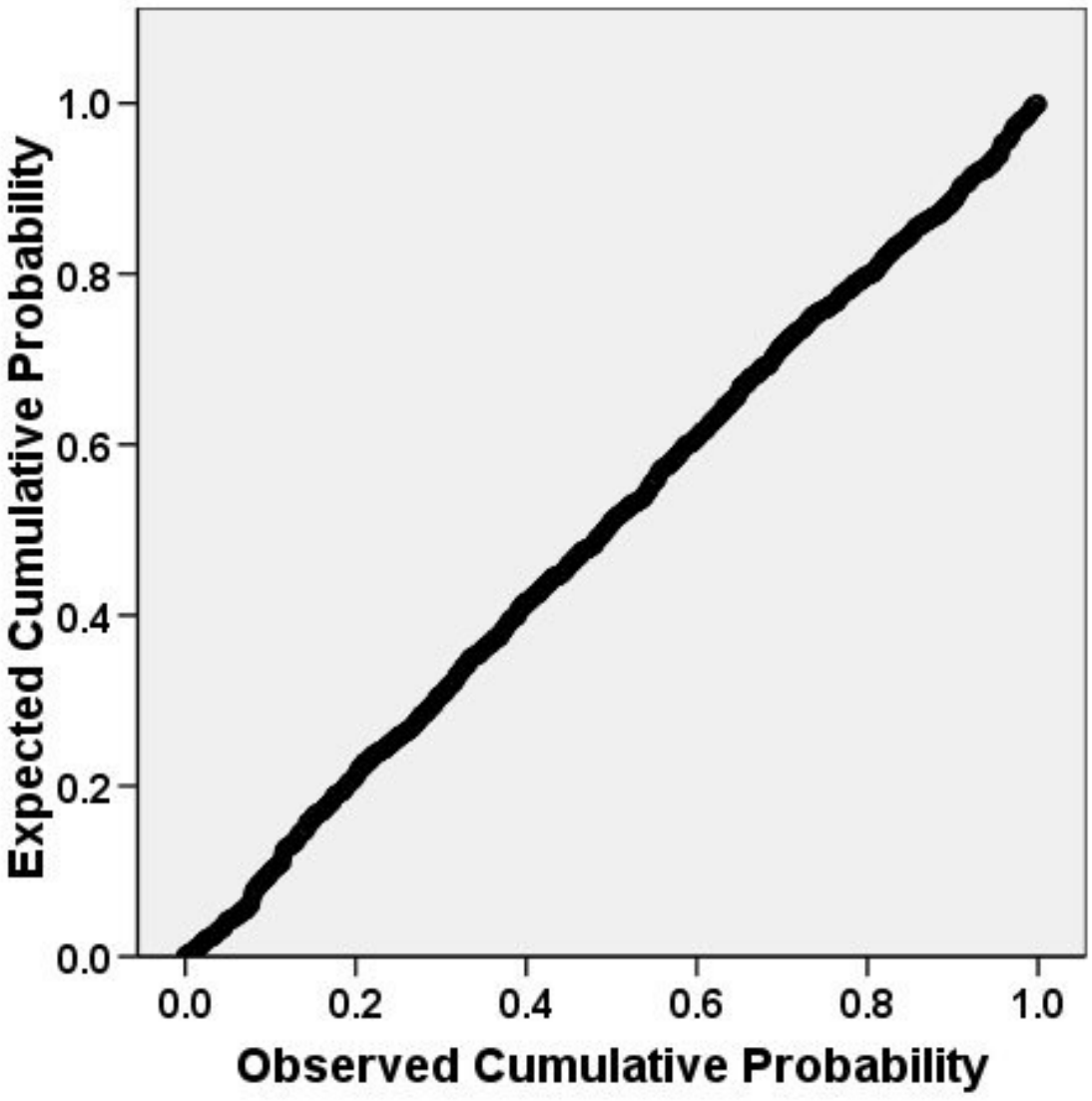


FIGURE 4.5

Pretest Normal Probability Plot of GM Fastlane Cognitive Diversity



Hypothesis Testing

Blog entry comments were characterized as influenced primarily by cultural tribalism or need-for-cognition and commenters in each blog were segmented into evangelists, idea seeders and flockers. Then, the models depicted in Figures 3.10 and

3.11 were estimated and hypotheses tested. Although this model is fully described in Chapter III, the segmentation of blog entry comments and commenters might be a source of confusion. So, before the estimation of the model is discussed the implementation details involving the segmented blog entry comments and commenters will be described.

Since the unit of analysis in the models of Figures 3.10 and 3.11 is the blog entry, values must be calculated for the 32 indicators used to reflect the eight latent variable constructs for each blog entry. As explained in Chapter III, the need for validated self-expression (as manifested in cultural tribalism) and need-for-cognition are competing motivations within blogs. Each blog entry is approached in retrospect, complete with all its comments. As each blog entry is examined, its comments are segmented based on equation 3.1 into those influenced primarily by cultural tribalism and those primarily influenced by need-for-cognition. Then summary measures of each segment, mean and standard deviations of individual and collective thought separation, are calculated to measure the effect of each segment, and its associated influence, on the whole blog entry. These summary measures are the indicators for the cultural tribalism and need-for-cognition latent variables in Figures 3.10 and 3.11.

The flocking, mass dissemination and idea seeding factors reflect relationships between summary measures of cognitive objects produced by individual commenters. For example, when calculating the values for the three indicators used to reflect flocking, all the data for the commenters previously identified as flockers who commented on the current blog entry are retrieved. From this data the mean number of clusters started and the mean number of comments posted, across all blog entries, is calculated.

Additionally, for each flocker who commented on the current blog entry, the standard deviation of the cognitive distance between all the comments they post in the lifetime of their participation and their associated cluster centroids is calculated. The joint reflective impact of the three indicators on the flocking latent variable becomes the impact of that construct on the cognitive diversity of the current blog entry.

The calculation of the indicators reflective of the other four latent variables is fully described in Chapter III and should be readily understood. As a result, the discussion will now turn to estimating the model.

TABLE 4.5

Pretest Multigroup Fitted Naïve Model Parameter Values

	Parameter Value		Hypothesis Support		LM Test
	Standardized	**Unstandardized**	**Statistical**	**Practical**	**Probability**
H_1	-0.0625	-0.0543 (t = -1.64)	Yes	Yes	0.000
H_2	0.6469	0.5435 (t = 9.93)	Yes	Yes	0.000
H_3	-0.0300	-0.0372 (t = -0.68)*	No	No	0.026
H_4	-0.1042	-0.0843 (t = -2.54)	Yes	Yes	0.000
H_5	-0.0017	-0.0017 (t = -0.04)*	No	No	0.000
H_6	-0.0748	-0.0721 (t = -1.61)	Yes	Yes	0.000
H_7	0.0015	0.0007 (t = 0.02)*	No	No	0.000

*Not significant

Prior to fitting the 32 indicator measurements to the models in Figures 3.10 and 3.11, with a pretest sample of 150 blog entries randomly selected from the GM Fastlane blog and the Blog for America, the models were translated into simultaneous equations. Then, multigroup structural equation modeling and maximum likelihood estimation (using EQS 6.1), with factor loadings and relationships between factors constrained

equal across blogs, was used to solve the naïve and alternative models and obtain the results reported in Tables 4.5 (Naïve: CFI = 0.568, RMSEA = 0.121, χ^2 (78) = 647) and 4.6 (Alternative: CFI = 0.883, RMSEA = 0.063, χ^2 (78) = 232), and Figures 4.6 and 4.7.

TABLE 4.6

Pretest Multigroup Fitted Alternative Model Parameter Values

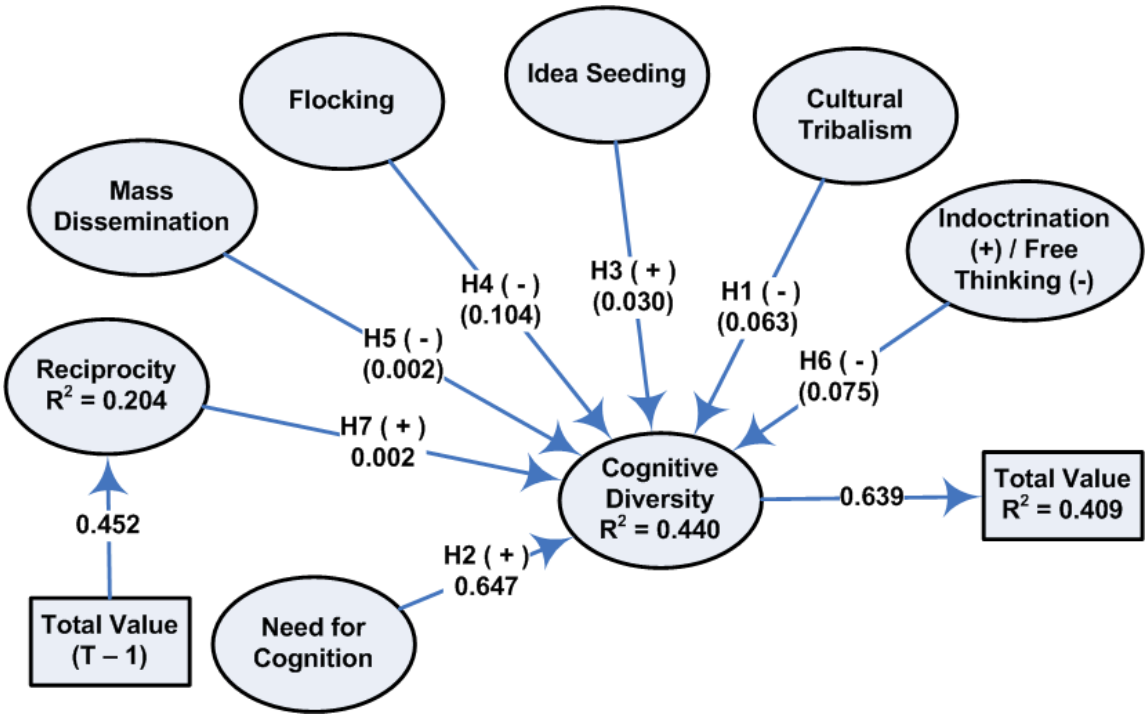
Path	Parameter Value		Relationship Support		LM Test
	Standardized	Unstandardized	Statistical	Practical	Probability
R_1	0.3757	0.4354 (t = 6.07)	Yes	Yes	0.000
R_2	-0.2108	-0.2427 (t = -4.34)	Yes	Yes	0.034
R_3	-0.0993	-0.1159 (t = -1.93)	Yes	Yes	0.000
R_4	0.1398	0.1720 (t = 3.00)	Yes	Yes	0.000
R_5	0.3021	0.3258 (t = 6.04)	Yes	Yes	0.003
R_6	-0.4416	-0.4498 (t = -8.13)	Yes	Yes	0.000
R_7	0.2761	0.3993 (t = 4.61)	Yes	Yes	0.010
R_8	-0.0862	-0.0770 (t = -2.24)	Yes	Yes	0.000
R_9	0.6483	0.5468 (t = 10.28)	Yes	Yes	0.000

As shown in Table 4.5, only H_1, H_2, H_4 and H_6 are supported as indicated by both statistical significance and the size of the unstandardized parameter (i.e., effect size). However, statistical significance is not noteworthy in the main part of this study, as the large sample size is assured to raise statistical power enough to make all the estimated values statistically significant. A more meaningful criterion is whether the sizes of the estimated parameter values (i.e., effect size) show they have *practical significance* (Kirk 1996). Kirk eschews the selection of a static and universal threshold value of practical significance to be applied to every scenario. This study arbitrarily defines practical significance as an unstandardized parameter estimate greater than 0.05.

In other words, an influence must have at least a 5% effect on a target to be considered practically significant. Tables 4.5 and 4.6 also show that none of the structural paths between factors pass the test for weight invariance across all blogs (LM Test Probability > 0.05). It is readily apparent that the alternative model fits significantly better than the naïve model (ΔCFI = 0.315, $\Delta\chi^2$ (0) = 415).

FIGURE 4.6

Naïve Factor Model Fitted with Pretest Data



Formal tests of mediation were performed on the model in Figure 3.11 whenever one construct influenced another through an intermediary construct (e.g., mass dissemination affects cognitive diversity through cultural tribalism in Figure 3.11). This

was done using Hardy and Bryman's (2004) procedure of bypassing each intermediary, one at a time, with a direct path and performing a χ^2-difference test. In accord with Cheung and Rensvold's (2002) recommendation, ΔCFI was also calculated. When bypassing a mediator construct fails to significantly improve model fit, the validity of the base model is supported. Baron and Kenny (1986) specified that mediation be tested by looking at the estimated parameter value of the path bypassing each mediator; values close to zero show the mediator is needed. The results of mediation tests are summarized in Table 4.7; all mediation tests support the proposed alternative model.

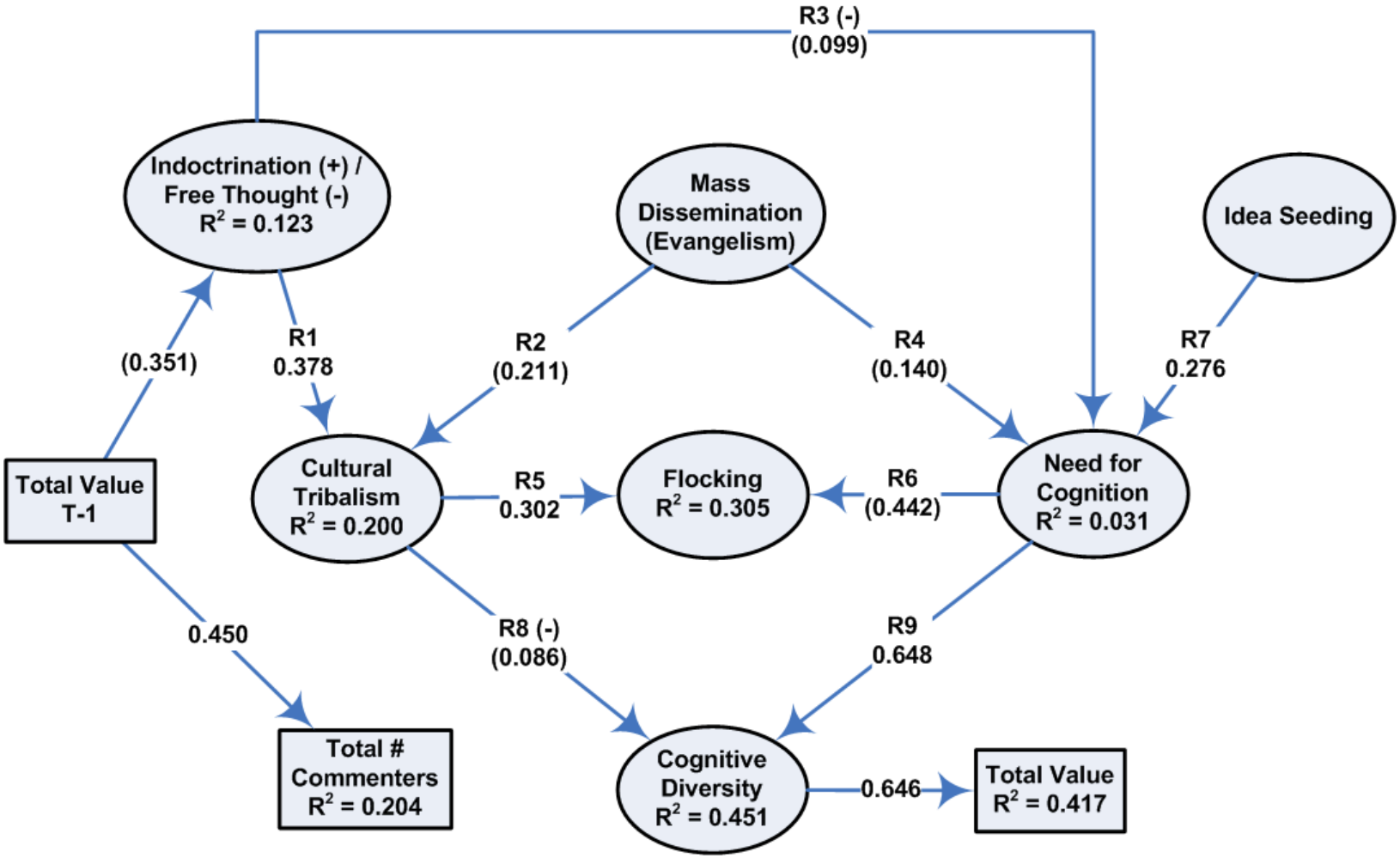
TABLE 4.7

Pretest Alternative Model Tests of Mediation

Relationship	CFI and X^2 Difference Test				Bypass Parameters
	CFI	X^2	Test of Significance ΔCFI	Test of Significance ΔX^2	
Baseline	0.883	χ^2 (78) = 232			
Indoctrination to Flocking	0.882	χ^2 (76) = 232	0.001	$\chi^2_d(2) = 0$	0.018 (t = 0.432)
Mass Dissemination to Flocking	0.883	χ^2 (76) = 230	0.000	$\chi^2_d(2) = 2, \rho < .5$	0.014 (t = 0.243)
Idea Seeding to Flocking	0.883	χ^2 (76) = 230	0.000	$\chi^2_d(2) = 2, \rho < .5$	0.104 (t = 1.156)
Indoctrination to Cognitive Diversity	0.890	χ^2 (76) = 221	0.007	$\chi^2_d(2) = 9, \rho < .02$	-0.034 (t = -0.972)
Mass Dissemination to Cognitive Diversity	0.883	χ^2 (76) = 230	0.000	$\chi^2_d(2) = 2, \rho < .5$	-0.119 (t = -2.333)
Idea Seeding to Cognitive Diversity	0.891	χ^2 (76) = 220	0.008	$\chi^2_d(2) = 12, \rho < .01$	0.034 (t = 0.437)

FIGURE 4.7

Alternative Factor Model Fitted with Pretest Data



Even though the fit of the naïve model to the pretest data is fairly good, when assessed in the context of the small sample size, three of the hypothesized relationships are both statistically insignificant and too small in effect size for the associated hypotheses to be considered supported at this stage. There are encouraging indications though, in the effect sizes and valance polarities of the relationships representing H_1, H_2, H_4 and H_6, where there is good support of the associated hypotheses.

The alternative model is both well fitted to the pretest data and has relationships between factors that are both statistically and practically significant. Note that the alternative model suggests explanations for the hypotheses not supported in the naïve model. The reason why the hypothesized association between idea seeding and cognitive diversity (H_3) was unsupported seems to be because the grassroots influence of need-for-cognition is required. Similar speculation seems justified for the hypothesized relationship between mass dissemination and cognitive diversity (H_5), the grassroots influences of cultural tribalism and need-for-cognition seem to be required. Such speculation also seems justified by the weakness of the hypothesized relationship between indoctrination and cognitive diversity (H_6) even though the hypothesis was statistically and practically supported. The unsupported hypothesized effect of reciprocity on cognitive diversity (H_7) may be due to an insufficient conceptualization of reciprocity. It seems the general principle behind the concept is sound, as lagged value was associated with increased numbers of participating commenters in the alternative model.

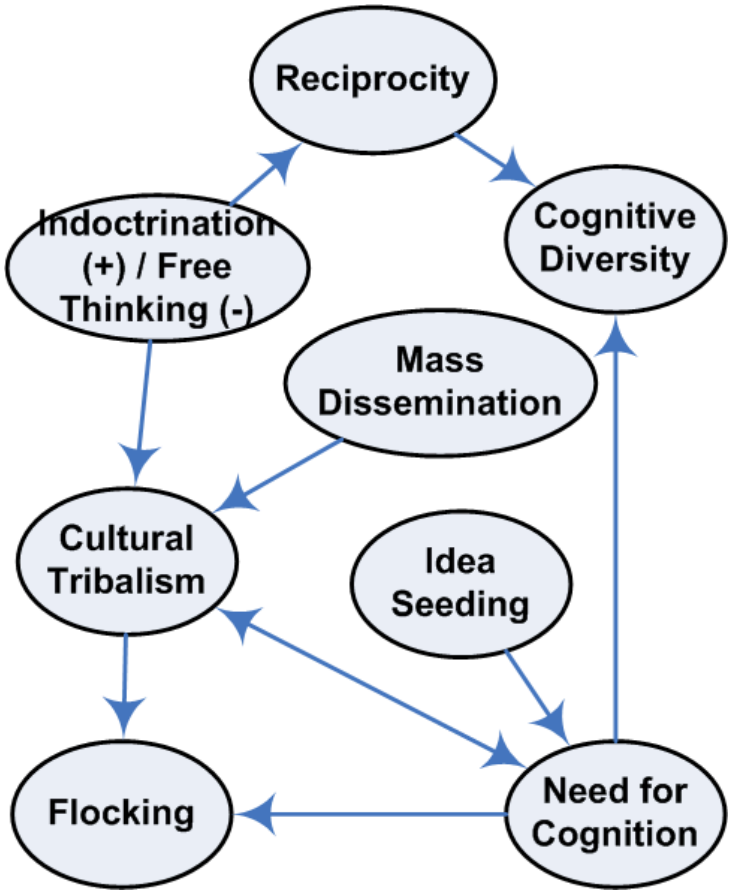
In this section of the chapter, two alternative models have been compared with varied results. It seems reasonable to ponder whether there is another model that would be superior compared to the models considered. In the next two sections, two novel methodologies are discussed that purport to find causal models from correlation data. These methodologies are used to search for better models that are then compared with the ones already described.

Directed Acyclic Graphs

Directed Acyclic Graphs (DAGs) are similar to structural equation models (SEMs) in that they are a pictorial diagram of the relationships between variables. The important difference is that where SEMs are representations of simultaneous linear equations, DAGs represent the findings of an artificial intelligence algorithm that uses correlations and metadata about temporal relationships between variables to propose an underlying causal structure. Swanson and Granger (1997) suggested that DAGs could be used as a data-driven means of inferring causal structure when little theory is available. Many alternative algorithms are used to construct DAGs; however the most well-known are Spirtes, Glymour and Scheines' (2000) PC search and Chickering's (2002) Greedy Equivalent Search (GES). Regular users of DAG algorithms (e.g., Zhang, Bessler and Leatham 2006) tend to place greater confidence in relationships found by multiple algorithms over those found by just one.

FIGURE 4.8

PC Search DAG

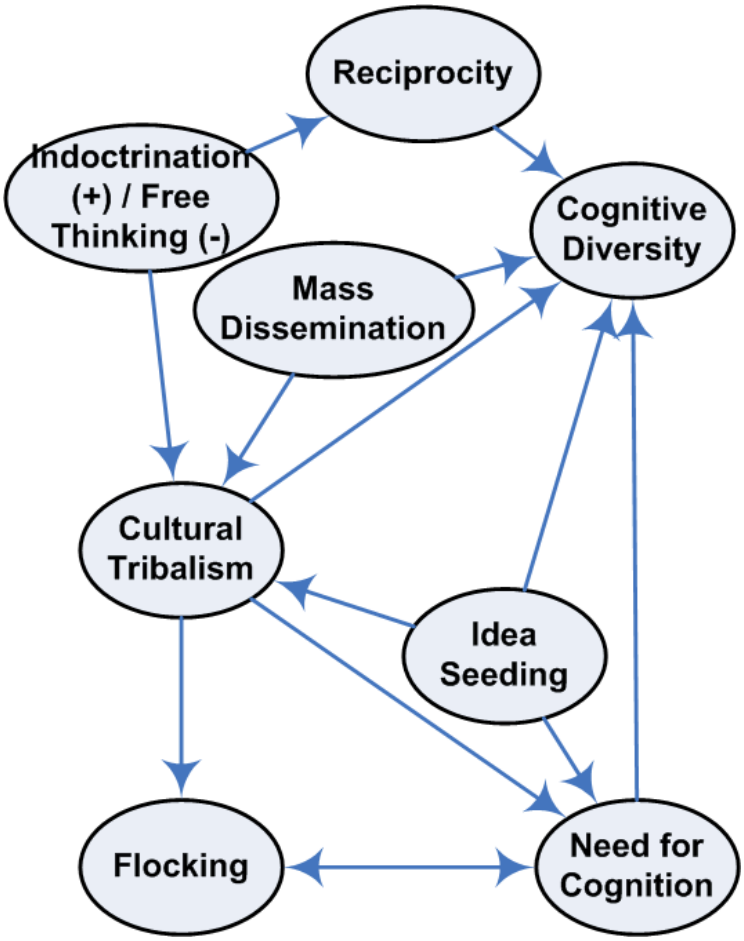


PC Search

The PC search algorithm begins by connecting every variable with an undirected edge (i.e., a line connecting them in the diagram). Edges are removed where correlation is either zero or conditional on other relationships between variables. Edges are made directed (i.e., given an arrow head) in a subsequent step too complex to be described here. When PC search was applied to the raw correlation matrix underlying that in Table 4.4, the DAG depicted in Figure 4.8 resulted.

FIGURE 4.9

Greedy Equivalent Search DAG



Greedy Equivalent Search

The Greedy Equivalent Search (GES) algorithm is a good complement to PC search because it begins by assuming that no edges connect the variables. It uses a two-step iterative procedure and a Bayesian analysis of probabilities to converge on a highest likelihood set of edges between the variables. When GES was used on the raw version of Table 4.4's correlation matrix, the DAG in Figure 4.9 emerged.

FIGURE 4.10

PC Search and GES Commonality DAG

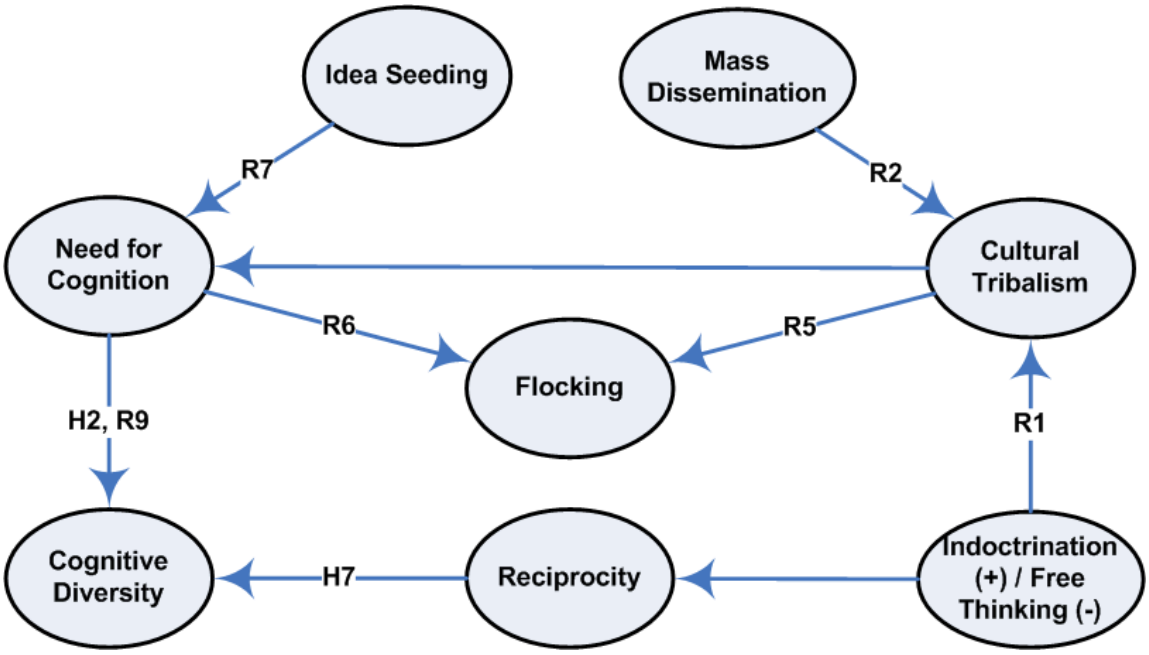


Figure 4.10 shows the edges found by both the PC Search and GES algorithms. The edges that match those on the naïve and alternative models are labeled. The findings of the data-driven algorithms lend support for some of the relationships proposed in the theory-based alternative model. The inclusion of one of the unsupported hypotheses from the naïve model (H_7) in a mediator role between indoctrination and cognitive diversity is interesting and worthy of further thought.

Limitations

Barnett and Woelfel (1979) point out that it can be difficult to select the optimal number of dimensions in the multidimensional space where cognitive objects are plotted.

I propose to use four dimensions in this study because as dimensionality increases, the computational tasks for self-assembling thematic clusters rise dramatically along with the computer memory required.

The results of this study, while relevant to the broad spectrum of marketing phenomena, may not be completely replicable in a blog study closer to the core business focus of mainstream marketing research. As stated in Chapter III, the social forms in this study are often highly dependent on initial conditions, are path dependent, nonlinear and only loosely coupled to each other. They often also involve random delays and different timescales. The more specific in context a future blog study may be, the more likely that context will be affected by a lesser number of social processes.

The purposive sample of blogs may cause the effects of processes that may be rare, such as indoctrination, to appear to have greater importance, relative to the other processes, than is true of the overall blogosphere.

Summary

In this chapter, a means of testing the models proposed in Chapter III was described and its effectiveness demonstrated with pretest data. In the next chapter all 15 blogs in the dataset are subjected to the testing procedure described here to create a basis for final determinations about the proposed hypotheses and models.

CHAPTER V

RESULTS

Chapter IV described this study's data set and a multifaceted series of procedures designed to test the hypotheses and models introduced in Chapter III. The data set is a diverse selection of 15 blogs, selected using guidelines specified by Kozinets (2002) to ensure a data set with high information content. The blogs were also selected so as to cover the full breadth of marketing activities: for-profit and non-profit, organizational and individual. Blog entries and associated comments were converted into word networks by Centering Resonance Analysis (CRA). The similarity (or resonance) between these word networks was measured and used as an input to multidimensional scaling to locate blog entries and comments as points in a cognitive space. Standard clustering algorithms were used to find thematic clusters among the points in cognitive space. The clusters were measured on such dimensions as size and density to derive a measure of cognitive diversity. Cognitive diversity metrics were then used to segment blog entry comments into those most influenced by cultural tribalism and need-for-cognition. Such metrics were also used to segment commenters into idea seeders, evangelists and flockers. Multi-group structural equation modeling was then used to test a model and hypotheses that related the joint influences of blog author and commenters on the diversity of thought expressed in blogs.

This chapter begins by describing the results of conducting the tests and procedures described in Chapter IV on all 15 of the blogs in the full dataset. On the basis

of those results, final determinations are made concerning support for the proposed hypothesis and models.

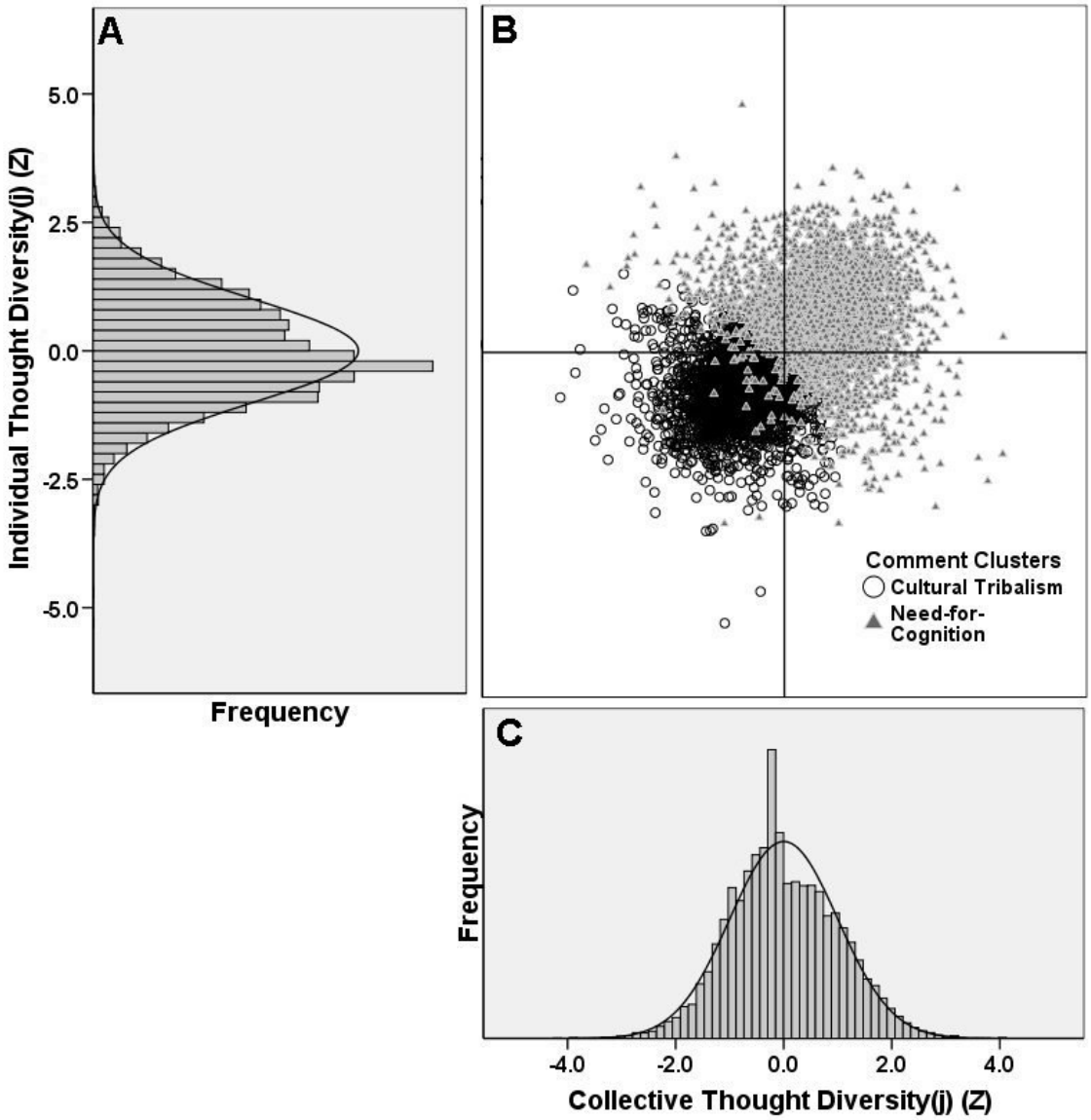
Classification of Blog Entry Comments

Chapter III described how blog entry comments can be classified as either influenced primarily by cultural tribalism or need-for-cognition using the discriminant model equation 3.1 or, by plotting the positions of comments in the conceptual map shown in Figure 3.2, derived from equation 3.1. The two comment types were distinguished, using Chui et al's (2001) clustering algorithm using all the blog entries and comments from each of the 15 blogs. Two clusters emerged and cluster membership was confirmed using McQueen's (1967) k-means. When a random sample of 500 comments was drawn from each blog in the dataset, pooled and plotted on the conceptual map of Figure 3.2, the result is as displayed in Figure 5.1 (B). While the results are not perfectly conforming to the proposed perceptual map regions, the higher density of data points within the so-called cultural tribalism and need-for-cognition quadrants suggests support for the proposed topology for blog entry comment segmentation. It is easy to see that the clustering algorithms merely cut the mass of comments in the middle. To get a more nuanced insight, it is necessary to look at the comment frequency histograms below and to the left of the scatter plot (A and C). Even though the two clusters contain approximately equal numbers of points (50.3% cultural tribalism), the cultural tribalism cluster is denser, as shown by its smaller footprint in the scatter plot and taller height in both the individual and collective thought diversity

histograms. This observation was confirmed by a statistically significant difference (F = 28.53) between the mean distance between cluster members and their cluster centers.

FIGURE 5.1

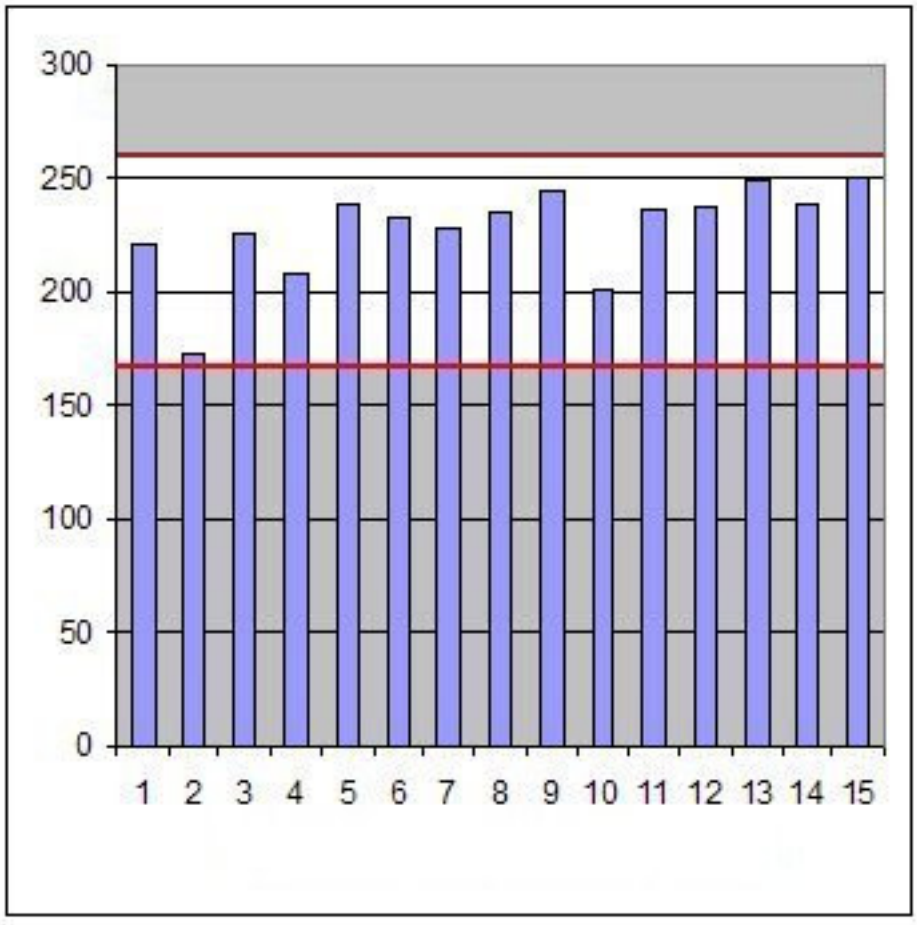
Blog Entry Comment Mapping



This observation is consistent with cultural tribalism theory; not only are both types of thought diversity low, but there is less variation even at those low levels. It is also interesting to note that high and low levels of individual and collective thought diversity seem to occur together; the two are moderately correlated 0.450 ($\rho < 0.001$).

FIGURE 5.2

Blog Cultural Tribalism Histogram

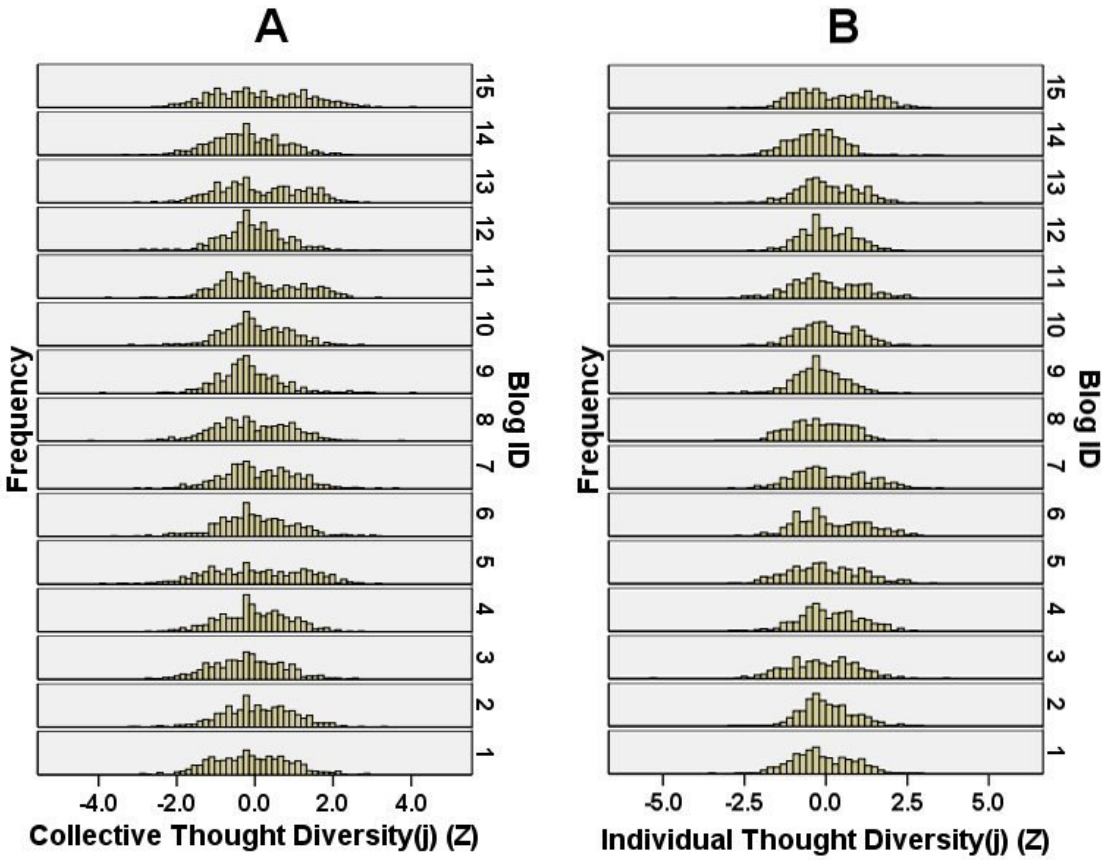


The histogram in Figure 5.2 addresses the question of whether the blogs differ in the number of comments classified as predominantly influenced by cultural tribalism. The white band indicates a 3σ confidence interval around the mean. There is no statistically significance difference in the level of cultural tribalism between the blogs.

To gain another perspective on the question of whether blogs differ in the relative effects of cultural tribalism versus need-for-cognition, the underlying distributions for individual and collective thought diversity were compared in the paneled histograms of Figure 5.3. In Figure 5.3, the blogs are identified with a number corresponding to those in the ID column of Table 4.1. It is apparent from Figure 5.3 A and B that the blogs differ in these underlying distributions, with some seeming to have skewed (e.g., 10, 11 in 5.2 A), normal (e.g., 1, 6) and bimodal (e.g., 5, 13, 15) distributions. The skewed distributions all favor the low end of the scale, perhaps indicating cases where the whole blog is characterized by cultural tribalism. Bimodal distributions may indicate blogs in transition, where idea seeding is in conflict with either indoctrination or mass dissemination for ideological dominance.

FIGURE 5.3

Paneled Histograms of Thought Diversity

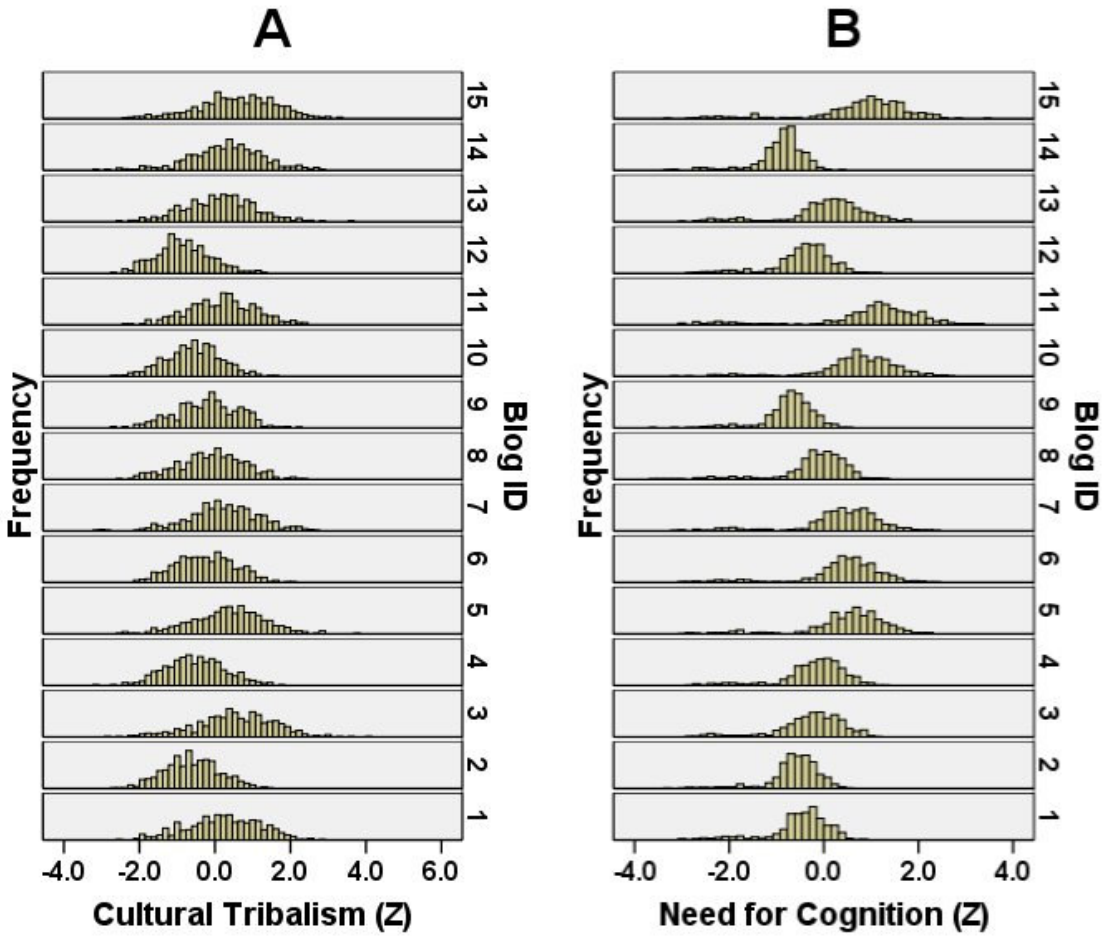


To address these speculations more definitively, paneled histograms of estimated values for the need-for-cognition and cultural tribalism latent variables were constructed (Figure 5.4). All the blogs seem to have similarly shaped distributions even though they are located on different parts of their underlying scale; they seem to have normally distributed levels of cultural tribalism and positively skewed levels of need-for-cognition. Since these two constructs share in common the individual and collective

thought diversity indicators, and are conceptualized as polar opposites, it may be that the normal distribution of cultural tribalism is a magnification of the elongated negative tail of the need-for-cognition distribution. Should this be true, then all the distributions of Figure 5.3 are bimodal, some merely more apparent than others.

FIGURE 5.4

Paneled Histograms of Cultural Tribalism and Need for Cognition



Regardless of whether the distributions of Figure 5.3 are bimodal or not, the evidence suggests that need-for-cognition is a universally dominant influence in blogs, something that bodes well for the possibility of high general levels of cognitive diversity.

The parameters of equation 3.1 were estimated from the dataset and the accuracy of the resulting model tested with a holdout sample. One discriminant function emerged from the estimation with a characteristic root eigenvalue greater than 1. This function correctly classified 90.2% of the holdout sample. The coefficients of the discriminant model are given in Table 5.1.

TABLE 5.1

Comment Discriminant Function Coefficients (Equation 3.1)

Parameter	Standardized	Unstandardized
β_4 (Stdev Individual)	0.457	0.541
β_3 (Mean Individual)	0.483	0.565
β_2 (Stdev Collective)	0.486	0.553
β_1 (Mean Collective)	0.467	0.547
β_0 (Constant)		-0.042

The only interesting theoretical insight from Table 5.1 is that all the parameters play a near equal role in separating comments primarily motivated by need-for-cognition from those motivated by cultural tribalism.

Classification of Commenters

Chapter III described how commenters can be characterized as evangelists, idea seeders and flockers by plotting their positions on a conceptual map (Figure 3.8, derived from equation 3.2). Idea seeders can be separated from evangelists by high diversity in thought expression: idea seeders express thoughts about a wide variety of things while evangelists always express similar thoughts. Idea seeders and evangelists can be separated from flockers on the basis of differences in boldness: idea seeders and evangelists are bolder when expressing their thoughts.

When the commenting activity of 7500 commenters, 500 randomly selected from each of the blogs in the dataset, was pooled, converted to z-scores and demarcated by Chiu et al's (2001) clustering algorithm, three clusters emerged (see Figure 5.5 B). These cluster assignments were confirmed using McQueen's (1967) k-means algorithm. Compare Figure 5.5 (A) with Figure 3.7, where idea seeders and evangelists are seen as occupying opposite tails of a normal distribution along an individual thought diversity axis. In Figure 5.5 (A) the distribution fits the rough pattern of a normal distribution but the center is distorted into a sharp peak. It seems as though commenters near the mean exhibit (or perhaps are drawn into) greater similarity in cognitive consistency (the cognitive distance between an individual's thought expressions) than commenters away from the mean on either tail. This higher density section appears to be within the cluster labeled "Flocking" or "Flockers" in Figure 5.5 (B). It was predicted that flockers would be below average in boldness but unrestricted in their overall thought diversity. While that appears to be generally true, it seems that the "average" flocker perceives and

maintains (perhaps through conscious effort or innate tendency to conform) an ideal level of consistency in their expressed thoughts. Figure 5.6 (B) shows this phenomenon is consistent across blogs.

FIGURE 5.5

All Blog Commenter Clusters

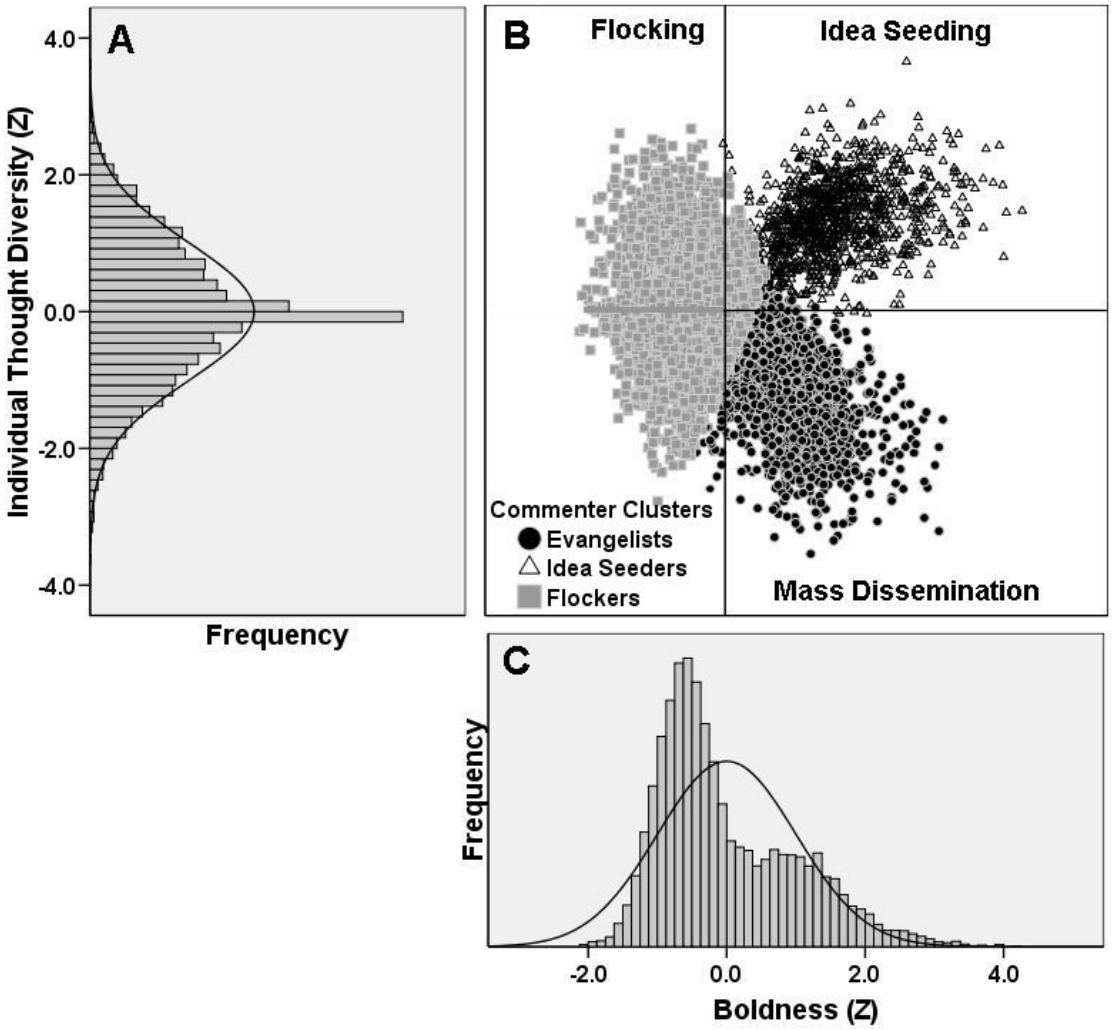
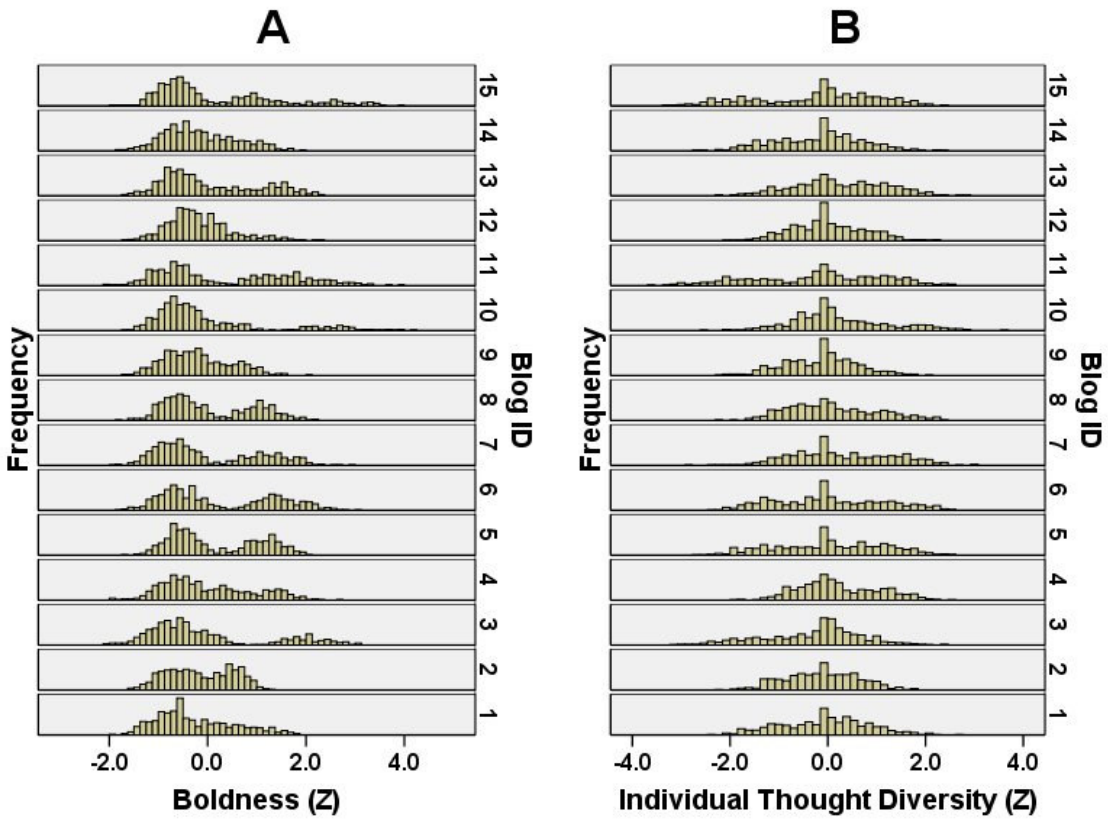


FIGURE 5.6

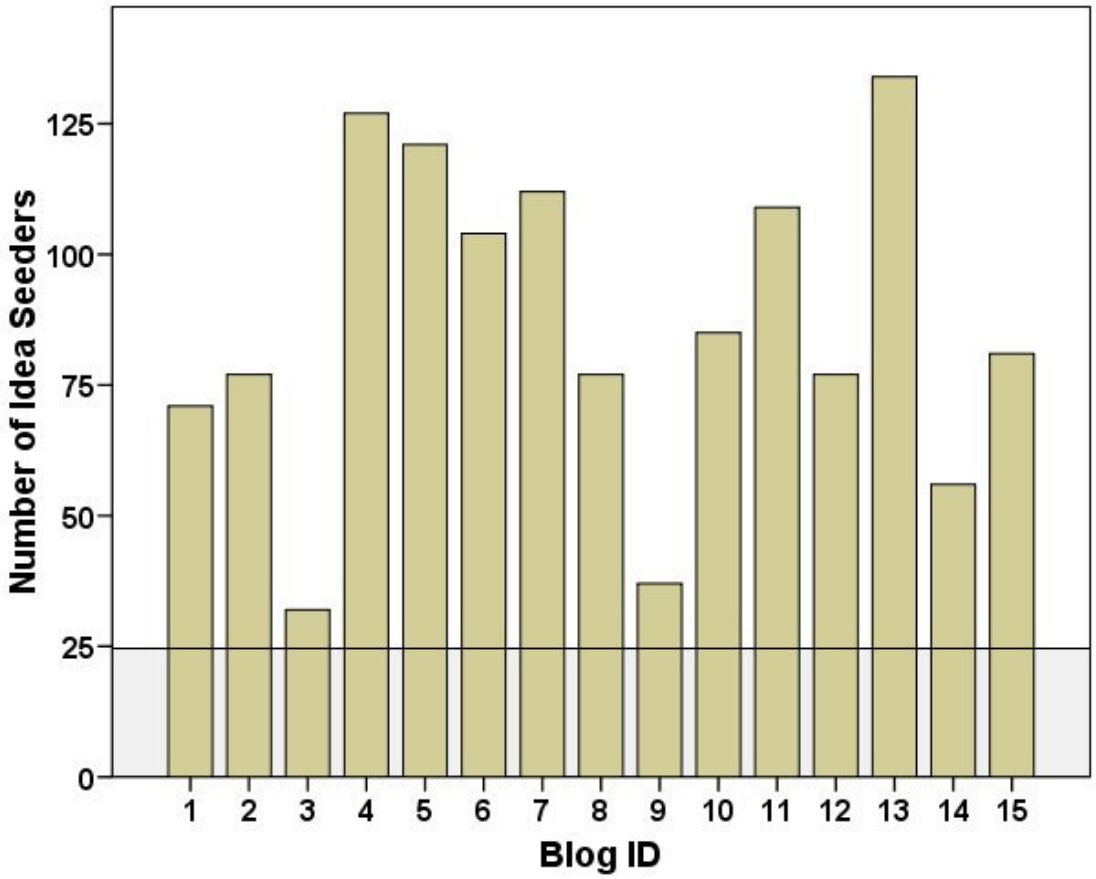
Paneled Histograms of Boldness and Individual Thought Diversity



This pattern of things in the same vicinity being drawn to a common point is similar to the physics concept of a *fixed point attractor*, a stationary point that exerts a gravity-like force on nearby objects, drawing them closer. Here, evidence of a *social attractor* may be present. A full discussion of these results here is out of context; however, they are discussed in Chapter VI. Figure 5.7 shows that unrestricted thought freedom, as personified by idea seeders, is consistent across all blogs (the white band is a 2σ (95%) confidence interval around the mean).

FIGURE 5.7

All Blog Idea Seeding Histogram

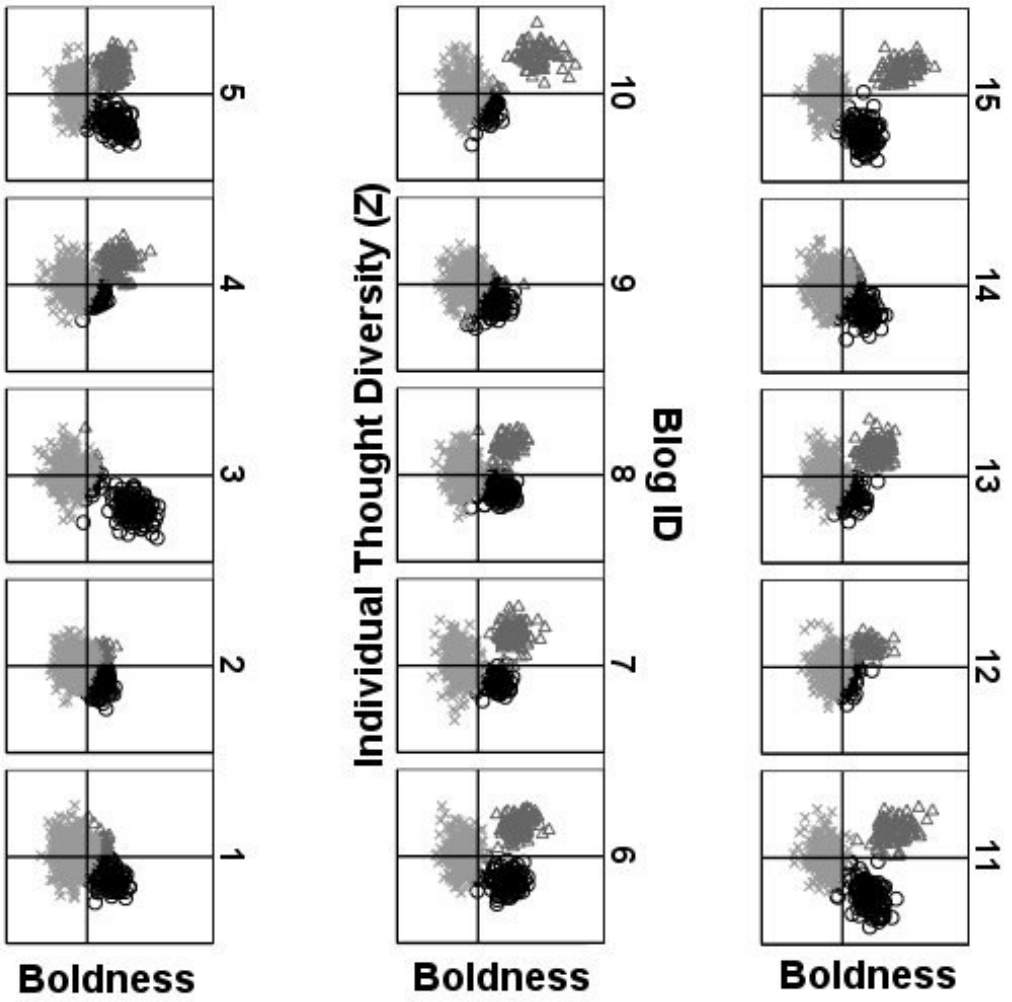


Since the scatter plot of Figure 5.5 (B) holds data from all 15 blogs, it might be considered indicative of the diversity of boldness and individual thought that characterizes the whole blogosphere. If that is so, then comparing it with similar plots using data from one blog might be a qualitative method of assessing cognitive diversity. Figure 5.8 contains such plots of commenters from all the individual blogs. While most plots contain the same three basic clusters the size and location of the clusters vary

noticeably. For example, it could be strongly argued that blog 6 has the profile of a blog that reflects the diversity present in the whole blogosphere while such blogs as 12, 13 and 14 are more limited. Note that no comparisons can be made with diversity of thought in the whole market; that must be deferred to future research.

FIGURE 5.8

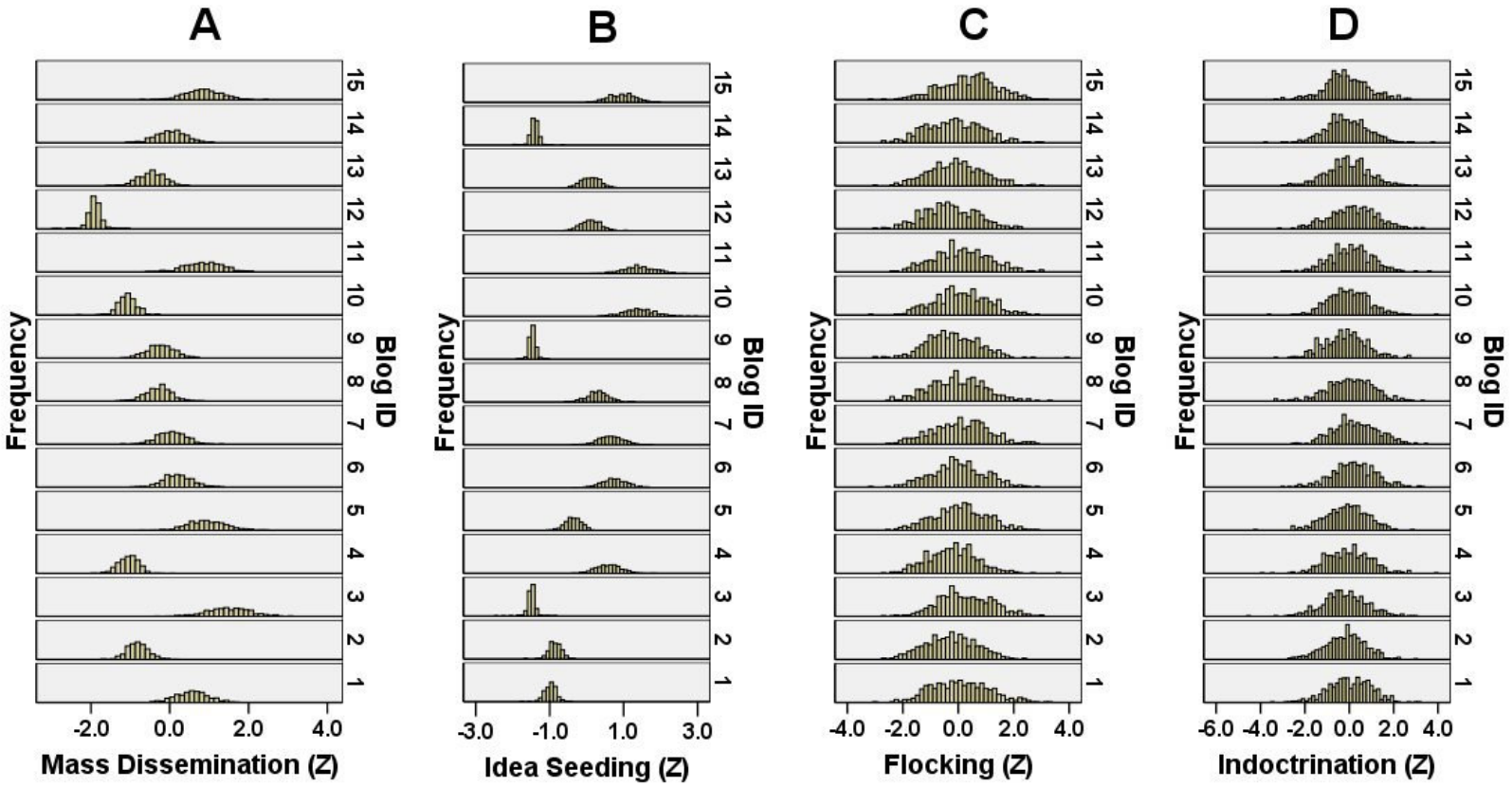
Individual Blog Commenter Clusters



In Figure 5.8, the clustering algorithms group commenters into similarly located clusters in all the blogs, with positions roughly as predicted in Figure 3.8. While interesting, the scatter plots of Figure 5.8 are more instructive when viewed in conjunction with Figures 5.6 and 5.9. The boldness histograms in Figure 5.6 (A) generally show bimodal distributions, two separate normally distributed populations on either side of the mean. The individual thought diversity histograms (Figure 5.6 B) are consistent in showing one wide normally distributed population, punctuated with a center peak. The alternative model depicted in Figure 3.11 proposed indoctrination, mass dissemination and idea seeding to be the exogenous influences driving the other constructs, Figure 5.9 shows that levels of those influences, as well as flocking, are roughly normally distributed. Indoctrination seems invariant across blogs, while flocking changes little. Variance in these influences does not seem to be an obvious explanation for the variation in commenter cluster configurations shown in Figure 5.8. However, levels of mass dissemination and idea seeding vary widely and seem logically correlated with changes in Figure 5.8. Note, for example, that low levels of mass dissemination in Figure 5.9 (A) blogs 10 and 12 match small mass dissemination clusters in Figure 5.8. A similar pattern exists for idea seeding in blogs 3 and 9.

FIGURE 5.9

Paneled Histograms of Blog Author and Commenter Types



In preparation for estimating the parameters of equation 3.2, the commenters were forcibly classified into the segments of Figure 3.8. Using these topology-based segment membership assignments, the parameters of equation 3.2 were estimated and the accuracy of the resulting model tested with a holdout sample. Two discriminant functions emerged from estimation with data from each of the blogs, correctly classifying an average of 92.1% of their respective holdout samples. Means of the coefficients of the respective discriminant models are given in Table 5.2.

TABLE 5.2

Commenter Discriminant Function Coefficients (Equation 3.2)

Parameter	Function 1		Function 2	
	Standardized	**Unstandardized**	**Standardized**	**Unstandardized**
B_5 (Stdev Comment to Cluster)	0.419	0.671	-0.010	-0.016
β_4 (# Clusters)	0.503	0.810	-0.012	-0.019
β_3 (Stdev Individual)	0.035	0.043	0.714	0.897
β_2 (Mean Individual)	0.038	0.046	0.639	0.777
β_1 (# Comments)	0.391	0.597	-0.051	-0.078
β_0 (Constant)		-0.063		-0.033

Construct Validity

As stated in Chapter IV, construct validity focuses on the extent to which data exhibits (1) convergent validity, the extent to which different assessment methods concur in their measurement of the same construct (also called factorial or measurement invariance), and (2) discriminant validity, the extent to which the measures of different constructs are distinct (Campbell and Fiske 1959). The demonstration of factorial invariance began by establishing baseline confirmatory factor analysis (CFA) results for each blog dataset. The results were then compared to see if the same indicators loaded on the same factors. After the factorial structure was found to be the same across blogs, hypotheses that each *free* factor loading (each factor's "first" indicator is constrained to unity and is thus not free) was equal across blogs was tested by forcing the factor loadings to be equal and then observing overall fit statistics as well as the LaGrange Multiplier (LM) Test univariate probability for each factor loading. This probability is a measure of significance for the LM Test, a test of the hypothesis that a factor loading varies across groups (in this case, blogs). LM Test probabilities less than 0.05 are considered statistically insignificant. That is, when the LM Test probability is less than 0.05 the parameter does not have the same value across the blogs. This procedure does not require equal sample sizes, nor does the test of causal structure invariance described below.

Invariance of causal structure was tested in a manner similar to the above. The same structural model was simultaneously fitted to the 15 blog datasets while constraining all factor loadings, structural path loadings and one error covariance equal.

Then multigroup model fit statistics were observed as well as the LM Test univariate probability for each constrained parameter.

TABLE 5.3

Normalized Estimates of Mardia's (1970) Multivariate Coefficient

Blog	Mardia's Coeff.	Blog	Mardia's Coeff.
AutoBlog	1.562	Gizmodo	3.146
Blog for America	0.857	Google Blogoscoped	2.244
Blog Maverick	2.561	Joystiq	6.435
The Consumerist	3.715	PaulStamatiou	2.359
EnGadget	2.762	Townhall	2.103
The Evangelical Outpost	4.243	TV Squad	3.204
Fastlane	1.208	The Unofficial Apple Weblog	3.401
Freakonomics	2.493		

Underlying the use of structural equation modeling to perform CFA is a strong assumption of multivariate normality. If evidence suggests that this assumption is violated it has important implications for the interpretability of the findings. None of the indicators in the data set demonstrated significant nonzero univariate skewness (one tail of the distribution is longer than the other) or kurtosis (variance is due to infrequent extreme deviations from the mean). Normalized estimates of Mardia's (1970) multivariate coefficient, an aggregate indicator of kurtosis, are given in Table 5.3. Bentler (2005) suggests that only values greater than 5.0 indicate non-normally distributed data. Another test of multivariate normality is discussed later.

The individual blog baseline CFA results showed the same indicators loading on the same factors as given in Table 5.4. Additionally, as Table 5.4 also shows, all but one of the 32 core factor construct indicators met Hair et al's (2005) criterion for statistical significance (factor loading > 0.4 with a sample size > 200) and 27 met the criterion for practical significance (loading > 0.5). It must be noted that Hair et al's criteria, as well as most of the reference values quoted in this study, applies to evaluating instruments used to gather experimental data. The circumstances of this study are much different as Hair et al's criteria are being used to evaluate a means of gaining insights from non-experimental, albeit primary, empirical data. Therefore, it is argued that Hair et al's criteria must be referenced with caution as the level of control over extraneous variation that characterizes experiments is absent here. That being understood, the general conformance of this study's factor loadings and other metrics to those recommended by Hair et al. (2005) and Byrne (2006) is worthy of note and indicative of a reasonable level of performance. The multigroup CFA with factor loadings constrained equal across blogs had good fit (CFI = 0.884, RMSEA = 0.025, χ^2 (3244) = 22,788) indicating that all blogs have similar, if not the same, factor structures. Table 5.4 also shows that all of the 32 factor loadings exhibited a less than 5% LM Test univariate probability, indicating that their factor loadings are not the same across the blogs even though the structure appears to be the same.

Discriminant validity was assessed in two stages: (1) correlations between constructs, corrected for attenuation, were confirmed significantly less than unity (Tables C-1 to C-15 in Appendix C contain the correlation matrices for all 15 blogs), and

(2) a χ^2 difference test was performed between the baseline multigroup CFA model (CFI = 0.875, χ^2 (436) = 35,515) and an alternative multigroup model (CFI = 0.403, χ^2 (464) = 167,750) where the correlation between each latent construct was constrained at unity. According to Cheung and Rensvold (2002), a large CFI (i.e., greater than 0.01) or χ^2 difference demonstrates discriminant validity. It is readily apparent that a ΔCFI of 0.472 and a $\Delta\chi^2$ of 132,235 with degrees of freedom equal to 28 argue strongly in favor of discriminant validity.

TABLE 5.4

Confirmatory Factor Analysis, Reliability, Variance Explained and Factorial Invariance

Main Factor and Reliability	Indicator	Loading	R^2	LM Test Probability
Cognitive Diversity r = 0.711	Mean Cluster Density	-0.651	0.424	0.000
	Mean Cluster Radius	0.593	0.352	0.000
	Number of Clusters	0.577	0.333	0.000
	Percentage of Outliers	0.645	0.417	N/A
Cultural Tribalism r = 0.683	Mean Collective Thought Separation	-0.525	0.276	0.000
	Stdev Collective Thought Separation	-0.518	0.268	0.000
	Mean Indiv. Thought Separation	-0.500	0.250	0.000
	Stdev Indiv. Thought Separation	-0.534	0.285	N/A
	Mean Commenter Longevity	0.533	0.284	0.000
	% Repeat Commenters (T - 1)	0.454	0.206	0.005
Flocking r = 0.558	Mean Clusters Started	-0.525	0.275	0.000
	Stdev Indiv. Thought Separation From Cluster Centroid	-0.585	0.342	N/A
	Mean Number of Comments	-0.518	0.269	0.000

TABLE 5.4 (CONT)

Main Factor and Reliability	Indicator	Loading	R^2	LM Test Probability
Idea Seeding	Mean Clusters Started	0.758	0.575	0.002
r = 0.852	Stdev Indiv. Thought Separation From Cluster Centroid	0.829	0.687	0.000
	Mean Number of Comments	0.748	0.559	0.000
	Mean Indiv. Thought Separation	0.537	0.289	0.000
	Stdev Indiv. Thought Separation	0.598	0.358	N/A
Indoctrination /	Mean Time Between Entries	-0.598	0.357	N/A
Free Thought	Mean Entry-to-Entry Separation	-0.578	0.334	0.000
r = 0.603	Stdev Entry-to-Entry Separation	-0.563	0.317	0.000
Mass	Mean Clusters Started	0.468	0.219	0.000
Dissemination	Stdev Indiv. Thought Separation From Cluster			
r = 0.690	Centroid	0.608	0.369	0.000
	Mean Number of Comments	0.534	0.286	0.000
	Mean Indiv. Thought Separation	-0.639	0.408	0.000
	Stdev Indiv. Thought Separation	-0.464	0.215	N/A
Need for	Mean Collective Thought			
Cognition	Separation	0.540	0.291	0.000
r = 0.635	Stdev Collective Thought Separation	0.384	0.148	0.000
	Mean Indiv. Thought Separation	0.405	0.164	0.000
	Stdev Indiv. Thought Separation	0.599	0.359	N/A
	% First Time Commenters	0.539	0.291	0.000
Reciprocity				
r = 1.0	Total Number of Commenters	1.0	1.0	N/A

Note: $\rho < .05$. Reliabilities are calculated from the indicators to each individual factor in Table 5.4. Pearson "r" is reported instead of Cronbach's α. N/A in the right hand column denotes the arbitrary first factor.

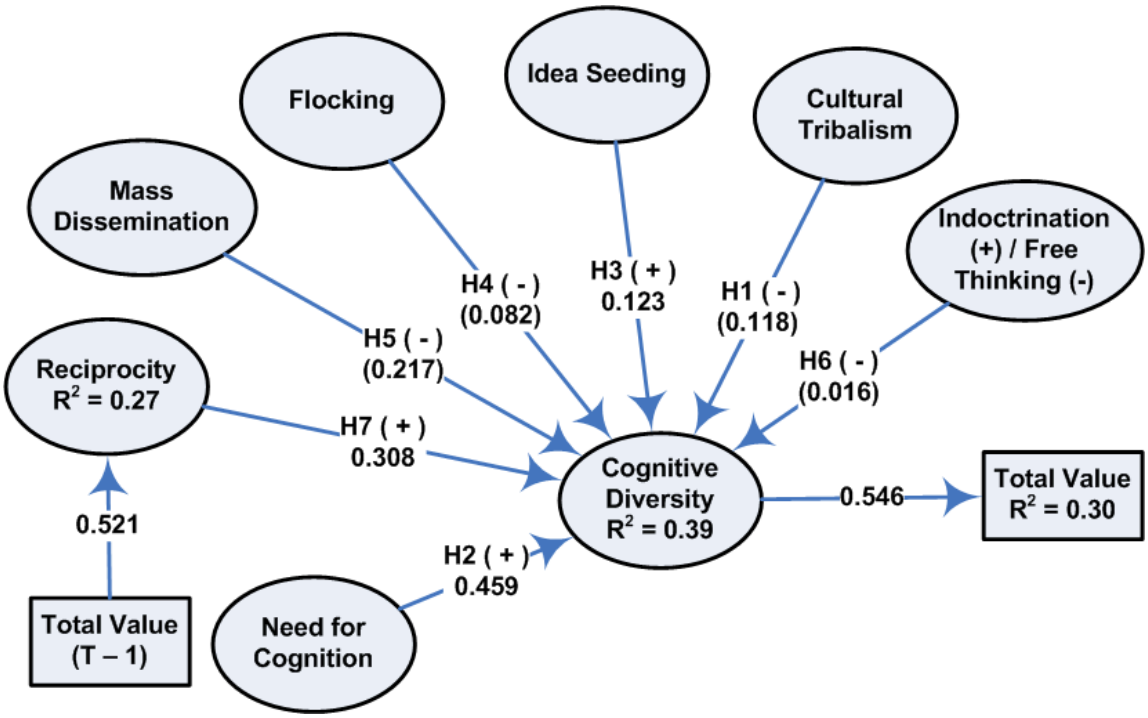
As has already been intimated, prior to conducting a confirmatory factor analysis, and the model estimation, it is important to test for multivariate normality in the factor indicators. This is particularly important in this study as many measures of collective phenomena have an underlying power-law distribution, as discussed in Chapter II, rather than a normal distribution. Hair et al. (2005) point out the difficulty in testing multivariate normality (the joint normality of more than one variable) and suggest that testing univariate normality is an acceptable substitute, especially if sample sizes are large. Hair et al. recommend graphical analysis using the normal probability plot as well as statistical techniques such as skewness and kurtosis (already tested with Mardia's (1970) coefficient as shown in Table 5.3). Figure C-1 in Appendix C shows a normal probability plot for estimated values of each blog's cognitive diversity latent variable. These plots show no normality problems in agreement with Table 5.3.

Hypothesis Testing

Blog entry comments were characterized as influenced primarily by cultural tribalism or need-for-cognition and commenters in each blog were segmented into evangelists, idea seeders and flockers. Then, the models depicted in Figures 3.10 and 3.11 were estimated and hypotheses tested.

FIGURE 5.10

Fitted Core Naïve Model



Prior to fitting the 32 indicator measurements to the models in Figures 3.10 and 3.11, with the blog data, models were translated into simultaneous equations. Then maximum likelihood estimation (using EQS 6.1) was used, with factor loadings and paths between factors constrained equal across blogs, to solve the multigroup models and obtain the results reported in Tables 5.5 (Naïve: CFI = 0.516, RMSEA = 0.126, χ^2 (639) = 102,255), 5.6 (Alternative: CFI = 0.810, RMSEA = 0.078, χ^2 (649) = 40,464) and Figures 5.10 and 5.11 (reflective parameters for both models are given in Appendix D). Tables 5.5 and 5.6 also show that none of the structural paths between factors pass the test for weight invariance across all blogs (LM Test Probability > 0.05). It is readily

apparent that the alternative model fits significantly better than the naïve model (ΔCFI = 0.294, $\Delta\chi^2$ (10) = 61,791).

TABLE 5.5

Multigroup Fitted Naïve Model Parameter Values

	Parameter Value		Hypothesis Support		LM Test
	Standardized	**Unstandardized**	**Statistical**	**Practical**	**Probability**
H_1	-0.1178	-0.1455 (t = -18.88)	Yes	Yes	0.000
H_2	0.4594	0.6259 (t = 52.00)	Yes	Yes	0.000
H_3	0.1228	0.1257 (t = 22.14)	Yes	Yes	0.000
H_4	-0.0816	-0.0960 (t = -11.98)	Yes	Yes	0.000
H_5	-0.2170	-0.2266 (t = -33.74)	Yes	Yes	0.000
H_6	-0.0165*	-0.0171 (t = -2.55)*	Yes	No	0.000
H_7	0.3083	0.4727 (t = 26.92)	Yes	Yes	0.000

*Not practically significant

TABLE 5.6

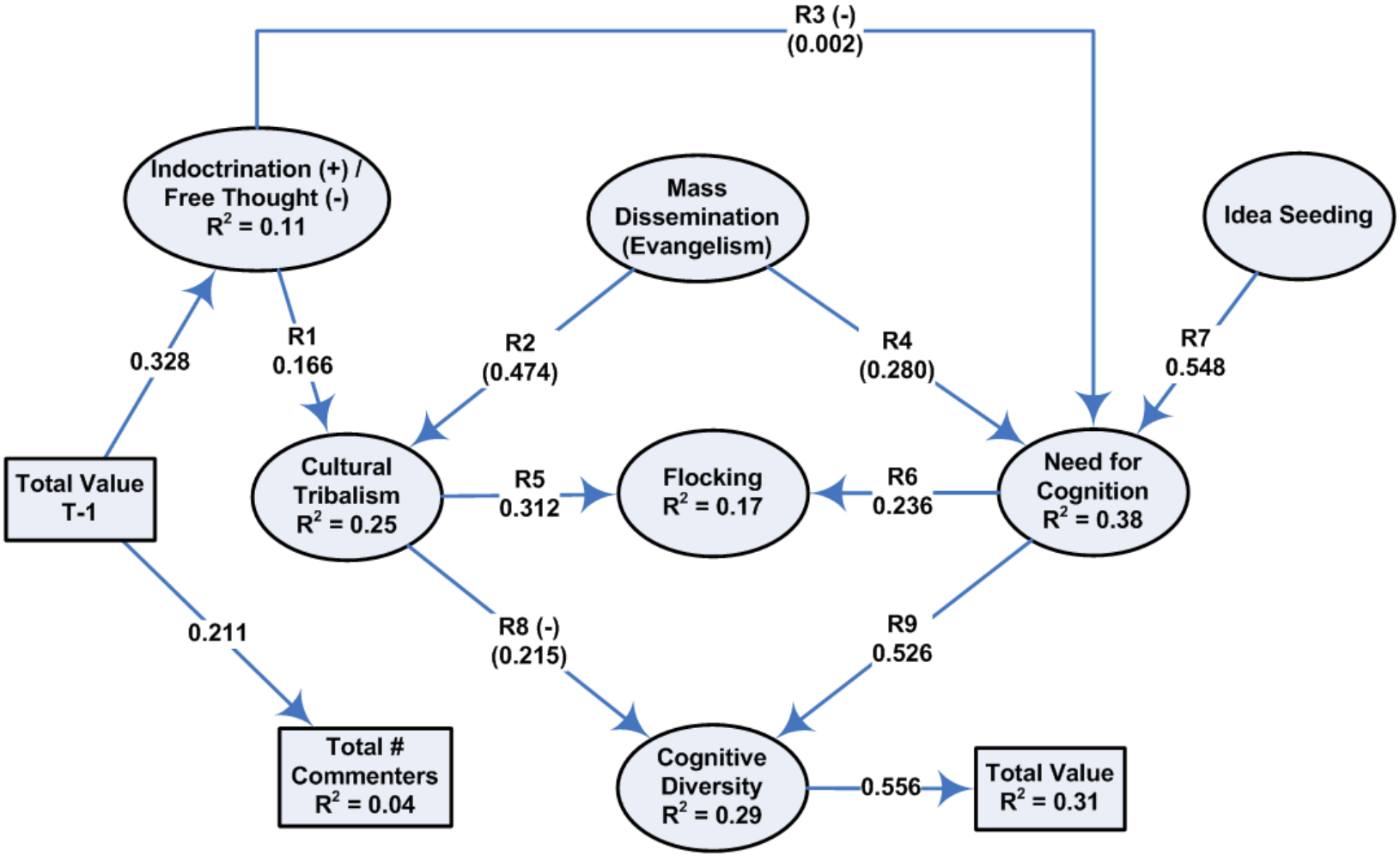
Multigroup Fitted Alternative Model Parameter Values

Path	Parameter Value		Relationship Support		LM Test
	Standardized	**Unstandardized**	**Statistical**	**Practical**	**Probability**
R_1	0.1662	0.1394 (t = 23.33)	Yes	Yes	0.000
R_2	-0.4743	-0.3936 (t = -54.75)	Yes	Yes	0.000
R_3	-0.0017*	-0.0012 (t = -0.26)*	No	No	0.000
R_4	-0.2796	-0.1936 (t = -38.30)	Yes	Yes	0.000
R_5	0.3124	0.3269 (t = 36.09)	Yes	Yes	0.000
R_6	0.2357	0.2956 (t = 28.63)	Yes	Yes	0.000
R_7	0.5478	0.3776 (t = 59.25)	Yes	Yes	0.000
R_8	-0.2150	-0.2731 (t = -32.22)	Yes	Yes	0.000
R_9	0.5262	0.8012 (t = 55.39)	Yes	Yes	0.000

*Neither practically nor statistically significant.

FIGURE 5.11

Fitted Alternative Model



Formal tests of mediation on the model in Figure 3.11 were conducted whenever one construct influenced another through an intermediary construct (e.g., mass dissemination affects cognitive diversity through cultural tribalism in Figure 3.11). This was done using Hardy and Bryman's (2004) procedure of bypassing each intermediary, one at a time, with a direct path and performing a χ^2-difference test. Cheung and Rensvold's (2002) ΔCFI test and Baron and Kenny's (1986) test of parameter values were also conducted. The results are summarized in Table 5.7; all mediation tests support the proposed alternative model.

TABLE 5.7

Alternative Model: Tests of Mediation

Relationship	CFI and X^2 Difference Test				Bypass Parameter
	CFI	X^2	Test of Significance ΔCFI	Test of Significance ΔX^2	
Baseline	0.810	χ^2 (649) = 40,464			
Indoctrination to Flocking	0.810	χ^2 (634) = 40,434	0.000	$\chi^2_d(15) = 30$, $\rho < .02$	0.002 (t = 0.136)
Mass Dissemination to Flocking	0.811	χ^2 (634) = 40,423	0.001	$\chi^2_d(15) = 41$, $\rho < .001$	0.263 (t = 0.8422)
Idea Seeding to Flocking	0.810	χ^2 (634) = 40,461	0.000	$\chi^2_d(15) = 3$, $\rho > .99$	0.143 (t = 1.376)
Indoctrination to Cognitive Diversity	0.815	χ^2 (634) = 39,414	0.005	$\chi^2_d(15) = 1050$, $\rho < .001$	-0.079 (t = -4.833)
Mass Dissemination to Cognitive Diversity	0.817	χ^2 (634) = 39,118	0.007	$\chi^2_d(15) = 1346$, $\rho < .001$	-1.811 (t = -4.198)
Idea Seeding to Cognitive Diversity	0.812	χ^2 (634) = 40,101	0.002	$\chi^2_d(15) = 363$, $\rho < .001$	-0.076 (t = -0.607)

As summarized in Table 5.5, H_1 to H_5 and H_7 are supported as indicated by both statistical significance and effect size. Statistical significance is not noteworthy in this study, as the large sample size is assured to raise statistical power enough to make all the estimated values statistically significant. The sizes of the estimated parameter values show they meet the level of practical significance arbitrarily set in Chapter IV in every case except H_6.

TABLE 5.8

Summary of Hypotheses and Results

	Discussion
H_1	*Cultural tribalism* is negatively related to *cognitive diversity*. Figure 5.10 and Table 5.5 show that H_1 is supported with both statistical and practical significance.
H_2	*Need-for-Cognition* is positively related to *cognitive diversity.* Figure 5.10 and Table 5.5 show that H_2 is supported with both statistical and practical significance.
H_3	*Idea seeding* is positively related to *cognitive diversity.* Figure 5.10 and Table 5.5 show that H_3 is supported with both statistical and practical significance.
H_4	*Flocking* behavior is negatively related to *cognitive diversity.* Figure 5.10 and Table 5.5 show that H_4 is supported with both statistical and practical significance.
H_5	*Mass dissemination* is negatively related to *cognitive diversity.* Figure 5.10 and Table 5.5 show that H_5 is supported with both statistical and practical significance.
H_6	*Indoctrination* is negatively related to *cognitive diversity.* Figure 5.10 and Table 5.5 show that although H_6 was estimated with the right valance (i.e., negative sign) and is supported by statistical significance, it fails the test of practical significance (i.e., parameter effect size > 0.05). However, Figure 5.11 and Table 5.6 indicate that when mediated by cultural tribalism, indoctrination has the hypothesized negative relationship with cultural tribalism.
H_7	*Reciprocity* is positively related to *cognitive diversity.* Figure 5.10 and Table 5.5 show that H_7 is supported with both statistical and practical significance.

Summary

In this chapter, the procedures detailed in Chapter IV were conducted on the entire data set. All but one hypothesis (H_6) in the naïve was supported; however, the essence of that hypothesis was supported in the alternative model. The alternative model of Figure 3.11 was shown to fit the data better than the naïve model of Figure 3.10. In the next chapter the implications of these findings are discussed. A summary of the hypotheses tested and their results is given in Table 5.8.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

Chapter V described the qualitative and quantitative results of this study. In this chapter, the research questions introduced in Chapter I are reviewed in the context of the findings of this study. Then the research implications of this study are discussed, along with suggestions for future research directions. Finally, the managerial implications of this study's findings are discussed and actionable recommendations are made.

Review of Research Questions

Mechanisms Explaining the Expression of Thought

In Chapter III, two models were introduced: (1) a naïve model that sought to explain the extent to which six social processes influenced cognitive diversity, the expression of diverse thought, and (2) an alternative model, that on the basis of early theoretical reasoning, sought to explain one way in which the six processes of the naïve model might interact. Since processes are dynamic, this study took data snapshots at blog entry intervals of summary measures (mean and standard deviation) of blog author and individual activity as well as collective activity under the assumption that processes can be measured indirectly by measuring their effects: regularity in the cognitive distance between, quantity and timing of thought expressions. In the following sections observations of the effects of the six social processes will be discussed.

Cultural Tribalism

In Chapter III, cultural tribalism was described as a process of individuals gravitating to blogs where they can express their views without compromise. Since individuals tend to temper their self-expression to be accepted by those around them, people who are highly sensitive to the dissonance resulting from compromised self-expression can never feel truly comfortable unless they are in the company of those whose views are similar. Thus cultural tribalism results in bringing like-minded people together. The same mechanism should also cause individuals to select a blog community where the blog author espouses views similar to their own.

It was expected, as shown in Figure 3.2, that contexts highly influenced by cultural tribalism would be characterized by relatively low levels of individual and collective thought diversity as like-minded people keep discussing the same ideas over and over. It was also expected, as shown in Figure 3.3 where Mean Commenter Longevity is an indicator of cultural tribalism, that such contexts would nurture long term associations manifested in a high proportion of serial commenters. It can be seen from Table 5.4 that these attributes were reliably ($r = 0.683$) associated together into the single latent variable construct denoted cultural tribalism.

It was expected that cultural tribalism would be negatively associated with cognitive diversity in general (H_1). This hypothesis was supported in terms of practical and statistical significance; however, its association with cognitive diversity was found to be weaker in the naïve model than in the alternative model (compare the relative magnitudes of H_1 and R_1) where it was conceived as being a grassroots mediator

between the exogenous influences of mass dissemination and indoctrination on cognitive diversity.

It was interesting to discover that cultural tribalism seemed to have an influence in all blogs even though blogs do differ in their relative influence (Figure 5.4 A). Scoble and Israel's (2006) concern that cultural tribalism is the central dynamic in the blogosphere seem unfounded as need-for-cognition seems to exert more than double the effect of cultural tribalism (compare the relative magnitudes of R_8 and R_9 in Figure 5.11). It does seem to be, however, the prime influence motivating flocking (compare R_5 with R_6). As discussed later, flocking seems to have a major presence in the blogosphere and it may well be that cultural tribalism's support of flocking is its underappreciated major consequence.

Need-for-Cognition

Need-for-cognition is widely discussed in the psychology literature as a motivation that underlies behavior, particularly the extent to which people analyze others' attempts at persuasion. In this study, the term is used to denote the counter-influence to cultural tribalism (recall the discussion surrounding Figure 3.2). The need-for-cognition motivation drives people to express a diverse array of thoughts as they discuss and respond to the joint stimuli of blog author entries and the comments of other blog readers.

As an influence that competes with cultural tribalism, need-for-cognition is conceptualized as being associated with relatively high levels of individual and collective thought diversity (see Figure 3.2) as well as a higher inflow of new voices into

the blog conversation (refer to Figure 3.3 where "% First Time Commenters" is an indicator of need-for-cognition). It can be seen from Table 5.4 that these indicators were reliably ($r = 0.635$) associated together into the single latent variable construct denoted need-for-cognition.

It was expected that need-for-cognition would be positively associated with cognitive diversity (H_2). This hypothesis was formally supported by both practical and statistical significance; indeed, its association with cognitive diversity was found to be strong in the naïve model (see Table 5.5), and even more so in the alternative model where it was conceived as being a grassroots mediator (see Figure 3.6 and related discussion) between the exogenous influences of mass dissemination and idea seeding on cognitive diversity.

Need-for-cognition is an influence present in all blogs; however, the strength of its presence varies more widely than that of cultural tribalism (compare Figure 5.4 A and B). Comparing Figure 5.5 B with Figure 5.9 B seems to indicate a strong correlation between the strong presence of idea seeding and need-for-cognition, a correlation confirmed by the strong weight on relationship R_7 in Figure 5.11.

Although no apriori hypothesis was advanced regarding the relationship, it was expected that need-for-cognition would be negatively associated with flocking (compare the negative polarity of H_2 with the positive polarity of H_4 in Figure 3.10). However, the study results show a practically and statistically significant positive relationship between them (R_6 in Figure 5.11). A plausible explanation for this observation follows. In Figure 3.8, flocking was distinguished from the other influences on cognitive diversity on the

basis of low boldness rather than constrained thought diversity. Since flocking can have a full spectrum of thought diversity, cultural tribalism can be associated with the lesser levels of flocking thought diversity (hereafter called "narrow" flocking) while need-for-cognition can be associated with the greater ("wide" flocking). These associations are based on the relationships between cultural tribalism, need-for-cognition and individual thought diversity depicted in Figure 3.2.

Thresholding and Idea Seeding

Idea seeders are conceptualized as one of three general categories of blog commenter. They were associated with the behavioral dynamic highlighted in Granovetter's (1978) *Threshold Models of Collective Behavior* because they, being high in boldness to express their thoughts, often act as an impetus that allows more timid readers to express similar thoughts, starting a cascade of thought expression as thresholds of inhibition are overcome.

Idea seeding is therefore associated with high levels of boldness and high levels of individual thought diversity as their boldness renders them uninhibited in expressing any thought that comes to mind. Unlike the evangelists of mass dissemination, idea seeders have no memeset they are devoted to propagating (recall Figure 3.7 and surrounding discussion); they are merely uninhibited in expressing whatever ideas come to mind. They are conceived as manifesting their boldness in a higher than normal cognitive distance between their expressed thought and that of others, and in a higher volume of comments (Figure 3.4). Since they are the first to broach a new topic of conversation, they start a higher than average number of new thematic clusters. It can be

seen from Table 5.4 that these indicators were very reliably (r = 0.852) associated together into the single latent variable construct denoted idea seeding.

It was expected that idea seeding would be positively associated with cognitive diversity (H_3). This hypothesis was formally supported by both practical and statistical significance (see Table 5.5); however its association with cognitive diversity was found to be weaker in the naïve model, than in the alternative model (compare the relative magnitudes of H_3 with R_7 in Figure 5.11) where it was conceived of as being an exogenous influence mediated by grassroots need-for-cognition (Figure 3.6) in its influence on cognitive diversity. Need-for-cognition among the mass of blog readers thus amplifies the influence of idea seeding on cognitive diversity as new thoughts expressed provoke the thinking of others. The necessity of this mediation was confirmed by a test whose result appears in Table 5.7 ("Idea Seeding to Cognitive Diversity") where bypassing need-for-cognition had an insignificant effect on model fit and a weak bypass parameter.

Idea seeding seems to have a presence in every blog but the extent of that presence varies widely (Figures 5.8 and 5.9 B). Although the model of Figure 3.11 portrays idea seeding as an exogenous influence, it is really a manifestation of an extreme side of the grassroots influences that affect blogs. As already stated, idea seeders have no agenda; they come to the blog as ordinary readers, have their thinking stirred and then express those thoughts. From Figure 5.11, it is apparent that their participation as an influence on the mob is solely responsible for any cognitive diversity in, and value resulting from, blogging.

Flocking

Flockers are conceptualized as one of the three general categories of blog commenter. They are distinguished by below average boldness in expressing their thoughts as they are motivated by Maslow's (1943) need-for-affiliation and a fear of expressing their thoughts in a manner that alienates others in whose group they wish to belong. This study seems to indicate that flockers comprise the largest segment of commenters. As a large population, it stands to reason that they would differ widely in their willingness to express their true thoughts and in how much they fear exclusion from the group of their choice. It can be seen from Table 5.4 that the indicators descriptive of boldness were associated together into the single latent variable construct denoted flocking, albeit with the lowest level of reliability ($r = 0.558$) among the constructs. It is surmised that the one-dimensional nature (i.e., a lack of boldness) of the construct was responsible for the lower reliability.

It was anticipated that flocking would be negatively associated with cognitive diversity (H_4). This hypothesis was formally supported by both practical and statistical significance (see Table 5.5); however, its association with cognitive diversity was found to be very weak in the naïve model. It was anticipated that its influence would be stronger because limiting the expression of diverse thought robs the collective of the full array of cognitive resources possessed by its members. However, the observed near neutral result can be explained by two perceptions: (1) flockers that express nearly their full scope of thought (wide flocking) counterbalance those that express little of that scope (narrow flocking), and (2) the system under study knows nothing of potential

contribution from its members, only their overt participation. Thus, there is no way for lost knowledge to be accounted for in the measure of cognitive diversity as when it was not expressed it never existed as far as the blogosphere is concerned.

Flocking's positive relationship with both cultural tribalism and need-for-cognition is interesting. As noted earlier, wide flockers, expected to be more diverse in their expression, are probably influenced primarily by need-for-cognition whereas narrow flockers are influenced by their attraction to groups based on cultural tribalism. Cultural tribalism and need-for-cognition thus mediate the effect of the independent influences of indoctrination, evangelism and idea seeding on flocking (see Table 5.7, where bypassing cultural tribalism's and need-for-cognition's mediation had no significant effect on model fit). As it was argued that idea seeding is but an extreme aspect of the effect of grassroots influences on blogging, so flocking is another aspect of grassroots influence.

Figures 5.8 and 5.9 (C) seem to suggest that flocking has a ubiquitous presence in the blogosphere that varies little between blogs. It is wrong then to conclude from flocking's neutral influence on cognitive diversity that it has no importance. Flocking is a manifestation of another form of value gained by blog participants: social belonging. The model of Figure 3.11 thus portrays flocking as a second dependent variable, that along with cognitive diversity create value for the community. Just as marketing theory maintains that consumers have diverse preferences that need to be met, so the grassroots evolution of the blogosphere has facilitated value creation for at least these two groups

of consumers, who to varying degrees substitute knowledge for social affiliation in their conscious and unconscious choices.

It has already been noted that the individual thought diversity distribution of Figure 5.5 (A) has an abrupt peak that occurs in all the blogs as shown in Figure 5.6 (B). The only location in Figure 5.5 (B) where this peak could occur would seem to be the flocking cluster. Figure 5.5 (A) shows a mainly normal distribution where people near the mean seem to have been drawn to the mean itself, thus eroding the top of the distribution and creating a dense pillar. A plausible explanation was suggested in Chapter V that this is the result of a *social attractor*, what Manzini (1994) describes as something that "orients the choices of a multiplicity of individuals" (p. 43). In this case it is suggested that people near the mean exert effort, consciously or unconsciously, to be "normal" in the extent to which their expressed thought diverges from their perception of the community mean. For this to have the observed effect, these people must be very accurate in their assessment of what "normal" divergence is and what the community norms are. Since all the blogs used in this study are English language, perceptions of "normal" may be a part of Western culture, learned adaptations that help individuals live in greater harmony in Western society. This control of expressed thought divergence is an integral aspect of how flocking was conceptualized in Chapter II, offering evidence that this study is truly measuring the flocking phenomenon.

Mass Dissemination

Mass dissemination was conceptualized as one of the two mechanisms used to spread memes or idea viruses. Mass dissemination may be covertly top-down or

hierarchically imposed by an entity with an agenda or it may be another grassroots influence where the charisma of the meme internally motivates people who embrace it to spread it. Agents of mass dissemination are denoted evangelists.

Mass dissemination is similar to idea seeding in the boldness of evangelists to express their thoughts regardless of context. Mass dissemination differs from idea seeding in the narrow diversity of thoughts expressed by evangelists. The mass dissemination construct is thus reflected by boldness (a history of being a prolific commenter expressing thoughts at high variance from the norm, initiating more thematic clusters) and low diversity of expressed thought. It can be seen from Table 5.4 that the indicators discussed above were reliably ($r = 0.690$) associated together into the single latent variable construct denoted mass dissemination.

It was expected that mass dissemination would be negatively associated with cognitive diversity (H_5). This hypothesis was formally supported by both practical and statistical significance in the naïve model (refer to Table 5.5). In the alternative model, mass dissemination's effect on cognitive diversity was modeled as mediated by cultural tribalism and need-for-cognition. On the cultural tribalism side, it was reasoned that the limited view point expressed by evangelists would have an effect similar to indoctrination in creating an intellectual home for people that embrace the meme. However, model estimation showed that mass dissemination is negatively associated with cultural tribalism (R_2 in Figure 5.11). Two reasons are offered for this unexpected finding: (1) a competing meme entrenched in a blog community by officially enacted indoctrination will actively oppose competing memes introduced by evangelists, and (2)

evangelists do not have the intellectual authority to set a blog's thematic tone. If evangelists are successful in drawing converts, they must migrate to other more agreeable blogs or start their own blogs thereby becoming indoctrinating blog authors.

On the need-for-cognition side, it was reasoned that blog readers could not readily identify a new evangelist, so evangelists would be mistaken for idea seeders and thereby initiate grassroots interest, increasing evidence of need-for-cognition in action. However, it may be that evangelists are too zealous and bore blog readers, causing them to withdraw from participation. Fortunately, in relative terms, evangelists seem less influential than true idea seeders and do not so affect the discussion as to dissuade it completely.

Indoctrination

Indoctrination was conceptualized as the second mechanism used to spread memes. Indoctrination is the only overtly top-down or hierarchically imposed influence proposed to be acting on blogs. It is seen as being the means of influence possessed by the blog author, exercised in the writing of blog entries that may set the thematic tone for the blog and establish the norm for thoughts expressed.

Indoctrination as a construct is similar to the individual thought diversity construct applied to commenters. It is indicated by the mean and standard deviation of the cognitive distance between blog entries with the assumption that the more similar the blog entries, the more intense the indoctrination. It is also indicated by the mean time between entries, under the assumption that greater indoctrination intensity will be accompanied by more frequent blog entries. As the indoctrination concept implies intent

to persuade, the absence of such intent must also be conceptualized. Since indoctrination, by definition, discourages critical thinking, its absence must embrace such thought and is therefore denoted “free thought” in the blog author thought diversity spectrum of Figure 3.5. It can be seen from Table 5.4 that the indicators discussed above were reliably ($r = 0.603$) associated together into the single latent variable construct denoted indoctrination / free thought.

It was expected that indoctrination would be negatively associated with cognitive diversity (H_6). This hypothesis was not supported by either practical or statistical significance in the naïve model (refer to Table 5.5). In the alternative model, its effect on cognitive diversity was seen as mediated by cultural tribalism and need-for-cognition, the indoctrination side of the spectrum was thought to be more associated with cultural tribalism (R_1 in Figure 5.11) while the free thought side was thought to be associated with need-for-cognition (R_3). Model estimation revealed that while indoctrination was associated with cultural tribalism (R_1), it had neither practical nor statistical effect on need-for-cognition (R_3). It is concluded that the cognitive diversity expressed by the blog author has only a one-sided impact on the cognitive diversity expressed by the blog community. It seems to be only able to create an intellectual home for people who agree with the blog author’s views and those who flock around them. The free thought side of blog entries may however be a signal of tolerance for diverse viewpoints that, while undetected by this study, are perceived by readers and result in elevated levels of idea seeding and wide flocking that, through need-for-cognition, expand cognitive diversity (R_9).

Reciprocity

Reciprocity was conceptualized as both a psychological and an economic influence. Blog readers who gain value from the blog conversation are envisioned as gradually accumulating a sense of obligation to contribute their own insights. Every individual has a different threshold of tolerance for the cognitive dissonance resulting from unreciprocated benefit. However, when that threshold is exceeded, individuals will write comments contributing their own knowledge to the collective, thereby repaying the community.

In the naïve model, reciprocity was conceived as being influenced by prior value created by the blog conversation and manifested by an increase in commenters. With only one indicator, the construct cannot be viewed as having much content validity; therefore, it was not included in the alternative model. However, prior value was included as an influence that might be associated with blog authors focusing on popular themes, and in so doing appear to be indoctrinators, and with an increase in the number of commenters.

It was posited that reciprocity will be positively associated with cognitive diversity (H_7), a hypothesis that was both statistically and practically supported in the naïve model (Table 5.5). In the alternative model, the influence of reciprocity's implied motivator, prior value received, was shown to be statistically and practically associated with indoctrination and the number of commenters (see Figure 5.11). However, even with these supportive results, it must be acknowledged that the hidden internal nature of reciprocity prevents a complete accounting of its influence on cognitive diversity, and

indeed its role in shaping the behavior observed in the blogosphere as a whole, to be gained though ethnography. As a result, it cannot be said that the full association between reciprocity and cognitive diversity was properly assessed here.

Processes that Expand Cognitive Diversity

The preceding discussion can be summarized in concluding that the key to achieving cognitive diversity is inspiring need-for-cognition in the blog community, and such need-for-cognition is primarily associated with the presence of idea seeding, that is, free thinking and uninhibitedly expressive commenters. Idea seeding and need-for-cognition is a grassroots stimulus-response pair that may indeed have a reciprocal relationship over time: idea seeding invokes need-for-cognition that generates thoughts that when expressed to the community results in the seeding of more ideas. Such a dynamic can be a proverbial perpetual motion machine of thought and cognitive diversity generation.

Processes that Limit Cognitive Diversity

It seems that cognitive diversity may be adversely affected by indoctrination, mass dissemination and cultural tribalism. As discussed above, there is evidence to suggest that mass dissemination reduces cognitive diversity by dampening need-for-cognition. The thematic repetition of evangelists may exhaust the interest of the community, causing readers to become uninspired and docile. Blog authors, whether motivated by an indoctrination agenda or by the desire to feed the community more of what it appears to enjoy, who limit the thematic content of their posts encourage the congregating of those who have a narrow interest in that theme, inspiring them to repeat

back the same ideas. Similar to the feedback mechanism posited to exist between idea seeding and need-for-cognition, a blog author who has nurtured a community characterized by cultural tribalism may, if monitoring the community, exacerbate the situation by adopting a populist stance in selecting themes for blog entries. It must be questioned though whether such communities have long term viability; since boredom is a universal human propensity, one would expect that eventually interest in communities dominated by cultural tribalism would wane unless they could attract a steady inflow of new adherents. The gradual decline and disappearance of cultural tribalism communities may also be a source of reduced cognitive diversity.

Most Important Processes

The resilience of the blogosphere seems strong since influences that seem to expand cognitive diversity are greater than those that seem to reduce it. It is interesting to note that indoctrination, the only top-down influence is exclusively associated with reducing cognitive diversity while grassroots emergent processes are behind the influences that expand it. Even what may be well-intentioned top-down attempts to give the community what it wants (prior value feedback to indoctrination), seem only to serve to damper the expression of diverse thought.

It must be remembered though, that this study narrowly focused on investigating influences that affect the expression of diverse thought under the assumption that the value that accrues from such cognitive diversity is of most interest to companies because they want a balanced and true perception of their image. Other types of value, seemingly of greater interest to other stakeholders, have been briefly mentioned: value from

satisfying need-for-esteem by exercising self-expression (through cultural tribalism), from satisfying need-for-affiliation (through flocking), and from satisfying need-for-cognition. It is apparent that the pursuit of these other forms of value differently affects cognitive diversity. Striving to satisfy need-for-cognition seems to positively influence cognitive diversity while the other pursuits seem to restrict it. Maslow (1943) leads to the conclusion that need-for-affiliation is a more basal and compelling need than the others. As a result, it must be acknowledged that by focusing its inquiry on processes that affect the filling of cognitive needs, this study may not have examined the major influences in the blogosphere.

Implications for Research in Marketing

In Chapter I, it was stated that this study sought to make substantive and methodological contributions to the field of marketing. In this section, the nature of these contributions is described.

Substantive Contributions

It must be acknowledged that this study was largely exploratory in nature. Its main contribution is an early-stage detection of patterns and phenomena that require further investigation in order for greater confidence to be placed in the findings. That caveat stated, this study does introduce some unique concepts to the marketing literature. One of those contributions is empirical evidence supporting the concept of flocking as a factor in group cognition. While this idea was proposed by Rosen (2002), it has not been hitherto empirically demonstrated in a socio-cognitive context. Additionally, this study introduces the concept of cultural tribalism into the marketing literature as distinct from

groupthink and group polarization. Cultural tribalism is a grassroots, long term phenomenon resulting from a desire for uncompromised self-expression, whereas, groupthink occurs because of an externally imposed requirement for consensus and group polarization occurs from people banding together to exercise greater influence.

This study adds incrementally to the literature investigating what motivates individuals to express their thoughts, as a form of word-of-mouth, in a public forum. It proposes and provides evidence for a reason why word-of-mouth might not reflect the true thoughts of participants: need-for-affiliation may cause some blog participants to compromise the expression of their thoughts and opinions to maintain social acceptance. It also introduces need-for-cognition as a word-of-mouth motivator. Furthermore, it evidences the presence of three distinct types of word-of-mouth participant (flockers, evangelists and idea seeders) and a model that shows how these types interact in the same context.

This study also contributes to the small body of literature that explores marketing phenomena from a complexity science perspective; that is, as primarily influenced by the grassroots actions of consumers rather than the top-down efforts of companies to drive consumers into desired behaviors. A common perception of companies is that a blog is merely another tool of persuasion and image construction. This study presents evidence to suggest that the readers, through the comments they write, have much greater power to provoke the thinking of other consumers than the blog author. The social process theory focus of this study also makes an unusual complexity science contribution to the literature as it investigates how known processes interact to create the complex behavior

of a dynamic context, whereas most marketing studies focus on single processes and simpler contexts.

This study also adds to the small body of memetic studies in the marketing literature. The concept of memetics, or ideas that spread like viruses with unusual ease, has intrigued researchers interested in advertising and persuasion. Unfortunately, it has been difficult to test models of memetic propagation because they involve unconscious cognitive processes. While this study does not provide definitive proof of the validity of the memetic concept or its theories of propagation, its fledgling detection of similarity in conversational themes and their spread through a group of people known to be in contact offers support for the proposed mass dissemination and indoctrination concepts.

One of the most interesting contributions of this study is the evidence provided for the finding of a social attractor in the distribution of individual thought diversity of Figure 5.5 (A). Like memetics, the concept of social attractors is difficult to validate because it assumes the operation of social processes that are difficult to empirically detect. This study documented a strange pattern in what was expected to be a normal distribution. The social attractor model is a plausible explanation for the pattern observed, that while inconclusive, certainly provides grounds for further investigation.

Watts and Dodds (2007) challenged the “two-step flow” model of communication advanced by Katz and Lazarsfeld (1955) and widely embraced by both the marketing (e.g., Van den Bulte and Joshi 2007; Vernette 2004) and communications (e.g., Weimann 1994) disciplines. The “two-step flow” model describes information flowing from mass media sources, through a small number of influential “opinion

leaders," to the public at large. Watts and Dodds (2007) demonstrated through computer simulations that such information flows depend, not on a few influential individuals (the classic "influentials hypothesis"), but on developing a critical mass of "easily influenced individuals" (p. 441) to begin a cascade of information flow with "easily influenced people influencing other easily influenced people" (p. 447). Watts' and Dodds' computational model is persuasive but offered without empirical support.

This study provides some empirical support for the Watts and Dodds (2007) model, in that it demonstrates the superiority of grassroots mechanisms in propagating ideas. Both studies draw on Granovetter's (1978) concept of cascades in collective behavior overwhelming individual behavioral thresholds to support their theories. It could be argued that flockers, the seeming majority, are a class of easily influenced people, as their spreading of ideas is less the result of reasoned conviction than the need for social acceptance. However, this study also demonstrates the influence of idea seeders, individuals whose novel ideas frequently start cascades of thinking. Idea seeders should be considered a hitherto unrecognized class of influential. They act, not as regular intermediaries between established idea creators and the masses, but as decentralized and intermittent introducers of diverse ideas who influence the thoughts of others. Some idea seeders are observed to be more reliable in their influence than others. These idea seeders may be the classically conceptualized opinion leaders identified by prior research (e.g., Katz and Lazarsfeld 1955). This phenomenon seems to warrant an expansion of our conceptualization of what the influentials hypothesis refers to.

This study's empirical findings seem to support the conceptualization of a complex environment that allows the simultaneous operation of the influentials and critical mass hypotheses. With regard to the critical mass hypothesis, this study's cultural tribalists may be a first wave of early converts to an ideology, while flockers act as a subsequent, easily influenced early majority that create the critical mass necessary for mass dissemination. Alternatively, perhaps diffusion begins with the attention of the need-for-cognition curious, who give a new idea enough visibility for it to attract the attention of easily influenced flockers who then create a critical mass. It seems more reasonable, given the history of research that supports the influentials hypothesis, to consider that both dynamics coexist in a complex environment, rather than to consider one model supplanting the other.

Methodological Contributions

As detailed in Chapter II, there is a long legacy of research in marketing that has investigated and attempted to model cognition. This study adds to the marketing literature a novel integrative application of textual analysis, network theory, multidimensional scaling and cluster analysis that creates cognitive maps with greater theoretical transparency than competing methodologies employing neural networks.

The methodology used by this study has garnered interesting, albeit embryonic, insights that argue its use in other circumstances where text is used to infer cognition. The methodology is a good quantitative complement to the qualitative analysis of text as each cognitive object in multidimensional space can be linked back to its underlying word content, matching cluster patterns to identifiable themes and even theme transitions

to changes in a group's cognitive structure (how knowledge objects were connected together).

One contemporary theme in the communications literature is discussing the measure of persuasion effectiveness. Particularly relevant to this study is measuring attitude toward a message. Often the literature discusses whether measuring perceived effectiveness (PE) before launching a costly campaign, is indicative of the actual effectiveness the campaign would have (e.g., Dillard, Weber and Vail 2007). Perceived effectiveness is generally assessed by surveying a group of message recipients and asking them whether they find the message persuasive (Dillard et al. 2007, p. 617). Seeding a candidate message in the form of a reader comment to one or more blogs where company products are discussed, and using a methodology similar to that employed in this study to monitor how its content is embraced by subsequent commenters, may prove to be a more reliable means of testing effectiveness than surveying PE.

Peterson (2005) posited "with rare exceptions, answers to questions asked in consumer behavior research studies … tend to be constructed rather than … directly retrieved from memory" (p. 352). He called for researchers to augment self-reports with alternative sources of data, including behavioral observation. Monitoring customer generated media (e.g., blogs and forums) may gain equal footing with (or even displace) the collection of survey-based respondent driven data as an observational, and thus more objective, data collection method. This study may be a step in the direction of giving the research community confidence to do that.

This study also contributes to the literature about research methods by being one of the few large-scale studies using multiple populations. It is thus an example of an approach that could be taken by those who seek to do a similar study.

Directions for Future Research

Recognizing synonyms in Centering Resonance Analysis' calculation of influence (repeated words that tend to connect ideas within a body of text) and resonance (comparing bodies of text on the basis of the same words used with similar influence) would be a powerful enhancement to the basic algorithm. However, that enhancement brings attendant complications as many words in the English language have different meanings even though they are spelled the same. Work on this enhancement is deferred to future research.

There seems to be some potential for further insight by examining and analyzing the data from a time series perspective. To this end the following research questions merit investigation:

1. Do blog participants grow in the diversity of their thought expression?
2. How do cognitive diversity and the principle constructs (e.g., indoctrination, mass dissemination and flocking) develop over time, starting with the blog entry and updated as each comment is added?
3. Is there trending in the major constructs as more blog entries are added to a blog?
4. How do blog communities evolve from inception to present with respect to membership size, and changes in relative dominance of cultural tribalism and need-for-cognition as motivating influences?

Progress in these areas might also reveal latent variable construct indicators on additional dimensions that would enhance the reliability and content validity of this study.

Managerial Implications

Statistically Improbable Words

In Chapter I, companies were portrayed as better able to perform image management with feedback from the blogosphere. Centering Resonance Analysis (CRA) has been demonstrated to be a means of identifying the most important keywords associated with a corporate image. Furthermore, the word network produced by CRA preserves the context in which the most influential words were used. This capability could be enhanced by looking for statistically improbable words (SIWs) among the most influential words demarcated by CRA. These are words that occur much more often than other words associated with a company, or a company and all its competitors in a market. The SIW concept is conceptually derived from Amazon.com's (2008) *Statistically Improbable Phrases (SIPs)*, the most distinctive phrases inside a particular book, relative to phrases found in all books. Amazon uses these phrases to hint at unique themes inside a book that might better separate it from others in the same genre. Similarly, if this concept is applied to CRA influential words, it may better identify the most unique words descriptive of corporate image. A company may also be able to see whether key words from marketing communications are making an impact with the public.

Does Diversity of Thought in the Blogosphere Echo that of the Market?

This study began by asking a very general research question and then proceeded to answer more specific questions that would hopefully lead to eventually answering the original question: How do companies know that the diversity of thought they see in the blogosphere reflects that present in the overall market? Preliminary evidence has been presented to suggest that a majority of blog commenters (i.e., flockers) compromise the expression of their true thoughts to gain social acceptance. This study proposes a means of identifying the most extreme of such people (flockers near and below the mean of individual thought diversity). Companies who monitor blogs for image insights may be justified in assigning less weight to image indications gained from extreme flockers as these commenters may not be expressing their true thoughts, merely thoughts that are intended to maintain social acceptance.

It remains uncertain whether the population of blog commenters is a representative sample of the whole market, as this study reveals no profile of cognitive diversity in the entire market. The only way a conclusion could be made is if it is assumed that all the measures used in this study (e.g., individual and collective thought diversity) conform to a certain distribution (e.g., normal) in the overall market. For example, in Figure 5.5 (A) individual thought diversity is depicted as generally normally distributed, but with a sharp discontinuity at the mean. It may be assumed that this attribute is normally distributed in the whole market and thereby conclude that the population of blog commenters does not represent the whole market (because of the sharp discontinuity). However, it may be that the whole market of all consumers exhibits

a similar discontinuity, supporting the conclusion that the population of blog commenters accurately represents the whole market. Theoretical argument cannot resolve this uncertainty as reasons can be offered for both assumptions. It is recommended that, until an alternative and independent means of corroborating this study's results is performed, generalizations remain tentative. Image insights gleaned from blog monitoring should be examined, but recognized as possessing some degree of uncertainty yet to be fully understood.

Evidence was presented to suggest that blogs whose content is characterized by a single theme attract a readership with similarly narrow interests. As Scoble and Israel (2006) cautioned, such cultural tribalism cannot be a source of balanced image insights. The nurturing of such communities will be exacerbated if the blog author recycles blog entry themes because they were popular in the past. If a diverse-thinking blog community is desired, then it seems that monitoring the proportion of idea seeders is important. Collectively, the blog populations studied here had an average of 15% idea seeders. If a blog has substantially less than that, cultural tribalism has probably become the dominant motivation among blog participants. It seems that idea seeders are nurtured if blog authors introduce diverse themes into their conversation starting entries.

Blogs as a Tool of Influence

This study's contribution to the reputation-building literature only addresses the question of whether blogs might be used as a persuasive tool to alter reputation. This study identifies a class of blog participant, idea seeders, whose diverse thought expression seems to reliably spark interaction among a wide variety of blog participants.

It is thought that this interaction is primarily of a sensemaking nature, participants trying to determine where novel thoughts fit into their existing cognitive structures. However, blog participants who generally express ideas on the same themes, denoted evangelists, whether or not they have an intent to persuade, seem only to provoke conversation among individuals with a history of interest in those themes. It seems then, that if idea seeders begin a practice of repeating ideas on the same theme, then their influence as conversation starters would diminish. This study observes this behavior pattern in the community's reaction to the arrival of new evangelists. The blog community cannot tell their true nature in the beginning and treats them as idea seeders. However, as their true nature is gradually revealed, they lose their ability to provoke diverse thought expression and thereby lose much of their influence in the community.

This study suggests two complementary interventions a company could make in the blog conversation to ensure it gets a more accurate perception of construed image, its reputation among consumers:

a) *Tactical manipulation*. This action is a variation on Dellarocas' (2006) *strategic manipulation* where firms post anonymous positive messages about their products to opinion forums. However, in this case the intent is not to persuade readers of the value of a product to induce purchase, but to give silent, less bold readers the courage to reveal their similar thoughts and thus start a cascade that reveals a hidden sentiment segment. To execute this tactic a company introduces a group of covert and seemingly independent evangelists. Each evangelist would initially be received as an idea seeder and have a limited opportunity to have the

community to consider their ideas. This limited period is the evangelist's primary opportunity to embolden silent blog community members into revealing their perspective. If each evangelist expresses thoughts narrowly on a common theme, the illusion of natural grassroots action may be lost. However, if each evangelist frames their message from a slightly different perspective, expressing understated interest or tentative support, they may be able to create the illusion of sufficient widespread agreement, Cialdini's (1984) *social proof* or *band-wagon effect*, to draw out enough other community members that the critical mass needed for a Watts and Dodds (2007) style information cascade is reached. It should be noted though that this tactic may primarily influence flockers, those motivated by a need for social belonging, and as a result, it may be superficial in its indications.

b) *Blog author encouragement.* This study suggests the blog author is more influential with cultural tribalists (i.e., the already persuaded) than those motivated by need-for-cognition. From Figure 5.8 it seems that some blog authors are able to act on the fine line between maintaining the reputational beliefs of the tribe and encouraging a free and open dialogue. An atmosphere welcoming free thought should attract and maintain a diverse-thinking population of readers. As long as this diversity is present, occasional tactical acts of manipulation may be effective in revealing new segments of product enthusiasts.

Godes et al. (2005) proposed a framework with four non-mutually exclusive ways that a company could be involved with online social interactions between

consumers: (1) Observer: a company uses their observation of interactions among consumers as a learning opportunity; (2) Moderator: a company encourages consumers to interact by establishing a blog or other form of online community; (3) Mediator: a company takes control of the information contained in online word-of-mouth and disseminates it as it desires (e.g., removing blog comments it sees as undesirable); and (4) Participant: a company enters into social interactions with consumers by acting as an anonymous consumer. Plainly, the latter two behaviors have the potential to manipulate consumer perceptions. Dellarocas (2006) analyzed the participant scenario from the perspective of it being a competitive strategy among firms. He concluded that when firms compete on the basis of manipulating the consumer, it is a costly distraction from making more substantive market offering improvements. Mayzlin (2006) also analyzed the participant scenario from the same perspective and found that "firms spend more resources on promoting inferior products," (p. 155) because "they don't get the free publicity from legitimate chatters." (p. 161)

Balasubramanian and Mahajan (2001) note that one of the primary advantages of community is that it allows individuals to build a reputation that streamlines exchange because trustworthiness does not have to be ascertained prior to every transaction. Bart et al. (2005) note that community features [e.g., weblogs] are a factor driving trust in websites. They note that "shared consciousness and a sense of moral responsibility and affinity enhance the consumer's level of trust" (p. 136) and may make consumers more confident in accepting reputational information from online communities. Therefore, any

corporate participatory behavior that undermines trust robs blogging of some of its ability to create positive outcomes for both firms and consumers.

This study's findings suggest that, since they tire of evangelists, blog participants have a resistance to sustained corporate manipulation. They may derive this resistance by matching a commenter's user name with memory of the content of their past contributions. Alternatively, they may simply remember the ideas heard without reference or association to who said them. The latter alternative is similar to the SIR epidemiology model (Kermack and McKendrick 1927) where individuals are in one of three states: susceptible, infected, and recovered. Variants of this model have been applied to ideas and influence in the marketing literature (e.g., Van den Bulte and Joshi 2007). If blog participants become resistant to repeated ideas without reference to source, sustained corporate manipulation, even in the guise of different evangelists, is likely to prove futile.

Alternative Sources of Value

Companies should be mindful that people participate in blog communities for reasons that might undermine thought diversity (e.g., need-for-affiliation). There may be value in servicing those needs in a corporate blog even though commenters might not be providing valuable insights into image or directions for product development. Such communities may enhance the loyalty of customers and have a positive effect on sales.

Managerial Implications of Research Findings Linked to Research Questions

In the *Review of Research Questions* section at the start of this chapter, the research questions were reviewed in light of the study's findings. In this section, the managerial insights of the research are linked to the research questions of the study:

1. *What mechanisms explain the expression of thoughts by individuals in a blog?* Figure 5.10, the fitted naïve model, shows that reciprocity and need-for-cognition are the most important direct influences on cognitive diversity. Figure 5.11, the fitted alternative model, shows that idea seeding is the primary influence on need-for-cognition. Therefore, reciprocity and idea seeding seem to be most relevant in determining whether or not comments are made, and if so, when they are made. This finding presents a challenge as blog authors may feel pressured to induce reciprocity by consistently posting content known to be of high value to their community and thereby fall into indoctrination behavior. Blog authors should be cognizant of what themes interest readers, but should be more focused on expanding the variety of relevant themes discussed in the blog, creating an atmosphere of free thought to encourage idea seeders. Let the value of spontaneous, albeit happenstance, contributions of a large population of idea seeders and wide flockers be the engine that induces reciprocity.
2. *What processes expand cognitive diversity, the expression of diverse thought?* The influences on cognitive diversity with the largest positive effect sizes were reciprocity and need-for-cognition (Figure 5.10), while need-for-cognition was primarily influenced by idea seeding (Figure 5.11). As discussed in response to

question 1 above, the most important influences in blogs are also the most positive in supporting diverse thought. Reciprocity and need-for-cognition seem to work together to create self-sustaining conversations that attract wide interest and diverse contributions.

3. *What processes limit cognitive diversity?* Mass dissemination (evangelists repeating the same ideas widely to influence those with the lowest thresholds of resistance), flocking (particularly, narrow flocking where blog participants seek social acceptance in groups characterized by cultural tribalism) and cultural tribalism (homogeneous groups formed to satisfy members' desire for self-expression in a non-judgmental, albeit echo chamber, context) are the primary influences reducing cognitive diversity. The implications of the findings relevant to these influences are discussed next:

    a) Evangelists are often received as idea seeders when they first start speaking their message; however they are quickly disregarded as they repeat the same ideas. Evangelists may initially expand cognitive diversity before they are recognized. This presents an opportunity for companies trying to test for the presence of readers who agree with the evangelized message but are too timid to express it. It is anticipated that converts to an evangelized message will leave a blog to found their own, or migrate to a blog that embraces those ideas.

    b) Although narrow flockers, as the most timid and least independent content contributors, are unlikely to introduce ideas leading to breakthrough products,

they may be a very profitable consumer segment if they see their group acceptance as tied to imitating the product purchases of others. Their profitability may derive from a magnification of sales effort, where sales effort expended on influencing more discriminating consumers may trigger multiple imitative sales. Narrow flockers are also unlikely to be demanding consumers as long as their need for affiliation is met.

c) Cultural tribalists are the diehard loyalists whose complete satisfaction should be least susceptible to fickle change. However, companies should not assume that there is no competition for the loyalty of these community members. Companies should use the opportunity the passionate participation of the tribe will provide to ensure their basic needs are not overlooked.

4. *What is the relative importance of these processes?* The most important influence in the blogosphere is that idea seeding invokes need-for-cognition dynamic discussed above. The secondary influences of indoctrination enabling cultural tribalism, free thought supporting need-for-cognition, and the imitation of both groups influencing flocking are also very important. Figure 5.4 shows that in all the blogs used in this study, cultural tribalism and need-for-cognition are able to co-exist to varied degrees even though they are conceptualized as opposing influences. It is proposed that the key to this balance is the blog author recognizing the ideas for which the tribe demands active support (through indoctrination), and then create an atmosphere of free reign for as wide as possible an array of other ideas that do not contradict those actively espoused. It is apparent from Figures 5.4 and 5.8 that not every blog is

equally successful in achieving a balance that assures high cognitive diversity. However, companies whose own blogs support little diversity can always monitor other blogs frequented by their target markets to round their perspective of image and market satisfaction trends.

Public Policy Implications

Although, public policy issues are not the focus of this study, it is readily apparent that measuring the diversity of thought in a population around a public policy issue would be of great interest to legislators and pollsters. All the managerial implications of monitoring image could apply to any public policy issue, as could this study's speculations of how tactical manipulation could be used to reveal points of view held by non-expressive consumer segments. Three of the blogs used in this study are host to political discussions and it seems that the same kinds of participants (i.e., idea seeders, evangelists and flockers) explain the expression of thought in political blogs as those that are more commercial. A full exploration of this issue is deferred to future research.

Limitations

While the findings of this study have important implications for marketing practice and research in marketing, certain limitations must be borne in mind. For instance, the concepts that the study tries to measure are inherently complicated. This study takes an admittedly simplistic, albeit good-faith, approach to representing them as constructs. If the explained variation (R^2) estimates were corrected to reflect this

uncertain content validity, their values would certainly be much lower. Not only does content validity need to be assessed and raised, but the reliability of most of the constructs needs to be increased as well.

Also, the blogs in the dataset are all from the leftmost side of Figure 2.3, the most popular blogs. The profile of blog populations presented in this study may not reflect that of less popular blogs, the blogs that collectively comprise the majority of content in the blogosphere.

This study began by asking: How do companies know that the diversity of thought they see in the blogosphere reflects that present in the overall market? It might be argued that the scattering of points in Figures 5.1 (B) and 5.5 (B) are sufficiently symmetric and well-distributed throughout their respective spaces to show that the full spectrum of thought has been captured. However, these scatter plots collectively only cover a three-dimensional space (boldness, individual and collective thought diversity). If more dimensions were added, Bellman's (1961) *curse of dimensionality* would come into play due to the exponential increase in volume. As volume increases, the distance between objects located in the space becomes greater and the population density becomes sparser. Thus, the number of data points required to adequately sample the space becomes larger. The cognitive space of the overall market is sure to be an unmanageably high-dimensional space. However, with further research it may be possible to identify its principal dimensions, define a manageable cognitive space and map what is sure to be an even larger sample of blog entries and comments to it, and

thereby claim to have adequately approximated the extent to which the blogosphere represents the diversity of thought in the market.

Summary

This chapter discussed the major findings of how the six social processes investigated in this study were observed to be associated with cognitive diversity, the expression of diverse thought in blogs. Idea seeding, or the comments of the boldest and most diverse thinkers, was observed to have the most powerful influence on overall cognitive diversity because of its ability to provoke the thinking of others, inspiring them to become commenters too.

The most surprising findings were: (1) the intellectual content of blog authors' posts is not as influential as the intellectual contributions of the commenters, however the post's affective content is thought to be very important in creating a collaborative atmosphere for commenters; (2) flockers whose thought diversity is near the mean, seem able to estimate the average level of flocker thought diversity. This estimate appears to be a social attractor, a cognitive magnet for flockers who aspire to be average in the diversity of their expressed thought; and (3) idea seeders may be a new class of influential consumers, distinct from classical opinion leaders in that, while they exert power as a class, they may hold little influence as individuals. However, the classic opinion leaders may be those idea seeders who have more consistent influence.

This chapter also discussed the contributions of this study to the marketing literature. This study explores a new subject and introduces a new methodology to the field of marketing. However, being largely exploratory in nature, it tended to find

evidence for fruitful paths of further inquiry rather than definitive conclusions. Finally, this chapter made recommendations to marketing managers who are thinking about using blogs to gain image insights and as a tool for influencing image perceptions. It was recommended that caution be exercised in using insights gained from blogs because participants often compromise the expression of their thoughts to maintain social acceptance. It was also recommended that blogs not be used as a means of persuasion as grassroots influences have the most effect, and detected persuasion attempts may undermine trust.

REFERENCES

Aaker, Jennifer L. (1997). "Dimensions of Brand Personality," *Journal of Marketing Research*, 34 (3), 347-356.

Abdi, Herve *(*2007*),* "Centroid, Center of Gravity, Center of Mass, Barycenter," in *Encyclopedia of Measurement and Statistics,* N.J. Salkind ed., Thousand Oaks, CA: Sage.

Ajzen, Icek (2001), "Nature and Operation of Attitudes," *Annual Review of Psychology*, 52, 27-58.

Amazon.com (2008), *Statistically Improbable Phrases*, (accessed April 12, 2008), [available at http://www.amazon.com/gp/search-inside/sipshelp.html]

Anderson, Chris (2004), "The Long Tail," *Wired*, 12.10, 15-20.

Arnould**,** Eric J., and Craig J. Thompson (2005). "Consumer Culture Theory (CCT): Twenty Years of Research," *Journal of Consumer Research*, 31 (4), 868-882.

Azouley, Audrey and Jean-Noel Kapferer (2003). "Do Brand Personality Scales Really Measure Brand Personality," *Brand Management*, 11 (2), 143-155.

Bagozzi, Richard P., Youjae Yi and Lynn W. Phillips (1991), "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, 36 (3), 421-458.

Balasubramanian, Sridhar and Vijay Mahajan (2001), "The Economic Leverage of the Virtual Community," *International Journal of Electronic Commerce*, 5 (3), 103-138.

Barnett, George A. and Joseph Woelfel (1979). "On the Dimensionality of Psychological Processes," *Quality and Quantity*, 13, 215-232.

Baron, Reuben M. and David A. Kenny (1986), "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, 51 (6), 1173-1182.

Bart, Yakov, Venkatesh Shankar, Fareena Sultan and Glen L. Urban (2005), "Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study," *Journal of Marketing*, 69 (4), 133–152.

Belk, Russell W. (1988). "Possessions and the Extended Self," *Journal of Consumer Research*, 21 (1), 55-70.

Bellman, Richard (1961), *Adaptive Control Processes: A Guided Tour*, Princeton, NJ: Princeton University Press.

Bentler, Peter M. (2005), *EQS 6 Structural Equations Program Manual*, Encino, CA: Multivariate Software.

Bianchini, Monica, Marco Gori, and Franco Scarselli (2005), "Inside PageRank," *ACM Transactions on Internet Technology*, 5 (1), 92–128.

Blackmore, Susan (2001), "Evolution and Memes: The Human Brain as a Selective Imitation Device," *Cybernetics and Systems*, 32, 225-255.

Bousfield, Weston A. and Barry H. Cohen (1955). "The Occurrence of Clustering in the Recall of Randomly Arranged Words of Different Frequencies of Usage," *Journal of General Psychology*, 30, 149-165.

----, H. William Sedgewick and Barry H. Cohen (1954), "Certain Temporal Characteristics of the Recall of Verbal Associates," *American Journal of Psychology*, 67, 111-118.

Brown, Tom J. and Peter A. Dacin (1997). "The Company and the Product: Corporate Associations and Consumer Product Responses," *Journal of Marketing*, 61 (January), 68-84.

----, Peter A. Dacin, Michael G. Pratt and David A. Whetten (2006). "Identity, Intended Image, Construed Image and Reputation: An Interdisciplinary Framework and Suggested Terminology," *Journal of the Academy of Marketing Science*, 34 (2), 99-106.

Brown, Steven, Robert V. Kozinets and John F. Sherry Jr. (2003). "Teaching Old Brands New Tricks: Retro Branding and the Revival of Brand Meaning," *Journal of Marketing*, 67 (July), 19-33.

Burnstein, Eugene and Keith Sentis (1981), "Attitude Polarization in Groups," in *Cognitive Responses in Persuasion*, Richard E. Petty, Thomas M. Ostrom and Timothy C. Brock, eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 197-216.

Burrell, Gibson and Gareth Morgan (1979), *Sociological Paradigms and Organizational Analysis*, London: Heinemann.

Cambel, Ali Bulent (1993), *Applied Chaos Theory: A Paradigm for Complexity*, Boston MA: Academic Press.

Campbell, D.T. and D.W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix," *Psychological Bulletin*, 56, 81-105.

---- and Julian C. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*, New York: Houghton Mifflin Company.

Cederman, Lars-Erik (2005), "Computational Models of Social Forms: Advancing Generative Process Theory," *American Journal of Sociology*, 110 (4), 864-893.

Chandrashekaran, Murali, Beth A. Walker, James C. Ward and Peter H. Reingen (1996), "Modeling Individual Preference Evolution and Choice in a Dynamic Group Setting," *Journal of Marketing Research*, 33 (May), 211-223.

Chiu, Tom, DongPing Fang, John Chen, Yao Wang and Christopher Jeris (2001), "A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment," *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263.

Cheung, Gordon W. and Roger B. Rensvold (2002), "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling: A Multidisciplinary Journal*, 9 (2), 233-255.

Cialdini, Robert B. (1984), *Influence: The Psychology of Persuasion*, New York: William Morrow and Company, Inc.

Cohen, Jack and Ian Stewart (1994), *The Collapse of Chaos: Discovering Simplicity in a Complex World*, London: Penguin Books.

Corman, Steven and Kevin J. Dooley (2008), *Crawdad Technologies, LLC,* (accessed May 11, 2008), [available at http://www.crawdadtech.com]

----, Timothy Kuhn, Robert D. McPhee and Kevin J. Dooley (2002). "Studying Complex Discursive Systems – Centering Resonance Analysis of Communication," *Human Communication Research*, 28 (2), 157-206.

Cronbach, Lee J. (1951), "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16, 297-334.

Dawkins, Richard (1976), *The Selfish Gene*, Oxford: Oxford University Press.

---- (2006), *The God Delusion*, New York: Houghton Mifflin.

Dellarocas, Chrysanthos (2006), "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science*, 52 (10), 1577-1593.

Dholakia, Utpal M., Richard P. Bagozzi and Lisa Klein Pearo (2004), "A Social Influence Model of Consumer Participation in Network- and Small-Group-Based Virtual Communities," *International Journal of Research in Marketing*, 21 (3), 241-263.

Dowling, Grahame (2006), "How Good Corporate Reputations Create Corporate Value," *Corporate Reputation Review*, 9 (2), 134-143.

Dwyer, Paul (2007), "Measuring the Value of Electronic Word-of-Mouth and Its Impact in Consumer Communities," *Journal of Interactive Marketing*, 21 (2), 63-79.

Dwyer, Tim (2004), *WilmaScope*, (accessed May 11, 2008), [available at http://wilma.sourceforge.net]

Eisenberg, Eric M. and Steven R. Phillips (1990), "What is Organizational Miscommunication?" in John Wiemann, Nikolas Coupland and Howard Giles (eds.) *Handbook of Miscommunication and Problematic Talk*, Oxford: Multilingual Matters, 85-103.

Fornell, Claes and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39-50.

Fournier, Susan (1998), "Consumers and Their Brands: Developing Relationship Theory in Consumer Research," *Journal of Consumer Research*, 24 (4), 343-373.

Freling, Traci H. and Lukas P. Forbes (2005). "An Examination of Brand Personality through Methodological Triangulation," *Brand Management*, 13 (2), 148-162.

Friendly, Michael L. (1979), "In Search of the M-gram: The Structure of Organization in Free Recall," *Cognitive Psychology*, 9, 188-249.

Gintis, Herbert (2004), "Modeling Cooperation among Self-Interested Agents: A Critique," *Journal of Socio-Economics*, 33 (6), 697-717.

Gioia, Dennis A., Maijken Schultz and Kevin G. Corley (2000). "Organizational Identity, Image, and Adaptive Instability," *Academy of Management Review*, 25 (January), 63-81.

Godes, David, Dina Mayzlin, Yubo Chen, Sanjiv Das, Chrysanthos Dellarocas, Bruce Pfeiffer, Barak Libai, Subrata Sen, Mengze Shi and Peeter Verlegh (2005), "The Firm's Management of Social Interactions," *Marketing Letters*, 16 (3/4), 415-428.

Govers, Pascalle C. M. and Jan P. L. Schoormans (2005), "Product Personality and Its Influence on Consumer Preference," *Journal of Consumer Marketing*, 22 (4), 189-197.

Granovetter, Mark (1978), "Threshold Models of Collective Behavior," *The American Journal of Sociology*, 83 (6), 1420-1443.

Grayson, Kent (1998). "The Icons of Consumer Research: Using Signs to Represent Consumers' Reality," in Stern, Barbara (Ed.) *Representing Consumers: Voices, Views and Visions*. London: Routledge, 27-43.

Green, Paul E., Yoram Wind and Henry J. Claycamp (1975), "Brand-Features Congruence Mapping," *Journal of Marketing Research*, 12 (August), 306-313.

Golder, Scott A. and Bernardo A. Huberman (2006), "Usage Patterns of Collaborative Tagging Systems," *Journal of Information Science*, 32 (2), 198-208.

Goldstein, Jeffery (1999), "Emergence as a Construct: History and Issues," *Emergence: Complexity and Organization*, 1, 49-72.

Hair, Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson and Ronald L. Tatham (2005), *Multivariate Data Analysis*, 6th Edition, Upper Saddle River, NJ: Prentice Hall.

Hall, Stuart (1980). "Encoding/Decoding," in Center for Contemporary Cultural Studies (Ed.): *Culture, Media, Language: Working Papers in Cultural Studies, 1972-79*, London: Hutchinson, 128-138.

Hardy, Melissa A. and Alan Bryman (2004), *Handbook of Data Analysis*, London: Sage Publications.

Holme, Petter (2003), "Network Dynamics of Ongoing Social Relationships," *Europhysics Letters*, 64 (3), 427-433.

Howard, John A., and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*, New York: John Wiley & Sons.

Hulberg, Jon (2006), "Integrating Corporate Branding and Sociological Paradigms: A Literature Study," *Brand Management*, 14 (1/2), 60-73.

Hutchinson, Wesley (1983), "Expertise and the Structure of Free Recall," in *Advances in Consumer Research*, Vol. 10, ed. Richard Bagozzi and Alice Tybout, Provo, UT: Association for Consumer Research, 585-587.

Janis, Irving Lester (1972), *Victims of Groupthink*, Boston: Houghton Mifflin.

John, Deborah Roedder (1999), "Consumer Socialization of Children: A Retrospective Look at Twenty-five Years of Research," *Journal of Consumer Research*, 26 (December), 183-213.

Jones, Thomas O. and Earl W. Sasser Jr. (1995), "Why Satisfied Customers Defect," *Harvard Business Review*, 73 (6), 88-102.

Kamada, Tomihisa and Satoru Kawai (1989), "An Algorithm for Drawing General Undirected Graphs," *Information Processing Letters*, 31 (1), 7-15.

Kano, Noriaki (1984), "Attractive Quality and Must-Be Quality," *The Journal of the Japanese Society for Quality Control*, April, 39-48.

Kapferer, Jean-Noel (1992), *Strategic Brand Management*, New York: Free Press.

---- (2004), *The New Strategic Brand Management: Creating and Sustaining Brand Equity Long Term*, London and Sterling, VA: Kogan Page.

Kassarjian, Harold H. (1977), "Content Analysis in Consumer Research," *Journal of Consumer Research*, 4 (June), 8-18.

Kawasaki, Guy (2004), *The Art of the Start: The Time-Tested, Battle-Hardened Guide for Anyone Starting Anything*, New York: Penguin Portfolio.

Keller, Kevin L. (2003), "Brand Synthesis: The Multidimensionality of Brand Knowledge," *Journal of Consumer Research*, 29 (March), 595-600.

Kerlinger, Fred H. (1964), *Foundations of Behavioral Research: Education and Psychological Inquiry*, New York: Holt, Rinehart and Winston.

Kermack, William O. and Anderson G. McKendrick (1927), "A Contribution to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society of London A*, 115, 700-721.

Kinkaid, D. Lawrence (1988), "The Convergence Theory and Intercultural Communication," in Young Y. Kim and William B. Gudykunst (eds.) *Theories in Intercultural Communication*, Thousand Oaks, CA: Sage, 280-298.

Kirk, Roger E. (1996), “Practical Significance: A Concept Whose Time Has Come,” *Educational and Psychological Measurement*, 56 (5), 746-759.

Kitchin, Robert (1998), *Cyberspace the World in the Wires*, Chichester: Wiley.

Kleine, Robert E., III, Susan Schultz Kleine and Jerome B. Kernan (1993), “Mundane Consumption and the Self: A Social-Identity Perspective,” *Journal of Consumer Psychology*, 2 (3), 209-235.

Kohonen, Teuvo (1990), “The Self-Organizing Map,” *Proceedings of the IEEE* 78 (9), 1464-1480.

Kozinets, Robert V. (2002). “The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities,” *Journal of Marketing Research*, 39 (February), 61-72.

Laughlin, Patrick R. and P.Christopher Earley. (1982), “Social Combination Models, Persuasive Arguments Theory, Social Comparison Theory, and Choice Shift,” *Journal of Personality and Social Psychology*, 42, 273-280.

Manzini, Ezio (1994), “Design, Environment and Social Quality: From “Existenzminimum” to “Quality Maximum,” *Design Issues*, 10 (1), 37-43.

March, James G. (1991), “Exploration and Exploitation in Organizational Learning,” *Organization Science*, 2, 71-87.

Mardia, Kanti V. (1970), “Measures of Multivariate Skewness and Kurtosis with Applications,” *Biometrika*, 57, 519-530.

Marsden, Paul S. (1998), “Memetics: a New Paradigm for Understanding Customer Behaviour and Influence,” *Marketing Intelligence & Planning*, 16 (6), 363-368.

Maslow, Abraham (1943), “A Theory of Human Motivation,” *Psychological Review*, 50, 370-396.

Matz, David C. and Wendy Wood (2005), “Cognitive Dissonance in Groups: The Consequences of Disagreement,” *Journal of Personality and Social Psychology*, 88 (1), 22-37.

Mayzlin, Dina (2006), “Promotional Chat on the Internet,” *Marketing Science*, 25 (2), 155-163.

McAlexander, James H., John W Schouten, and Harold F Koening (2002), "Building Brand Community," *Journal of Marketing*, 66 (January), 38-54.

McQueen, James B. (1967), "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.

Merriam-Webster (2007), *Merriam-Webster's Online Dictionary*, (accessed April 23, 2007), [available at http://www.m-w.com]

Mitchell, Andrew A. and Peter A. Dacin (1996), "The Assessment of Alternative Measures of Consumer Expertise," *Journal of Consumer Research*, 23 (3), 219-239.

Moscovici, Serge and Marisa Zavalloni (1969), "The Group as a Polarizer of Attitudes," *Journal of Personality and Social Psychology*, 12, 125-135.

Muniz, Albert and Thomas O'Guinn (2001), "Brand Community," *Journal of Consumer Research*, 27 (March), 412-432.

Murray, Richard (2007), "The Horizontal Axis of Communication," *A Corporate Guide to the Global Blogosphere*, an Edelman White Paper (accessed May 3, 2008), [available at http://www.edelman.com/image/insights/content/WhitePaper011107sm.pdf]

Obloj, Tomasz and Krzysztof Obloj (2006), "Diminishing Returns from Reputation: Do Followers Have a Competitive Advantage?" *Corporate Reputation Review*, 9 (4), 213-224.

Olson, Jerry C. and Aydin Muderrisoglu (1979), "The Stability of Responses Obtained by Free Elicitation: Implications for Measuring Attribute Salience and Memory Structure," *Advances in Consumer Research*, 6 (1), 269-275.

Page, Scott (2007), *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton, NJ: Princeton University Press

Pereira, Paulo T., Nuno Silva and Joao Andrade e Silva (2006), "Positive and Negative Reciprocity in the Labor Market," *Journal of Economic Behavior & Organization*, 59, 406-422.

Peterson, Robert A. (2005), "Response Construction in Consumer Behavior Research," *Journal of Business Research*, 58 (3), 348-353.

Petty, Richard E. and John T. Cacioppo (1986), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, New York: Springer-Verlag.

----, Rao H. Unnava, and Alan J. Strathman (1991), "Theories of Attitude Change," in Harold H. Kassarjian and Thomas S. Robertson (Eds.), *Handbook of Consumer Behavior*, New York: Prentice-Hall, 241-280.

Phelan, Steven E. (2001), "What is Complexity Science, Really?" *Emergence*, 3 (1), 120-136.

Pink, Daniel H. (2005), *A Whole New Mind*, New York: Riverhead Books.

Ratneshwar, S. (Ratti) and Allan D. Shocker, (1991), "Substitution in Use and the Role of Usage Context in Product Category Structures," *Journal of Marketing Research*, 28 (August), 281-295.

Reynolds, Craig W. (1987), "Flocks, Herds, and Schools: A Distributed Behavioral Model," *Computer Graphics*, 21(4) (SIGGRAPH '87 Conference Proceedings), 25-34.

Roberts, Peter W. and Grahame Dowling (2002), "Corporate Reputation and Sustained Superior Financial Performance," *Strategic Management Journal*, 23 (12), 1077-1093.

Roget, Peter Mark (1962), *Roget's Thesaurus*, 3th ed. New York: Thomas Y. Crowell.

Rosa, Jose Antonio, Jelena Spanjol and Michael S. Saxon (1999), "Sociocognitive Dynamics in a Product Market," *Journal of Marketing*, 63 (Special Issue), 64-77.

Rosen, Devon (2002), "Flock Theory: Cooperative Evolution and Self-Organization of Social Systems," *Proceedings of the 2002 CASOS (Computational Analysis of Social and Organizational Systems) Conference*. Carnegie Mellon University, Pittsburgh, PA.

Sabate, Juan Manuel de la Fuente and Ester de Quevedo Puente (2003), "Empirical Analysis of the Relationship between Corporate Reputation and Financial Performance: A Survey of the Literature," *Corporate Reputation Review*, 6 (2), 147-160.

Satorra, Albert, and Peter M. Bentler (1994), "Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis," in A. von Eye and C. C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research*, Thousand Oaks, CA: Sage, 399-419.

Sawyer, R. Keith (1999), "The Emergence of Creativity," *Philosophical Psychology*, 12 (4), 447-469.

Schwarz, Gideon, (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6 (2), 461-464.

Scoble, Robert and Shel Israel (2006). *Naked Conversations*, Hoboken, NJ: John Wiley & Sons

Shannon, Claude E. and Warren Weaver (1949). *A Mathematical Model of Communication*. Urbana, IL: University of Illinois Press.

Shirky, Clay (2003), *Power Laws, Weblogs and Inequality,* (accessed May 8, 2007), [available at http://www.shirky.com/writings/powerlaw_weblog.html]

Shoham, Aviv (2004), "Flow Experiences and Image Making: An Online Chat-Room Ethnography," *Psychology & Marketing*, 21 (10), 855-882.

Simmel, Georg and Donald Levine (1972), *Georg Simmel on Individuality and Social Forms*, Chicago, IL: University of Chicago Press.

Simon, Julian (2007), *Hierarchical Clustering*, (accessed April 23, 2007), [available at http://www.resample.com/xlminer/help/HClst/HClst_intro.htm]

Singer, Judith D. (1998), "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models," *Journal of Educational and Behavioral Statistics*, 24 (4), 323-355.

Sujan, Mita (1985), "Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgment," *Journal of Consumer Research*, 12 (June), 31-46.

Summers, John O. (2001), "Guidelines for Conducting Research and Publishing in Marketing: From Conceptualization Through the Review Process," *Journal of the Academy of Marketing Science*, 29 (5), 405-415.

Surowiecki, James (2004), *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, New York: Doubleday.

Technorati (2007), *About Us*, (accessed Aug 21, 2007), [available at http://technorati.com/about]

Thompson, Craig J. (1997), "Interpreting Consumers: A Hermeneutical Framework for Deriving Marketing Insights from the Texts of Consumers' Consumption Stories," *Journal of Marketing Research*, 34 (November), 438-455.

Thorndike, Robert M. (1996), *Measurement and Evaluation in Psychology and Education*, 6th ed. Upper Saddle River, NJ: Prentice Hall.

Toner, John and Yuhai Tu (1998), "Flocks, Herds, and Schools: A Quantitative Theory of Flocking," *Physical Review E*, 58 (4), 4828-4858.

Van den Bulte, Christophe and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26(3), 400-421.

Vernette, Eric (2004), "Targeting Women's Clothing Fashion Opinion Leaders in Media Planning: An Application for Magazines," *Journal of Advertising Research*, 44 (1), 90-107.

Waldrop, M. Mitchell (1992), *Complexity: The Emerging Science at the Edge of Order and Chaos*, New York: Simon & Schuster.

Wallace, Patricia M. (1999), *The Psychology of the Internet*, Cambridge, UK: Cambridge University Press.

Ward, James C. and Peter H. Reingen (1990), "Sociocognitive Analysis of Group Decision Making among Consumers," *Journal of Consumer Research*, 17 (3), 245-262.

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge, UK: Cambridge University Press.

Watts, Duncan J. and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34, 441-458.

Weick, Karl E., Kathlen M. Sutcliffe and David Obstfeld (2005), "Organizing and the Process of Sensemaking," *Organization Science*, 16 (4), 409-421.

Wikipedia (2007), *Wikipedia, The Free Encyclopedia,* (accessed April 23, 2007), [available at http://en.wikipedia.org]

Woelfel, Joseph (2007), *Who are we?*, (accessed March 7, 2007), [available at http://www.galileoco.com/N_who.asp]

---- and Edward L. Fink (1980), *The Measurement of Communication Processes: Galileo Theory and Method*. New York: Academic Press.

Zaltman, Gerald (1982), *Theory Construction in Marketing*, New York: Wiley.

Zhang, Jin, David A. Bessler and David J. Leatham (2006), "Does Consumer Debt Cause Economic Recession? Evidence using Directed Acyclic Graphs," *Applied Economics Letters*, 13, 401-407.

Zhang, Tian, Raghu Ramakrishnon and Miron Livny (1996). "BIRCH: An Efficient Data Clustering Method for Very Large Databases," *Proceedings of the ACM SIGMOD Conference on Management of Data*, Montreal, Canada, 103-114.

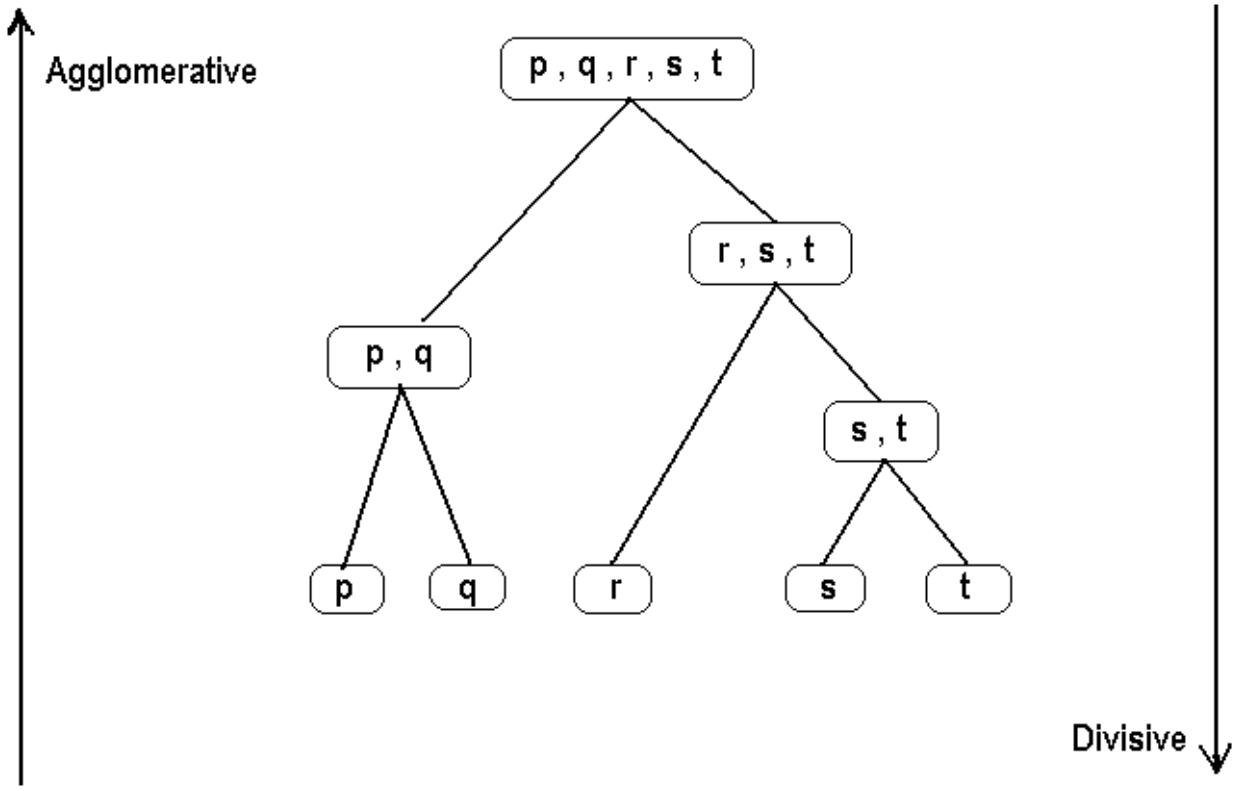
APPENDIX A

GLOSSARY

Agglomerative Hierarchical Clustering: Simon (2007) describe agglomerative hierarchical clustering as a bottom-up clustering algorithm where new clusters are formed one-at-a-time from existing objects and clusters, generally by combining objects or *clusters* closest in proximity to each other. Each step in the clustering process becomes a level in a hierarchy. Such hierarchies are often presented in two-dimensional diagrams known as dendrograms. An example of a dendrogram showing an agglomerative hierarchical cluster is presented in Figure A-1.

FIGURE A-1

Dendrogram of an Agglomerative Hierarchical Cluster



Alignment: One of Reynolds' (1987) three simple "steering behaviors" that characterize *flocking theory*. Alignment refers to the behavior of steering toward the average heading of local flock mates.

Allegory: One of Brown et al's (2003) four classes of meaning that can be assigned to a *brand*. Specifically, an allegory is a story with symbolic content where the brand can be readily associated with one of the symbols and thus be endowed with meaning by the story. For example, a financial services firm used a story called *The "Tree" and I* that conveyed the message that if clients nurture the tree, a symbol for an investment, the tree would take care of them when they got older.

Antinomy: One of Brown et al's (2003) four classes of meaning that can be assigned to a *brand*. Antinomy refers to a contradiction or paradox. The term often refers to attempts to revive an old *brand* as a new product such as was done with the restyled 2005 Ford Mustang which combined modern technology with styling reminiscent of the classic Mustangs of the 1960's.

Anthropomorphize: Merriam-Webster (2007) describes anthropomorphication as the assigning of human characteristics to an inanimate object. Freling and Forbes' (2005) found anthropomorphication to be a common phenomenon when people form a relationship with a *brand*. The *brand* is seen as having a personality and thus personhood.

Arcadia: One of Brown et al's (2003) four classes of meaning that can be assigned to a *brand*. In a general sense the word "arcadia" is used to refer to a place characterized by peace, tranquility and simplicity. It is often used to refer to someplace at a time in the past when popular memory sees conditions as being better. A *brand* can be linked with such time and place associations, such as the *brand* of oatmeal your grandmother served you in the happy days of your childhood.

Associations: Merriam-Webster (2007) describes association as one or more ideas, images, emotions or thoughts connected to something other than itself. This paper uses the word in two ways: *brand* associations (thoughts connected to a *brand*) and *cognitive associations* (the connection of different thoughts). Keller (2003) note these connections can be the result of personal contemplation or they can be prompted by an outside influence such as an advertisement.

Attitude: Petty et al (1991) describe attitude as a relatively enduring disposition of thought toward a person or object, usually the result of evaluation. The word is often used to denote a global evaluation but can be used in a more qualified sense: *brand* A is good but only useful in situation X.

Aura: One of Brown et al's (2003) four classes of meaning that can be assigned to a *brand*. Aura refers specifically to values that a *brand* is thought to stand for. The term is often applied to retro branding scenarios where old and established *brands* are so associated with an adherence to old-world standards of uncompromising quality that the brand itself conveys that meaning. It is not uncommon for an expression like "*brand* A is the Rolls Royce of garbage disposals" to be used. The invocation of the Rolls Royce automobile *brand* is irrelevant to garbage disposals except for the aura of almost holy reverence the *brand* name evokes.

Autonomy: Merriam-Webster (2007) describes autonomy as a state of being characterized by independence and freedom. In a culture where independence is prized over more collective concepts, it is often necessary for individuals to strike a balance between the autonomy of unique self-expression and adherence to community standards of acceptable expression. Flocking theory is one model of how individuals can strike such a balance and still function effectively in helping a community achieve its goals.

Bayesian Information Criterion: An alternative name for Schwarz' (1978) criterion, a statistical measure used to compare alternative models proposed to explain a set of data. The model with the lowest Schwarz criterion value fits the data best and is deemed to be more correct.

Betweenness: An abbreviated form of the term *betweenness centrality*, defined below.

Betweenness Centrality: Corman et al (2002) describes betweenness centrality as a term from network science used to measure the degree to which a member of a network connects otherwise disparate parts of the network together. Without this member, the network parts (cliques or sub-networks) would be separate or substantially less connected than they are with the member in place.

Blog: An abbreviated form of *weblog*, defined later.

Blogosphere: Wikipedia (2007) uses the term "blogosphere" to denote the collection of all *weblogs*.

Boldness: *Idea seeders* and *evangelists* are distinguished from *flockers* by one general attribute: boldness, a willingness to express thought with little regard to the opinions of others. Boldness is used in this study as an informal construct created by combining three parameters from equation 3.2: the number of *thematic clusters* started, the standard deviation in comment-to-closest-thematic-cluster separation, and the number of comments contributed. Boldness as an informal construct is used as an axis in Figure 3.8 to show how *evangelists* and *idea seeders* differ from *flockers*.

Brand: Kapferer (1992) notes the term "brand" is often is used to refer to concrete symbols, such as a name, that are legally associated with an economic entity like a company. However, these literal symbols often acquire certain cognitive associations (thoughts, images and ideas) in the minds of people to such an extent that the thoughts, symbols and entity are inseparable.

Brand Equity: Kapferer (1992) describes brand equity as the value associated with a brand. It is seen as the additional money a consumer will spend to buy a branded product over an identical unbranded product.

Brand Identity Prism: The *brand* identity prism is a model proposed by Kapferer (1992) that introduced the idea of "brand personality," a term denoting the common human behavior of associating human personality traits with a brand.

Centering Resonance Analysis: Corman et al (2002) introduced centering resonance analysis as a *content analysis* methodology where a body of text is transformed into a network of nouns and adjectives based on the position of those words in the sentences composing that body of text. Words can be given a measure of *betweenness centrality*, called *influence*. Bodies of text can be compared on the basis of their containing the same words in similar positions of influence using a measure called *resonance* (each italicized word is defined in this glossary).

Centroid: Abdi (2007) describes the centroid or barycenter of a physical object as the coordinate in n-dimensional space from which the mass of the object can be conceived as originating. Since this proposal deals exclusively with intangible objects, the centroid is a geometrically calculated center of a cluster of such objects.

Circuit of Communication: The circuit of communication is a model of a communication process proposed by Hall (1980) where through an ongoing cycle of feedback and clarification, parties in conversation converge on a shared understanding.

Cluster: Chiu et al (2001) describe a cluster as a set of objects, tangible or intangible, that by virtue of their proximity in some dimensional space of reference are regarded as being associable into one entity that comes to represent them all.

Cluster Feature Tree: Zhang et al (1996) describe a *cluster* feature tree as a compact way of storing information about a cluster or series of nested clusters (clusters within clusters). Its conceptualization is based on the metaphor of a tree where the objects in the *clusters* are like leaves and the ways the objects are connected into *clusters* are like the branches.

Cluster Sample: Wikipedia (2007) describes cluster sampling as two-stage sampling: first you select a sample of areas of a population, and then you sample subjects within that area. This study seeks to make general findings about the whole blogosphere. The

first sample is the selection of the 14 blogs in Table 4.1 from all blogs in the blogosphere using *netnography's* criterion. The second sample would be the selection of commenters within the blogs. However, in this study all the commenters in each blog are used, except in the pretest dataset. Cluster samples differ from random samples in that it is understood that the subjects within each cluster are more similar than they would be in a random sample.

Cognitive Association: Brown et al (2006) describe a cognitive association as the establishing of a memory link between two or more *objects of cognition*, representing a relationship of juxtaposition (the objects are always together) or fusion (the objects are the same).

Cognitive Association Maps: Keller (2003) describes a cognitive association map as a dimensional space of reference where *objects of cognition* can be located and their proximity to other *objects of cognition* assessed.

Cognitive Diversity: In this paper cognitive diversity is used to denote the many cognitive styles or ways individuals think, perceive and remember information and then use that information to solve problems or, in general, enact their behavior. There are many theories that try to explain differences in cognitive style: the nature (genetic predisposition) versus nurture (result of formative experiences) debate is a well-known comparison of two such theories. Contemporary organizational theory takes the pragmatic view that regardless of why they exist, assembling teams based on cognitive diversity results in better problem solving.

Cognitive Structure: Ward and Reingen (1990) describe a cognitive structure as an encoding of information in the mind. This encoding can be a representation of some tangible or intangible object, relationship, concept or body of knowledge. It can represent one simple object, a theory of how the universe works and everything in between. Note that a distinction is made between the brain (a physical organ) and the mind (the result of electro-chemical reactions that form the operating system of the brain and give us consciousness).

Cohesion: One of Reynolds' (1987) three simple "steering behaviors" that characterize flocking theory. Cohesion refers to the behavior of steering toward the average position of local flock mates.

Collective Thought Diversity: Collective Thought Diversity is used as an informal construct created by combining two parameters from equation 3.2: Mean and Stdev Collective Thought Separation. It is designed to represent the degree of thematic difference between individuals' expressions of thought that are in the same collaboration context (e.g., writing comments to the same blog entry). One way the term is used in this study is as a latent sub-factor whose value differentiates *cultural tribalism* (the same

ideas are repeatedly echoed) from *need-for-cognition* (intellectual curiosity stimulates diverse ideas).

Comments: Comments are the means whereby readers express their thoughts on the *post* or *entry* written by the *weblog author* and the comments written by other readers.

Communication Convergence: A network-based model proposed by Kinkaid (1988) that explains how people congregate about a central theme in conversation. Kinkaid found that the density of communication networks indicates social cohesion.

Community: Kleine et al (1993) describe a community as a group of people who share in common things (e.g., goals, beliefs, resources, preferences, needs, risks, challenges) that reflect their individual identities. While some research may ascribe individual identity to group membership, this study focuses on *emergence*, demonstrated in this context by individuals choosing to associate based on commonalities shared prior to group membership.

Complexity: Waldrop (1992) describes complexity as an abbreviated reference to complex systems, mechanisms that tend to be high in dimensionality (many underlying factors), non-linear and therefore difficult to model. The behavior of a complex system is often attributed to *emergence*, a phenomenon whereby complex systems behavior is the result of the interaction of simple systems. That is a point-of-view embraced in this study.

Construed Image: Brown et al (2006) describe how companies try to determine how they are viewed by consumers as a part of the process of corporate image-building (Figure 2.1). Any such impression will be inaccurate to some extent; however, it is desirable to minimize this inaccuracy by looking for better ways to measure it. The image thought to be in the mind of the consumer is compared with the image the company wants to convey on an ongoing basis. Mismatches between the two images are used to plan future image-building activities.

Content Analysis: Kassarijian (1977) notes that content analysis is generally defined as the study of written communications. The three common goals of content analysis are to determine: the cognitive state (emotions, motivations, knowledge) of the writer, the intended meaning of what was written, and the effect the writing probably had on its readers. Content analysis usually is focused on counting the use of specific words and their synonyms, assuming the most used words reflect the most important ideas intended to be conveyed.

Consumer-Generated Media: Consumer-generated media is a synonym for *user-generated media*, defined later.

Critical Thinking: Facione (2007) describes critical thinking as a form of judgment that reflects on evidence, context and theory to understand the nature of the problem at hand and then gathers and evaluates information needed to come to a well-justified conclusion.

Cultural Tribalism: Kitchin (1998) describes cultural tribalism as a state that a *virtual community* can be in where the members share a strong ideological identity that welcomes the few new members who agree but eschews the many who do not. As a result, conversation tends to be between the same people, and is thus confined to the repetition of the same ideas; as a result, the term *echo chamber* is used synonymously. Cultural tribalism evolves out of people settling into communities where the opinions expressed match what they want to hear. Blogs characterized by cultural tribalism may also be those where blog authors are exercising *indoctrination*, an ongoing attempt to convert people to a way of thinking. What the author posts draws a homogenous group of vocal devotees.

Echo Chamber: The term "echo chamber" is a synonym for *cultural tribalism*, already defined.

Eigenvalue: Measurements of *complex* phenomena are thought to reflect the operation of simpler, non-divisible influences. Although these influences may be psychological in nature, their operation is often modeled as though they were physical forces. Physical forces are modeled as vectors, that is, they operate along a line with a magnitude and direction. While such forces can act in any direction, when dissected into their most simple form they are the result of a set of orthogonal (at right angles) vectors (called eigenvectors) that act along the axes of a dimensional space. Measurements of magnitude on theses axes are eigenvalues.

Elaboration Continuum: Petty and Cacioppo's (1986) Elaboration Likelihood Model of attitude change describes an elaboration continuum, a spectrum of thought depth in response to attempts at persuasion. How deeply a target of persuasion is motivated to think depends principally on assessments of personal relevance and *need-for-cognition*, a liking of deep thinking.

Emergence: Goldstein (1999) offers a contemporary definition of emergence: "the arising of novel and coherent structures, patterns and properties during the process of self-organization in *complex systems*" (p. 51). The key factors of this definition are: novelty (features not previously seen), coherence (self-maintaining), evolution and self-organization (simple entities or processes collectively enact *complex* behaviors).

Empirical: Summers (2001) described three kinds of scientific contribution that can be made: *theoretical*, *methodological* and empirical. The word empirical is used as an adjective to describe results derived from real world observation or experimentation. The word is used in contrast to "analytical", the results of a mathematical or theoretical proof

and "simulated", results derived from observing a model such as a computer program designed to mimic a real world behavior.

Entry: Scoble and Israel (2006) use the words entry and *post* as synonyms for the content a *weblog author* introduces to a blog to set the subject of conversation. The author posts content in expectation that readers will reply with comments thus enabling the author to learn from the reader community.

Entropy: Chiu et al (2001) note that "entropy" is used to refer to the loss of information that would occur if two *objects of cognition* or *clusters* of such objects were combined into one *cluster*. Stating that two objects are unique enough to persist as separate entities implies that these objects represent dissimilar information. If two objects are judged similar enough to combine into a *cluster*, and then always referenced as a *cluster*, some unique information belonging to the separate objects is lost. The clustering technique used in this study seeks to minimize such information loss.

Entropy-based Hierarchical Clustering: Entropy-based hierarchical clustering is similar to the *agglomerative hierarchical clustering* already described except proximity between objects is assessed using *entropy*, or information similarity, rather than some other measure of cognitive or literal distance.

Equilibrium: Devon (2002) describes equilibrium as a state of rest due to the balance of opposing forces. In this study, the expression of thought in groups is investigated. It is proposed that some people will either consciously or subconsciously express their thoughts in a manner designed to balance their desire to make a unique contribution with their desire to support norms of community conformity.

Evangelists: In this study, evangelists are the agents in the process of *mass dissemination*. They are highly prolific commenters that exhibit a low diversity of thought expression in their comments across blog entries. They are similar to *idea seeders* in that they are prolific but they differ in the narrow variation in ideas expressed.

Evolution: This study uses the term "evolution" in the same way as Dawkins (1979): a process of change whereby a species increases its survivability from generation to generation by slow changes to its genetic structure (the basic units of heredity) under the influence of random mutation, procreation strategies and natural selection (survival of the fittest). Dawkins applied the evolutionary process to explaining the way culture is transmitted over time in his description of the meme, the idea equivalent of the gene. Some ideas spread like a virus, mutated as they spread by people that express them in ways that make them palatable to the hearers.

Expectation Maximization: Chiu et al (2001) describe expectation maximization as an iterative algorithm used in a variety of applications including clustering. The algorithm iterates through two phases: expectation and maximization. In the expectation

phase of clustering, the likelihood of the objects belonging to the clusters already specified is calculated. In the maximization phase, the way objects are assigned to clusters is altered to maximize the likelihood values calculated in the expectation step. The two steps are iterated until no more likely way of assigning objects to clusters is found.

Expertise: Ratneshwar and Shocker (1991) describe expertise as the possession of a high level of knowledge and the skill needed to apply it. In this study, expertise is seen as one type of brand association where a person has acquired a high degree of knowledge related to the unique features of a product or product family and how to use them. This kind of brand association can be very valuable to the holder and constitute the basis for the way they make a living. This is a common situation for technology products with complex functionality like database software and management information systems.

Fixed Point Attractor: Attractors are states toward which some dynamic systems move. A fixed point attractor is one type of attractor; it is an equilibrium state where a system comes to rest. Cambel (1993) characterizes the movement of a system toward an attractor to be like a firefly attracted to a light. Cohen and Stewart (1994) note that being in an attractor state does not imply being stationary. They use as an example a table tennis ball floating in the ocean: push the ball down it always rises to the surface (the attractor state), however on the surface it is subject to all the forces in the environment (e.g., wind and current).

Flocker: This study classifies commenters as *idea seeders*, *evangelists* and flockers. Flockers are significantly more inhibited in expressing their thoughts in a group forum than the other commenter types. They are more concerned with maintaining their acceptance within a group and therefore control the degree to which their expressed thought diverges from that of the group.

Flocking Theory: Flocking theory is a model of how groups of birds and animals move in groups that Rosen (2002) applied to explain how humans express their thoughts in a group context. Every member of the flock employs three simple "steering behaviors": *separation*, *alignment* and *cohesion* which, taken together, allows each member to be in their desired location, relative to the group, one unit of time in the future. This model is used to explain the balance between expressing individualism and collective solidarity that some researchers have observed in human discourse.

Forum: Wikipedia (2007) describes a forum as a website on the Internet that hosts discussions and the posting of *user-generated content*. Forums often become *virtual communities* as they attract a group of regular readers who enjoy contact with people who share one or more interests. Forums are often referred to as discussion boards and message boards.

Free Elicitation: "Free elicitation" is a technique for getting consumers to reveal their thoughts introduced by researchers Olson and Muderrisoglu (1979). The technique allows the consumer to say anything that comes to mind in response to a question referred to as a "stimulus probe cue."

Free Rider: Pereira et al (2006) use the term "free rider" to refer to economic actors who consume more than their fair share of a resource while paying less than their fair share of the cost of production. Free riding is often regarded as a problem because it can cause essential *public goods* to be under-produced.

Fungible: Merriam-Webster (2007) describe the term "fungible" as an economic term usually applied to goods that can easily be exchanged for another instance of the same good. Fungibility is often confused with the term liquidity, the ease of exchanging a good for money. In this study, the term is used to apply to unique associations that consumers make with *brands* that make it difficult for other *brands* to be substituted.

Generative Process: Phelan (2001) notes that in *social process theory* and *complexity* theory in general, phenomena that are very difficult to model and explain are often seen as being the result of the interaction of simple processes referred to as "generative processes."

Generative Rules: Cederman (2005) notes that sometimes phenomena are explained by appealing to universal or high-level laws that force certain outcomes. These explanations are very common in explaining physical phenomena like the conversion of friction into heat by appealing to the laws of thermodynamics. However, social phenomena resist being robustly explained by appeals to universal laws even though attempts to do so are common. An alternative view explains *complex* phenomena by attributing them to the interaction of *generative processes*. When specific processes interact, specific outcomes can be predicted as described by generative rules, well-substantiated cause-and-effect relationships that have assumed the status of expressions of certainty.

Grassroots: A movement driven by the spontaneous, bottom-up organizing of the members of a community, usually around some issue. Such movements draw their credibility from the perception that the focal issue is so important to the members of a community that it independently motivates individuals into collective action.

Group Polarization: Burnstein and Sentis (1981) use the term "group polarization" to refer to a migration to extreme positions during a discussion. The result is that discussants gradually come to advocate more extreme positions than they would before the group met. Researchers have attributed the effect to the influence of persuasive argument combined with the desire for people to feel socially accepted. People self-assemble into positional clusters that tend to represent extreme versions of the arguments they find most plausible. *Groupthink* is often confused with group polarization and

indeed the two can occur together. As group polarization becomes evident within the group, people suppress the desire to introduce information inconsistent with the position held by the people they feel most attached to.

Groupthink: Groupthink is sometimes regarded as synonym for *group polarization*, already defined. However, as Ward and Reingen (1990) note, the term describes the specific situation where in group decision-making the members consciously strive to go along with what they believe is a consensus. It is a situation commonly found in highly cohesive groups where unanimity is a more important goal than properly appraising alternatives.

Hermeneutic: Thompson (1997) describes hermeneutics as the study of text. However, Thompson also notes it has come to be associated with the term "hermeneutic circle" which maintains that a text cannot be interpreted without reference to its cultural, literary (one part cannot be understood without reference to the whole) and historical contexts.

Holdout Sample: Often a dataset is divided into two parts: a training or estimation dataset and a holdout or testing dataset. The training data set is used to estimate the parameters of a model, while the holdout dataset is used to determine the predictive accuracy of the model.

Hypothesis: Zaltman (1982) describes a hypothesis as a tentative assertion that requires validation through testing. Such assertions usually propose either a causal relationship, such as A causes B, or a softer proposition of correlation, such as A is related to B.

Idea Seeder: "Idea seeder" is a term used in this study to denote those regular and prolific commenters who are least constrained by any norms of conformity and therefore freely express comments on a variety of themes, regardless of the theme of the blog entry. These people are hypothesized to be serial starters of thematic clusters that differ from the cluster containing the blog entry.

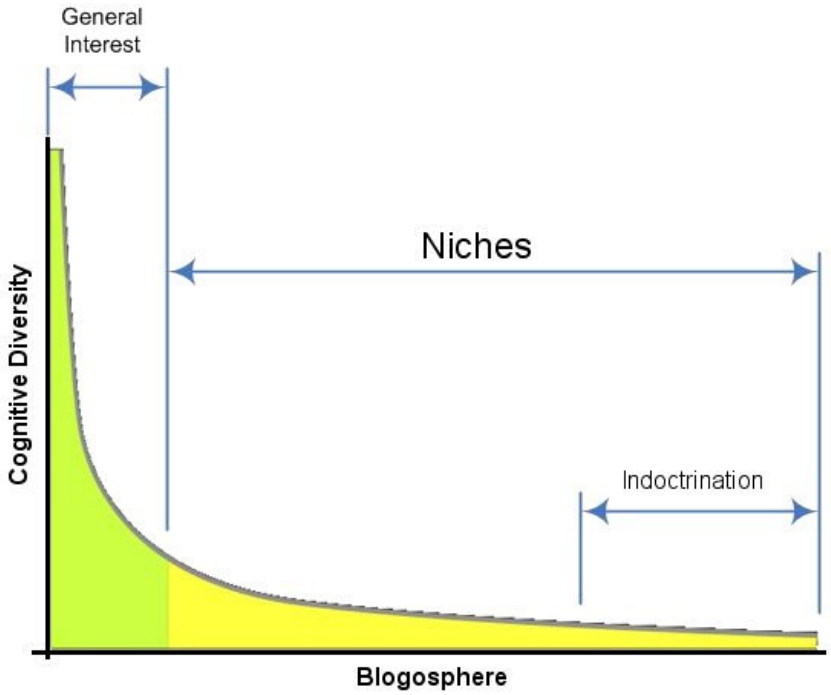
Identity: Brown et al (2006) use the term "identity" to denote how an organization or person views the self. It is distinct from any use of the term "image" which denotes how the self is viewed by others. Entities can undergo a process of image-crafting to control how others view them. However, they also participate in identity-crafting where they work to achieve some desired state of self as assessed by self-reflection.

Image: Gioia et al (2000) use the term "image" to refer to how, by way of cognitive associations, others view the self. It is described as being different from the concept of *identity*, defined above. There are three variations of the concept of image used in this paper: *intended*, *construed* and actual. The actual image of some entity is only known to the holder. The entity can *construe* this image and try to influence it to become more like its *intended* image.

Individual Thought Diversity: Individual Thought Diversity is used as an informal construct created by combining two parameters from equation 3.2: Mean and Stdev Individual Thought Separation. It is designed to represent the freedom with which an individual expresses thought relative to prior thoughts expressed. One way the term is used in this study is as a latent sub-factor whose value differentiates diverse-thinking *idea seeders* from the limited thinking *evangelists*.

FIGURE A-2

The Long Tail of the Blogosphere



Indoctrination: Dawkins (1976) note that for memes to be the cognitive equivalent of genes and evolve under the influence of an evolutionary process, they require a means of procreation. Dawkins (2006) proposed two basic forms of meme propagation: indoctrination (intense repetition) and *mass dissemination* (spread it widely). Another attribute of the process of indoctrination is that critical thinking is discouraged: accept the ideology without question. Shirky (2003) observed from his survey of a wide variety of research, that the number of readers per blog across the blogosphere has a distribution similar to that presented in Figure A-2. A few blogs attract general interest and many readers while most appeal to a diverse array of small readership niche interests. This study assumes, on the basis of the same logic used when describing *cultural tribalism*,

that cognitive diversity across the blogosphere has the same distribution. As a result, indoctrination blogs are predicted to come from the extreme end of the tail of that distribution.

Influence: In Corman et al's (2002) *Centering Resonance Analysis (CRA)*, the *betweenness centrality* of a word is called its influence. According to CRA, words of high influence are those that are repeatedly used to connect other words together and are thus the words that most reveal cognitive themes.

Inherent Classification: Chiu et al (2001) use the term "inherent classification" to denote clustering algorithms that do not require the user to specify beforehand the number of clusters to be found. These algorithms try to find the clusters that are naturally present in the data.

Intangible: Arnould and Thompson (2005) note that if something is intangible it cannot be perceived by the senses; it is incapable of being realized. They further note that many cognitive structures are intangible.

Intended Image: Brown et al (2006) call the *cognitive associations* that a company would like to instill in the minds of consumers its intended image. This image may or may not bear a resemblance to that company's *identity*, its view of itself. Both companies and individuals often try to say that their intended image and their *identity* are the same: "come and get to know the real me."

Interclass Correlation: Muthen (1997) describes interclass correlation is a measure of how alike one group, as a whole, is to all groups.

Interpretive Repertoire: A term introduced by Grayson (1998) to explain why receivers may not understand the message transmitted by a sender. When a company attempts to convey its *intended image* it creates a message that it can encode in a variety of ways (Figure 2.1), unless the receiver understands the way the message is encoded (words, figures of speech, visual metaphors, etc.) they cannot decode the message and adopt the *intended image*. The resources the receiver will draw from to understand and decode the message are termed the receiver's interpretive repertoire.

Intraclass Correlation: Muthen (1997) describes intraclass correlation is a measure of how alike members of a group are to each other, relative to how alike all members, irrespective of group, are.

K-Means: McQueen (1967) introduced K-Means as a clustering algorithm that uses an iterative approach to assigning objects to a user supplied number (K) of clusters. The algorithm assigns the objects to K clusters at random and then calculates the centroid of each cluster. The algorithm then reassigns objects to clusters by trying to minimize distances between objects and cluster centroids. The algorithm is fast but not necessarily

good at finding the global optima. Many users repeat the algorithm several times and then pick the best result.

Knowledge: Keller (2003) notes knowledge is one type of brand association that a person can have. Knowledge is often related to *expertise*, defined above.

Latent Content Analysis: Kassarjian (1977) notes that latent and *manifest* content analyses are the two main methodologies for analyzing text. Kassarjian explains that the latent method attempts to capture the themes embedded in the text and thus is generally considered to require human interpreters. Conclusions are based on inter-rater or inter-coder reliability, that is, the level of agreement between the interpreters.

Long Tail: Wikipedia (2007) notes that the phrase *The Long Tail* was originated by Anderson (2004) when describing the business model of firms that earn most of their revenue catering to the needs of many people with diverse niche interests rather than large mass markets. It is a reference to the shape of a *power-law* distribution, presented in Figure A-2, where most of the phenomena depicted inhabits a long tail of low frequency events.

Manifest Content Analysis: Kassarjian (1977) notes manifest and *latent* content analyses are the two main methodologies for analyzing text. The manifest method involves the counting of words and their frequencies, therefore ignoring the deeper meanings of the text.

Mass Dissemination: Dawkins (2006) maintains that mass dissemination and *indoctrination* are the primary *memetic* propagation mechanisms. This study calls the agents of mass dissemination in the *blogosphere, evangelists*.

Meaning: Brown et al (2003) describe meaning as a type of *cognitive association* that can be attached to a *brand*. There are four types of meaning that prior research has identified: a*llegory* (a brand story), *aura* (core values), *arcadia* (idealized community) and *antinomy* (paradox: old style, new technology).

Meme: The word "meme" was introduced by Dawkins (1976) in his book *The Selfish Gene*. It refers to ideas, particularly cultural norms, which are propagated from generation to generation in a manner similar to biological genes using the evolutionary process of random mutation, propagation and natural selection. Ideas that lead to behaviors that help humans be more fitted to survival are continuously adapted to circumstance and passed down through the generations as culture. The word is also applied to ideas that people spread in the same way they might spread the flu virus: "idea viruses."

Memeset: Blackmore (2001) describes a memeset or memeplex as a system of related *memes* that reinforce each other and increase their chances of propagation.

Memetics: Blackmore (2001) describes memetics is the science that studies memes and memetic algorithms, the process of adapting memes to increase their survivability. Blackmore notes that the process of meme propagation is different from that of biological genes in that memes can be altered by the people that hold them. People can alter the ideas they spread to make them more applicable to changing circumstances, including changes in what people are willing to accept.

Message Board: Wikipedia (2007) describes a message board as a website that allows people to hold discussions and post *user-generated media*. Message boards are also called *forums*, as already discussed.

Methodological: Summers (2001) described three kinds of scientific contribution that can be made: *theoretical*, methodological and *empirical*. A methodology is a process or procedure that is accepted by a discipline as a means of knowledge discovery. A methodological scientific contribution is the introduction of a new procedure to a discipline. Methodological contributions must be based on accepted theory and offer some demonstrable advantage over current methods.

Multidimensional Scaling: Green et al (1975) describe multidimensional scaling (MDS) as a means of modeling relationships between objects that can be represented spatially. Traditionally, MDS has only involved two- or three-dimensional representations because they can be visualized. However, data mining technology has allowed the detection of sophisticated patterns in data without human agency so the dimensionality of data spaces can be unconstrained. Any object that can be expressed as spatial coordinates using *eigenvalue*-based decomposition can be located and analyzed within a multidimensional space.

Need-for-Cognition: Petty and Cacioppo (1986) described their Elaboration Likelihood Model (ELM) of persuasion as a model of attitudinal formation and change. When someone receives a message intended to evoke attitudinal change, message effectiveness may well depend on triggering an intrinsic need-for-cognition motivation in the hearer, that is, an innate enjoyment of thinking. In this study, need-for-cognition causes people to deliberate over new ideas, expressing diverse thoughts as they try to make sense of new ideas and fit them into their conceptual structures.

Netnography: Kozinets (2002) intoduced *netnography* as a methodology where the principles of ethnography, or unobtrusive observation, were applied to the study of virtual community. He specified a selection criterion for subject communities that differed from the practice of standard ethnography: select communities that are focused on a research question-relevant topic, receive above average posting traffic, have a large number of contributing members, contain descriptively rich content and thus, in general, enjoy a high level of member-to-member interaction.

Network-based Model: Wasserman and Faust (1994) note that a network-based or simply a network model represents the relationships between objects by connecting them in a spider web-like fashion with ties. These ties can be weighted to represent the strength of relationships and they can be set to a length that conveys spatial proximity. Network models can be analyzed by the methodologies that have become the tools of network science, the study of network phenomena. A principal reference is Wasserman and Faust's (1994) book: *Social Network Analysis.*

News Group: Wikipedia (2007) describes a news group as a website that acts as a repository for messages posted by users. News groups are similar to *forums* or *message boards* but are regarded as being technically distinct because they are often named and organized as a hierarchy. An example: "rec.arts.movies" where "rec" is the general hierarchy name, "arts" and then "movies" denote a gradual narrowing of the content subject-matter. Messages can be posted at all levels of the hierarchy based on the author's assessment of best fit. New levels in the hierarchy can be created by users as needed or desired.

Noise: Shannon and Weaver (1949) describe noise as random interference that corrupts a message while it is being transmitted between a sender and receiver. While noise exists as a physical phenomena in any form of energy (e.g., audio, electromagnetic, heat), the term has been used in a social and psychological sense to denote any exogenous influence that corrupts messages in transmission. Noise can interact with endogenous deficiencies in *interpretive repertoire* to cause messages to be misinterpreted.

Objects of Cognition: "Objects of cognition" is a term introduced by Woelfel and Fink (1980) and used in this study to denote any idea, image or concept present in the mind or stored in memory.

Organizational Associations: Brown and Dacin (1997) note that people can link different thoughts with organizations, brands and products. It is apparent that these are levels in a hierarchy: organizations can have many brands while many products can bear the same branding. It is important to distinguish between the thoughts people link to the entities at the three levels because they may be very different. For example, a person may have a positive attitude toward one product within a brand while having a negative attitude toward the brand or organization as a whole.

Outdegree: Wasserman and Faust (1994) describe outdegree as a term from network science that refers to the number of ties an entity has initiated with other entities. For example, if that entity is a webpage that has five (5) links to other webpages, then the outdegree of that webpage is five (5).

Outliers. Chiu et al (2001) describe outliers as objects located far away from any natural object clusters. Many clustering algorithms will assign outliers to the nearest cluster but their distance to the cluster *centroid* is significantly greater than that of the other objects in the cluster.

Personality: Acker (1997) describes brand personality as a type of *cognitive association* that assigns human characteristics, and even identity, to a brand.

Post: Synonymous with *entry*, already described.

Power Law: A power-law is a class of statistical distributions where small occurrences are common and large events are rare. One example of a power-law distribution is presented in Figure A-2. Some examples of power-law distributions are the Pareto, Zipf and Lognormal distributions.

Proportioning Factor: Bianchini et al (2005) note that when Google™ calculates a web page's PageRank™, it uses a portion of the PageRank™ of web pages that it links to. The linking page inherits some of the importance of pages it references. This idea can be transferred to social networks: your importance is partially based on the importance of the people you know. Google™ keeps secret the exact proportion it uses. This paper uses an arbitrary value of 15%.

Psychological Construct: Thorndike (1996) describes a construct as a measurable representation of a concept. Since psychological concepts, such as trust and love, are intangible they cannot be directly measured. Thorndike notes that to measure concepts they must be mapped to a measurement instrument like a series of Likert scale questions. What the measurement instrument really measures is the construct. Construct validity is the extent to which the construct matches the concept.

Public Good: Gintis (2004) describes a public good as a resource that is not reduced by being consumed. *Blogs* and *forums* are public goods because any number of people can read them without diminishing the value received by any specific reader. The problem is that public goods are not costlessly created and it is difficult to get people to incur a cost for something that *free riders* can have for free. However, without some people being willing to incur the cost, the public good cannot be created and no one gets any value. *Blogs* and *forums* are started in the hope that as individuals benefit, enough of them will contribute sufficient intellectual capital to ensure a positive return to everyone regardless of the presence of *free riders*.

Quasi-experimental: Campbell and Stanley (1963) note that experiments are critical to the empirical approach of the scientific method, aimed at supporting or falsifying hypotheses while controlling for extraneous (unrelated to the manipulated behavior) variation. However, some research questions require investigation in natural contexts where extraneous variation cannot be controlled, nor can the behavior of a random selection of subjects be manipulated while withholding such manipulation from others (i.e., a control group). Such contexts are termed quasi-experimental. Since extraneous variation cannot be controlled, it must be identified and its effects determined as completely as possible.

Reciprocity: Gintis (2004) describes reciprocity as an in-kind response to the actions of others and notes it is a standard part of most ethical systems. This study builds on Pereira et al (2006) that found that the practice of reciprocity can be expected as long as the price is not too high.

Recursive: Merriam-Webster (2007) describes recursion as a mathematical technique where a function is defined in terms of itself. In this study the value of a blog entry depends on the value of comments that are similar to it, the value of these comments also depend on the comments similar to them, and so on.

Relationship: Fournier (1998) noted that relationship is one type of *cognitive association* people often have with *brands*. The term implies some degree of co-dependence and often involves the *anthropomorphication* (humanizing) of the brand.

Reliability: Hair et al (2005) define reliability as a "measure of the degree to which a set of indicators of a latent construct is internally consistent in their measurements. The indicators of highly reliable constructs are highly interrelated, indicating that they all seem to measure the same thing." (p. 710)

Replicator: Blackmore (2001) describes the term "replicator" within the context of *memetics* as the means whereby *memes* are copied and dispersed. Blackmore suggests that humanity's main evolutionary purpose is the propagation of cultural memes.

Reputation: Gioia et al (2000) describe a *brand's* reputation as the long term *cognitive associations* held by consumers. Reputation is widely considered to be the most fundamental instrument of social control. If a brand has had a good reputation for a long time, it has probably acquired substantial *brand equity* as a result, the company that owns the brand will naturally be reluctant to squander its reputation for the sake of short term gain.

Resonance: Corman et al (2002) introduced the term "resonance" as one of the metrics associated with *Centering Resonance Analysis*. Resonance is the degree of similarity between two bodies of text based on the extent to which they both use the same words with similar *influence*. Resonance therefore measures the cognitive distance

between two documents. Resonance can be used as a Euclidean distance in a cognitive space.

Science of Complexity: Phelan (2001) differentiated between traditional and the science of complexity: "Traditional science seeks direct causal relations between elements in the universe whereas complexity theory drops down a level to explain the rules that govern the interactions between lower-order elements that in aggregate create emergent properties in higher-level systems" (p. 8). See the description of *generative rules* discussed later.

Self-organizing Map: Kohonen (1990) introduced the self-organizing map (SOM) as a way of visualizing a *network model* where the links between objects have a desired length. If the network model is metaphorically shaken every time a new object is added to the network, the objects will jostle around under the influence of the links that are trying to assume their desired length until the objects assume the position relative to the other objects that they should have. The dimensionality of the space is an important factor in the self-organizing process. The space has to have enough dimensions to allow the objects to move freely enough under the influence of the links to assume their proper positions. However, it is desirable to have the minimum number of dimensions possible to keep the space from becoming too sparse.

Sensemaking: Rosa et al (1999) found that when consumers are introduced to a new product category they often go through a process of trying to determine what value the category can offer them. In our interconnected world this process involves an indirect conversation through the media and a more direct conversation through the Internet between consumers and producers. The producers wish to clarify the nature of the intended value offering and consumers want to assess whether one of their needs has really been met. This process of clarification is what this study calls sensemaking.

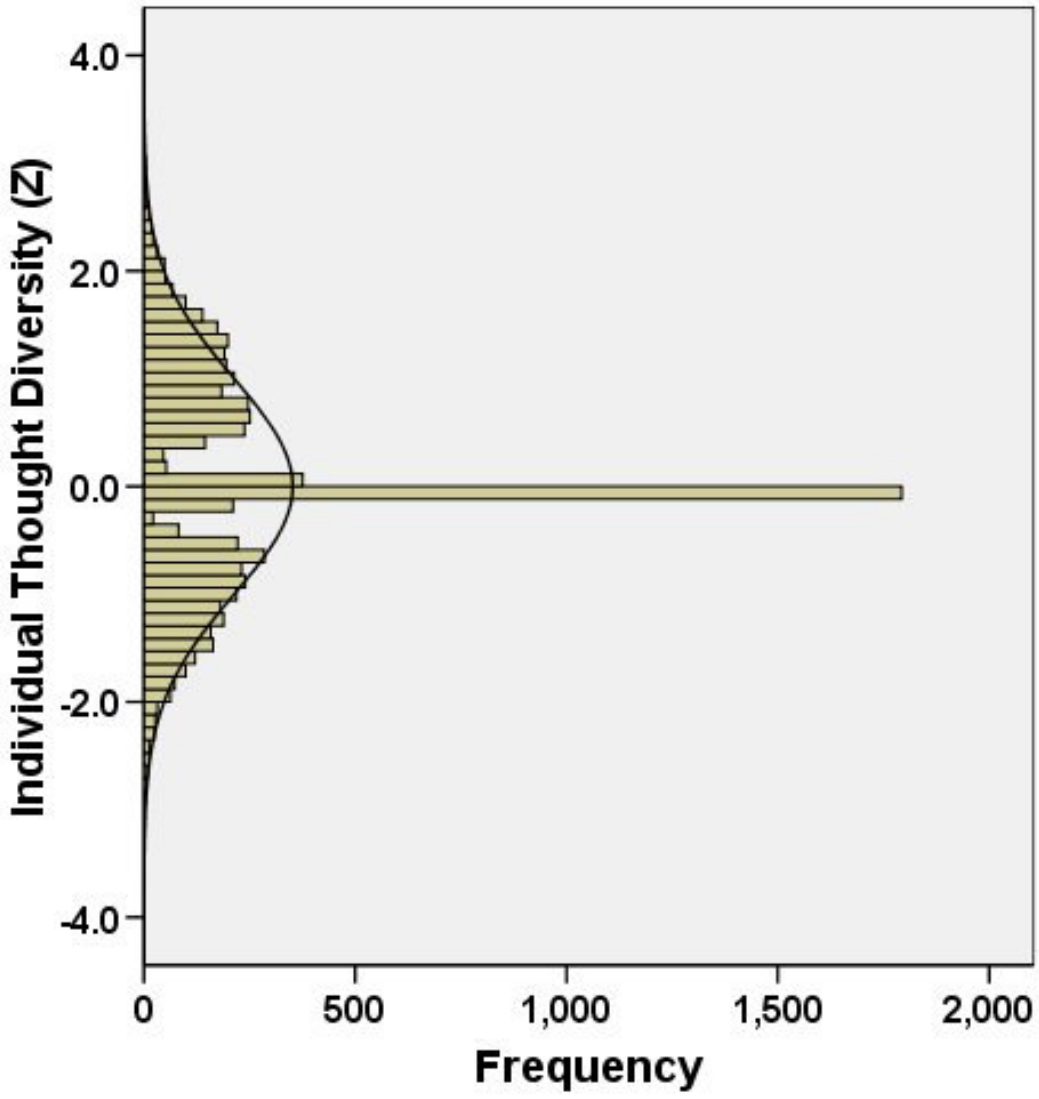
Separation: One of Reynolds' (1987) three simple "steering behaviors" that characterize flocking theory. Separation refers to the behavior of steering to avoid crowding local flock mates.

Social Attractor: The term "social attractor" is widely used in the sociology literature in a variety of ways, although it is seldom explicitly defined. The most generic definition seems to be Manzini's (1994) description of something that "orients the choices of a multiplicity of individuals" (p. 43). A social attractor is an attractor in the same sense as a *fixed point attractor*. Here it is used to describe a phenomenon observed in this study where the mean point in the spectrum of *individual thought diversity* draws people to an unexpected consistency. In Figure A-3, people near the mean seem drawn together into a commonality in the extent to which they will maintain a consistency in their thought expression. It might be expected that the consistency with which people express thought would be normally distributed, some few always repeating the same thoughts (i.e., *evangelists*), some few expressing a very wide diversity of thoughts (*idea seeders*),

while the majority fit neatly under the normal curve, in the middle. However, this study observes an unexpected attraction to the mean that distorts the normal distribution. This is discussed in greater depth in Chapter VI.

FIGURE A-3

A Social Attractor in Individual Thought Diversity



Social Forms: Social forms are synonymous with *generative processes*, already described.

Social Process Theory: Cederman (2005) noted that the physical sciences have many universal laws that explain observed behavior among tangible objects. Cederman also noted that the social sciences are still looking for equivalent general theories that explain large segments of observed phenomena. Cederman described Social Process Theory as a promising candidate for such a general theory as it explains the *complexity* of social phenomena by saying it is the *emergent* result of interacting *generative processes*.

Thematic Cluster: This study uses *self-organizing maps* to locate cognitive objects in proper proximity to each other. Once the proper locations are found, the next step is a cluster discovery phase where it is determined whether the cognitive objects form *clusters* that indicate commonality in thematic content. The number of such clusters and their size are indications of *cognitive diversity*.

Theoretical: Summers (2001) described three kinds of scientific contribution that can be made: theoretical, *methodological* and *empirical*. A theoretical contribution can be a new explanation for relationships between variables, a new insight into the boundaries of an existing theory or a better explanation for a relationship explained by existing theory.

Threshold Model: Granovetter (1978) proposed his *Threshold Model of Collective Behavior* to explain why crowds will often behave in ways the individuals in the crowd never would on their own. He describes how for many behaviors the threshold of individual internal resistance is normally distributed throughout a population. If the right combination of individual thresholds is present in a crowd then behavior started by one person can propagate through a crowd as the growing number of participants cause thresholds of resistance to be overcome in a cascading fashion.

Trackbacks: Wikipedia (2007) notes that the entries posted by blog authors are often referenced, through hyperlinks, in the work of other blog authors and online news articles. These links, much like journal citations, are generally regarded as an indication of the quality of the referenced blog entry. Such links to a blog entry are called trackbacks.

Unconstrained Free Response: Unconstrained free response is closely connected to the idea of *free elicitation*, already defined, where people are allowed to speak as much, or as little, as the want in response to a question or "stimulus probe cue."

User-generated Media: Wikipedia (2007) indicates that the term "user-generated media" is used interchangeably with the terms "user-generated content" and "consumer-generated media." Wikipedia also indicates that the three synonym phrases refer to online content that is produced by people who were hitherto assumed to be only users or

consumers of online content. The phenomenon reflects the availability of affordable new tools for authoring content that can be easily disseminated through the Internet. Such content includes *blogs*, podcasts (an audio file, usually containing commentary or entertainment content), video, cellular phone photos, word-of-mouth and wikis (a website that allows readers to edit the content and thus be a tool for collaborative authoring).

Virtual Community: Wikipedia (2007) describes a virtual community as a group of people who primarily interact via the Internet. Such communities are often very topic oriented and therefore form around *blogs* and *forums* that the community members like to read.

Weblog: Wikipedia (2007) describes a weblog, or *blog*, as a website where one or more regular authors initiate discussion on a topic of their choosing. The website allows *comments* to be added to the end of the blog author's *entry* thus allowing a two-way conversation between author and reader and a many-to-many conversation among the readers. A weblog is usually distinguished from a *forum* by the existence of the blog entry which is intended to control the subject of conversation.

Weblog Author: Scoble and Israel (2006) denote the person who writes the conversation-starting *entries* for a *blog* as the blog author. In a corporate *blog*, the author is often a high-level executive who is perceived by the readers as someone who has the power to use the insights gained from the *blog* conversation.

Wikipedia: Wikipedia (2007) describes itself as a web-based encyclopedia whose knowledge content is created and edited collectively by the users. The content of Wikipedia is constantly in a state of flux as new information is added and existing information is made more correct or complete by readers of varying expertise. Wikipedia is a good example of *user-generated media* and an example of *emergence* in action.

APPENDIX B

LATENT VARIABLE INDICATORS

Mean Cluster Density: Mean Cluster Density is one of the overall indicators of diversity in thought expression. After the comment clusters for each blog entry are found, the distance from each comment's location in cognitive space to the centroid of its cluster can be calculated. The standard deviation of each cluster's comment-to-centroid distances is proposed as a measure of how dense the comments are in their cluster. A low standard deviation indicates that the comments are closely packed and diversity in thought expression is predicted to be low.

Mean Cluster Radius: Mean Cluster Radius is one of the overall indicators of diversity in thought expression. After the comment clusters for each blog entry are found, the distance from each comment's location in cognitive space to the centroid of its cluster can be calculated. The mean of each cluster's comment-to-centroid distances is proposed as a measure of how different the comments are in their cluster. A low mean indicates that the comments are located close to their cluster centroid and diversity in thought expression is predicted to be low.

Mean # Clusters Started: The mean number of clusters (i.e., thematic clusters) started over the participation lifetime of a blog commenter is one of the measures used to segment commenters into *idea seeder*, *evangelist* or *flock follower* categories. For each blog entry, the commenters to that entry are grouped by category. Then, the mean number of clusters started by each category group is one metric used to estimate the influence of *flocking*, *mass dissemination* and *idea seeding* on a blog entry. Blog entries with many commenters who have started a low number of clusters may be highly influenced by flocking. However, blog entries whose commenters have started a large number of clusters may be more influenced by mass dissemination or idea seeding than flocking.

Mean Collective Thought Separation: Mean Collective Thought Separation is used as an indicator that separates contexts characterized by *cultural tribalism* from those characterized by *need-for-cognition*. It is the mean of entry-to-comment and comment-to-comment cognitive distances for the current blog entry. Thus, the typical cognitive distance between units of communication (i.e., comments and entry) is measured. It is proposed that a low mean is indicative of little diversity in thought expression and, moreover, a low tolerance for such diversity. This is one of the predicted attributes of a context where cultural tribalism is taking place.

Mean Commenter Longevity: Mean Commenter Longevity is used as an indicator of *cultural tribalism* as it is the mean time between the first comment and the current comment of all commenters to the current blog entry. It thus measures the mean lifespan of the active commenters to the blog.

Mean Entry-to-Entry Separation: Mean Entry-to-Entry Separation is used as an indicator of *indoctrination* in blog entries and as a means of separating indoctrination from non-indoctrination, or *free thinking*, blogs. It is calculated as the mean pair-wise cognitive distances between entries to the same blog. It is predicted that if a blog author is attempting to use indoctrination, the intense repetition of the same theme or a limited set of themes, in a blog then the mean cognitive distance between entries will be smaller than normal.

Mean Individual Thought Separation: Mean Individual Thought Expression is used as an indicator whose level distinguishes *evangelists* from *idea seeders* and contexts characterized by *cultural tribalism* from those characterized by *need-for-cognition*. This metric is the mean comment-to-comment cognitive distance, across blog entries, between comments made by the same individual, a commenter on the current blog entry. Thus, the metric measures how far, in cognitive distance, an individual commenter typically strays from a favorite theme. It is proposed that a low mean is indicative of little willingness to express diversity in thought expression, one of the predicted attributes of a context where *cultural tribalism* is taking place.

Mean # Comments: The mean number of comments (across blog entries) created by individuals commenting on the current blog entry is proposed to be an indicator that separates idea seeders and evangelists from the other commenters. Idea seeders and evangelists are predicted to more prolific than flock followers.

Mean Time between Entries: Mean Time between Entries is used as an indicator of *indoctrination* in blog entries and as a means of separating indoctrination from non-indoctrination, or *free thinking*, blogs. It is calculated as the mean interval between successive entries posted to a blog. It is proposed that the authors of blogs who are employing indoctrination will post entries more often than other authors, whether motivated by zeal or an intention to increase the intensity of the indoctrination effect.

Number of Clusters: Number of Clusters is one of the overall indicators of diversity in thought expression. After the comment clusters for each blog entry are found, the number of clusters is proposed to be indicative of the number of conversational themes. It is proposed that the more clusters, the greater the diversity in thought expression.

Percentage of Outliers: Percentage of Outliers is one of the overall indicators of diversity in thought expression. After the comment clusters for each blog entry are found, the distance from each comment's location in cognitive space to the centroid of its cluster can be calculated. The mean and standard deviation of each cluster's

comment-to-centroid distances is calculated and a z-score calculated for each comment. Any comment with a z-score greater than 2.0 is considered an outlier. It is proposed that the more outliers, the greater the diversity in thought expression.

% First Time Commenters: The number of commenters contributing to the current blog entry who have never contributed to previous entries is predicted to have an expansive effect on the overall diversity of thought expression in a blog. This metric is proposed to be an indication of *need-for-cognition*.

% Repeat Commenters (T-1): The number of commenters contributing to the current blog entry who contributed to the previous blog entry are predicted to impose a limiting effect on the overall diversity of thought expression in a blog. This metric is proposed to be an indication of *cultural tribalism*.

Stdev Collective Thought Separation: Stdev (i.e., standard deviation) Collective Thought Separation is used as an indicator that separates contexts characterized by *cultural tribalism* from those characterized by *need-for-cognition*. It is the standard deviation of entry-to-comment and comment-to-comment cognitive distances for the current blog entry. Thus, this metric measures the typical cognitive variation between units of communication (i.e., comments and entry). It is proposed that a low variation is indicative of low diversity in expressed thought, and perhaps a low tolerance for such diversity. This is one of the predicted attributes of a context where cultural tribalism is taking place.

Stdev Entry-to-Entry Separation: Stdev (i.e., standard deviation) Entry-to-Entry Separation is used as an indicator of *indoctrination* in blog entries and as a means of separating indoctrination from non-indoctrination, or *free thinking*, blogs. It is calculated as the standard deviation in the pair-wise cognitive distances between entries to the same blog. It is predicted that if a blog author is attempting to use indoctrination, the intense repetition of the same theme or a limited set of themes, in a blog then the standard deviation in cognitive distance between entries will be smaller than normal.

Stdev Individual Comment-to-Cluster Separation: The standard deviation of the cognitive distance between a commenter's comments and the centroids of the closest thematic cluster across blog entries over the participation lifetime of a blog commenter is one of the measures used to segment commenters into *idea seeder*, *evangelist* or *flock follower* categories. For each blog entry, the commenters to that entry are grouped by category. Then, the standard deviation of the comment-to-cluster separation for each commenter in each category group is calculated as one metric used to estimate the influence of *flocking*, *mass dissemination* and *idea seeding* on a blog entry. Blog entries with commenters who have low variation in comment-to-cluster separation may be highly influenced by flocking; as such commenters may be exercising deliberate control over the extent to which their expressed thought differs from that of others. However,

blog entries whose commenters have a high variation in comment-to-cluster separation may be more influenced by mass dissemination or idea seeding than flocking.

Stdev Individual Thought Separation: Stdev (i.e., standard deviation) Individual Thought Expression is used as an indicator that separates contexts characterized by *cultural tribalism* from those characterized by *need-for-cognition*. It is the standard deviation of entry-to-comment and comment-to-comment cognitive distances for the current blog entry. Thus, the typical variation in cognitive distances between units of communication (i.e., comments and entry) is measured. It is proposed that a low standard deviation is indicative of little diversity in thought expression and, moreover, a low tolerance for such diversity. This is one of the predicted attributes of a context where cultural tribalism is taking place**.**

Total # Commenters: The total number of unique commenters to a blog entry is hypothesized to determine the probability of there being an *idea seeder* among the commenters.

Total Value: Total value is a metric derived from Google's PageRank that is proposed to measure the value of the content in the blog entry and its associated comments. The lagged measure of this value (i.e., the total value of the previous blog entry) is proposed to indicate reciprocation activity among commenters who feel compelled to contribute content in return for the value they have received from the blog author and the other commenters.

APPENDIX C

BLOG CORRELATION MATRICES, NORMAL PROBABILITY PLOTS AND FULL MODELS

Blog Correlation Matrices

TABLE C-1

Correlation and Φ-matrix of Primary Constructs for AutoBlog

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.109	-					
3	Flocking	0.221	0.669	-				
4	IdeaSeed	0.023	0.005	0.045	-			
5	Indoc	-0.044	0.264	0.038	-0.022	-		
6	MassDiss	-0.021	0.284	0.153	0.000	0.009	-	
7	NFC	0.722	0.631	0.613	0.141	-0.238	0.282	-
8	Recip	0.235	0.078	0.046	-0.024	0.149	0.005	0.035

*Note: Correlations are corrected for attenuation. All values are significant to $\rho < .05$. CogDiv = Cognitive Diversity, CultTrib = Cultural Tribalism, IdeaSeed = Idea Seeding, Indoc = Indoctrination, MassDiss = Mass Dissemination, NFC = Need-for-Cognition, Recip = Reciprocity.

TABLE C-2

Correlation and Φ-matrix of Primary Constructs for Blog for America

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	-0.273	-					
3	Flocking	0.096	0.398	-				
4	IdeaSeed	0.038	-0.021	0.040	-			
5	Indoc	-0.087	0.242	0.012	-0.005	-		
6	MassDiss	0.033	0.116	0.077	0.001	0.005	-	
7	NFC	0.536	0.277	0.472	0.183	-0.245	0.187	-
8	Recip	0.261	0.052	0.026	-0.010	0.077	-0.006	0.045

TABLE C-3

Correlation and Φ-matrix of Primary Constructs for Blog Maverick

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	-0.180	-					
3	Flocking	0.058	0.631	-				
4	IdeaSeed	0.010	0.014	0.020	-			
5	Indoc	-0.122	0.235	-0.025	0.013	-		
6	MassDiss	-0.030	0.314	0.224	0.014	0.002	-	
7	NFC	0.361	0.667	0.710	0.053	-0.294	0.364	-
8	Recip	0.154	0.017	0.009	0.009	0.087	0.001	-0.026

TABLE C-4

Correlation and Φ-matrix of Primary Constructs for The Consumerist

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.088	-					
3	Flocking	0.294	0.437	-				
4	IdeaSeed	0.073	0.008	0.100	-			
5	Indoc	-0.007	0.420	0.073	0.027	-		
6	MassDiss	0.002	0.166	0.091	0.000	0.019	-	
7	NFC	0.901	0.332	0.530	0.298	-0.069	0.129	-
8	Recip	0.221	0.055	0.031	-0.008	0.065	-0.004	0.040

TABLE C-5

Correlation and Φ-matrix of Primary Constructs for EnGadget

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.061	-					
3	Flocking	0.260	0.551	-				
4	IdeaSeed	0.057	0.023	0.051	-			
5	Indoc	-0.070	0.292	0.024	0.006	-		
6	MassDiss	-0.006	0.270	0.199	0.001	-0.011	-	
7	NFC	0.738	0.715	0.677	0.153	-0.086	0.321	-
8	Recip	0.164	0.034	0.008	0.027	0.093	0.003	0.036

TABLE C-6

Correlation and Φ-matrix of Primary Constructs for The Evangelical Outpost

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.189	-					
3	Flocking	0.394	0.467	-				
4	IdeaSeed	0.058	0.018	0.066	-			
5	Indoc	0.101	0.363	0.063	-0.014	-		
6	MassDiss	0.013	0.170	0.092	-0.002	-0.016	-	
7	NFC	0.995	0.492	0.626	0.232	-0.025	0.202	-
8	Recip	0.297	0.125	0.094	-0.003	0.258	-0.023	0.179

TABLE C-7

Correlation and Φ-matrix of Primary Constructs for GM Fastlane

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.169	-					
3	Flocking	0.292	0.615	-				
4	IdeaSeed	0.047	-0.004	0.063	-			
5	Indoc	0.048	0.451	0.130	0.002	-		
6	MassDiss	-0.022	0.260	0.155	-0.011	-0.020	-	
7	NFC	0.838	0.638	0.640	0.200	-0.004	0.264	-
8	Recip	0.256	0.148	0.041	-0.005	0.266	-0.011	0.137

TABLE C-8

Correlation and Φ-matrix of Primary Constructs for Freakonomics

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.204	-					
3	Flocking	0.332	0.525	-				
4	IdeaSeed	0.058	0.017	0.070	-			
5	Indoc	0.041	0.282	0.026	-0.001	-		
6	MassDiss	-0.013	0.218	0.115	0.004	0.006	-	
7	NFC	0.929	0.557	0.607	0.230	-0.080	0.198	-
8	Recip	0.263	0.071	0.055	-0.015	0.172	0.000	0.090

TABLE C-9

Correlation and Φ-matrix of Primary Constructs for Gizmodo

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	-0.204	-					
3	Flocking	0.039	0.538	-				
4	IdeaSeed	0.025	0.012	0.022	-			
5	Indoc	-0.132	0.335	0.105	0.037	-		
6	MassDiss	-0.018	0.221	0.126	0.000	-0.018	-	
7	NFC	0.439	0.485	0.482	0.068	-0.183	0.309	-
8	Recip	0.191	0.053	0.001	0.009	0.100	-0.006	-0.010

TABLE C-10

Correlation and Φ-matrix of Primary Constructs for Google Blogoscoped

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.038	-					
3	Flocking	0.437	0.397	-				
4	IdeaSeed	0.113	0.005	0.124	-			
5	Indoc	0.013	0.380	0.060	0.000	-		
6	MassDiss	0.023	0.134	0.075	0.001	-0.001	-	
7	NFC	0.977	0.367	0.704	0.318	-0.005	0.114	-
8	Recip	0.328	0.074	0.091	-0.003	0.146	0.008	0.154

TABLE C-11

Correlation and Φ-matrix of Primary Constructs for Joystiq

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.388	-					
3	Flocking	0.460	0.537	-				
4	IdeaSeed	0.042	-0.031	0.067	-			
5	Indoc	0.041	0.362	0.097	0.028	-		
6	MassDiss	0.034	0.283	0.184	-0.016	-0.019	-	
7	NFC	0.993	0.664	0.699	0.213	-0.005	0.260	-
8	Recip	0.182	0.082	0.042	0.016	0.125	0.016	0.060

TABLE C-12

Correlation and Φ-matrix of Primary Constructs for PaulStamatiou

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	-0.033	-					
3	Flocking	0.283	0.365	-				
4	IdeaSeed	0.095	-0.001	0.086	-			
5	Indoc	0.043	0.492	0.098	-0.020	-		
6	MassDiss	0.018	0.101	0.062	-0.012	0.004	-	
7	NFC	0.891	0.198	0.518	0.346	-0.031	0.073	-
8	Recip	0.231	0.079	0.044	0.008	0.217	0.023	0.084

TABLE C-13

Correlation and Φ-matrix of Primary Constructs for Townhall

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.056	-					
3	Flocking	0.199	0.548	-				
4	IdeaSeed	0.092	-0.007	0.075	-			
5	Indoc	-0.032	0.242	-0.006	-0.023	-		
6	MassDiss	-0.017	0.224	0.101	0.009	-0.013	-	
7	NFC	0.727	0.590	0.576	0.269	-0.172	0.122	-
8	Recip	0.266	0.074	0.015	-0.018	0.157	0.010	0.052

TABLE C-14

Correlation and Φ-matrix of Primary Constructs for TV Squad

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.010	-					
3	Flocking	0.115	0.631	-				
4	IdeaSeed	0.003	-0.023	0.007	-			
5	Indoc	-0.073	0.244	0.036	0.010	-		
6	MassDiss	-0.024	0.281	0.187	-0.011	-0.030	-	
7	NFC	0.519	0.624	0.517	0.024	-0.261	0.298	-
8	Recip	0.201	0.065	0.009	-0.006	0.105	0.003	0.038

TABLE C-15

Correlation and Φ-matrix of Primary Constructs for The Unofficial Apple Weblog

		1	2	3	4	5	6	7
1	CogDiv	-						
2	CultTrib	0.071	-					
3	Flocking	0.274	0.713	-				
4	IdeaSeed	0.093	0.018	0.100	-			
5	Indoc	-0.113	0.341	0.086	0.007	-		
6	MassDiss	-0.033	0.322	0.162	-0.009	0.003	-	
7	NFC	0.767	0.739	0.761	0.247	-0.102	0.200	-
8	Recip	0.221	0.088	0.041	-0.010	0.151	-0.010	0.052

FIGURE C-1

Individual Blog Normal Probability Plots of the Cognitive Diversity Factor

(Expected versus Observed Cumulative Probability)

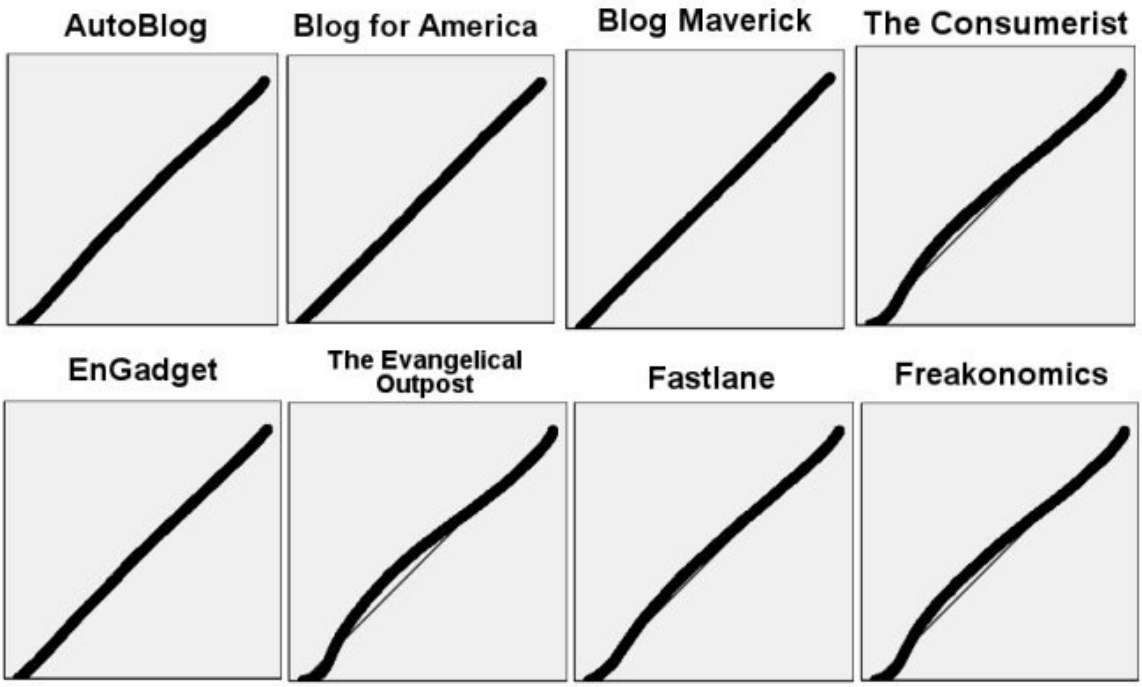


FIGURE C-1 (CONT)

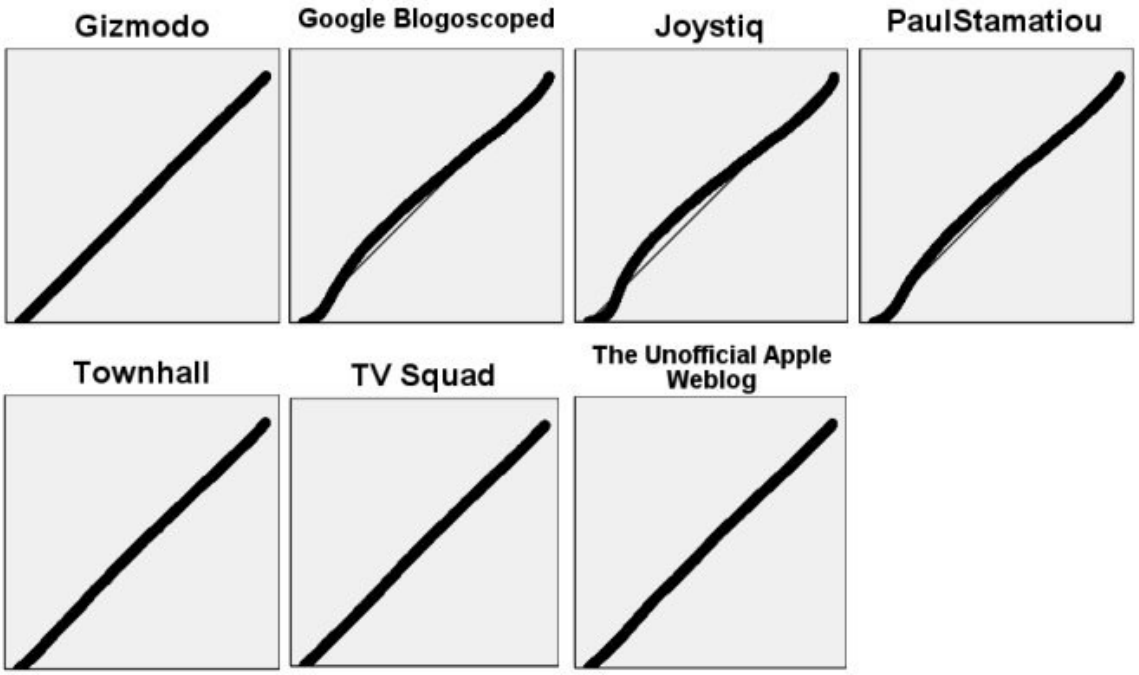


FIGURE C-2

Full Reflective and Formative Naïve Model

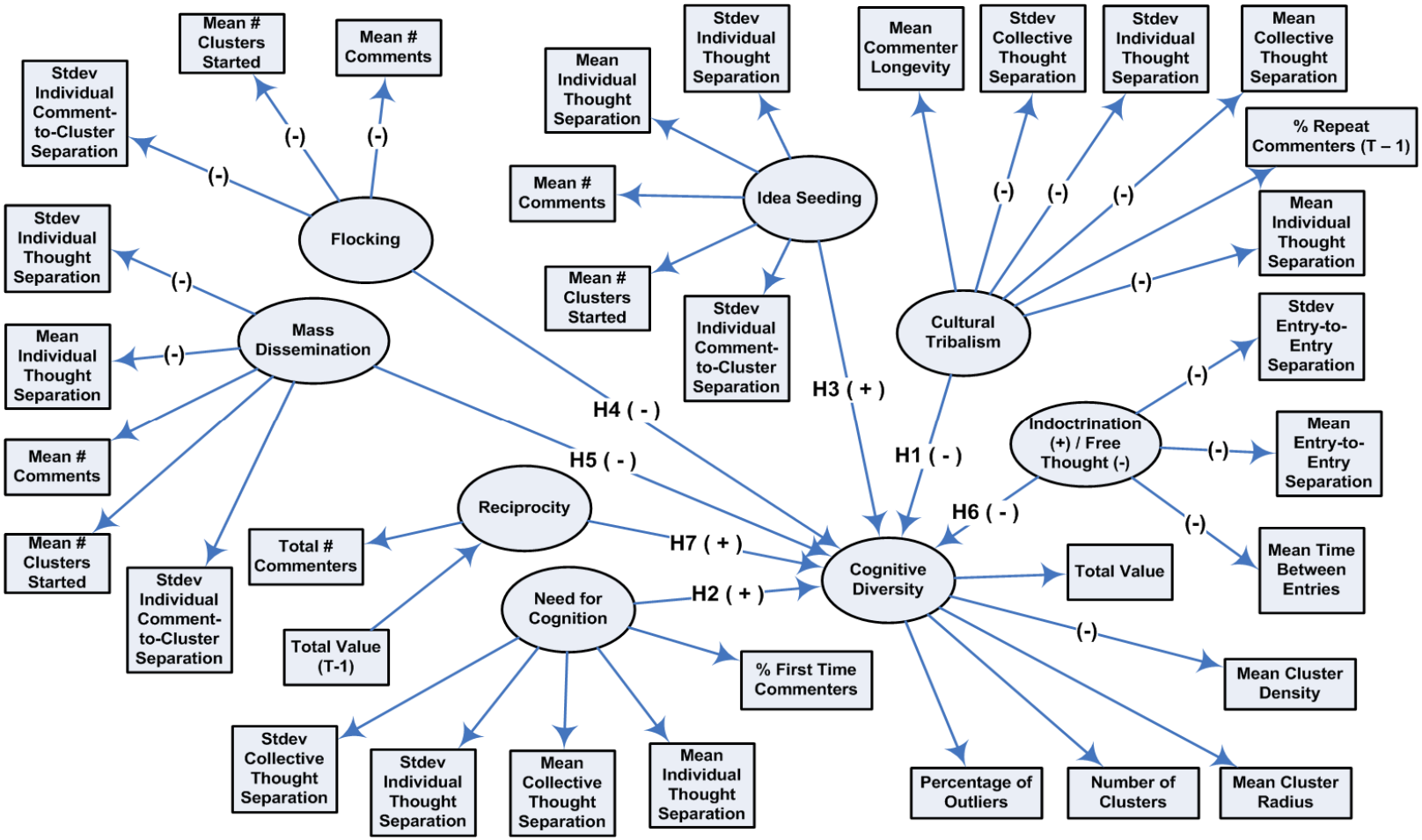
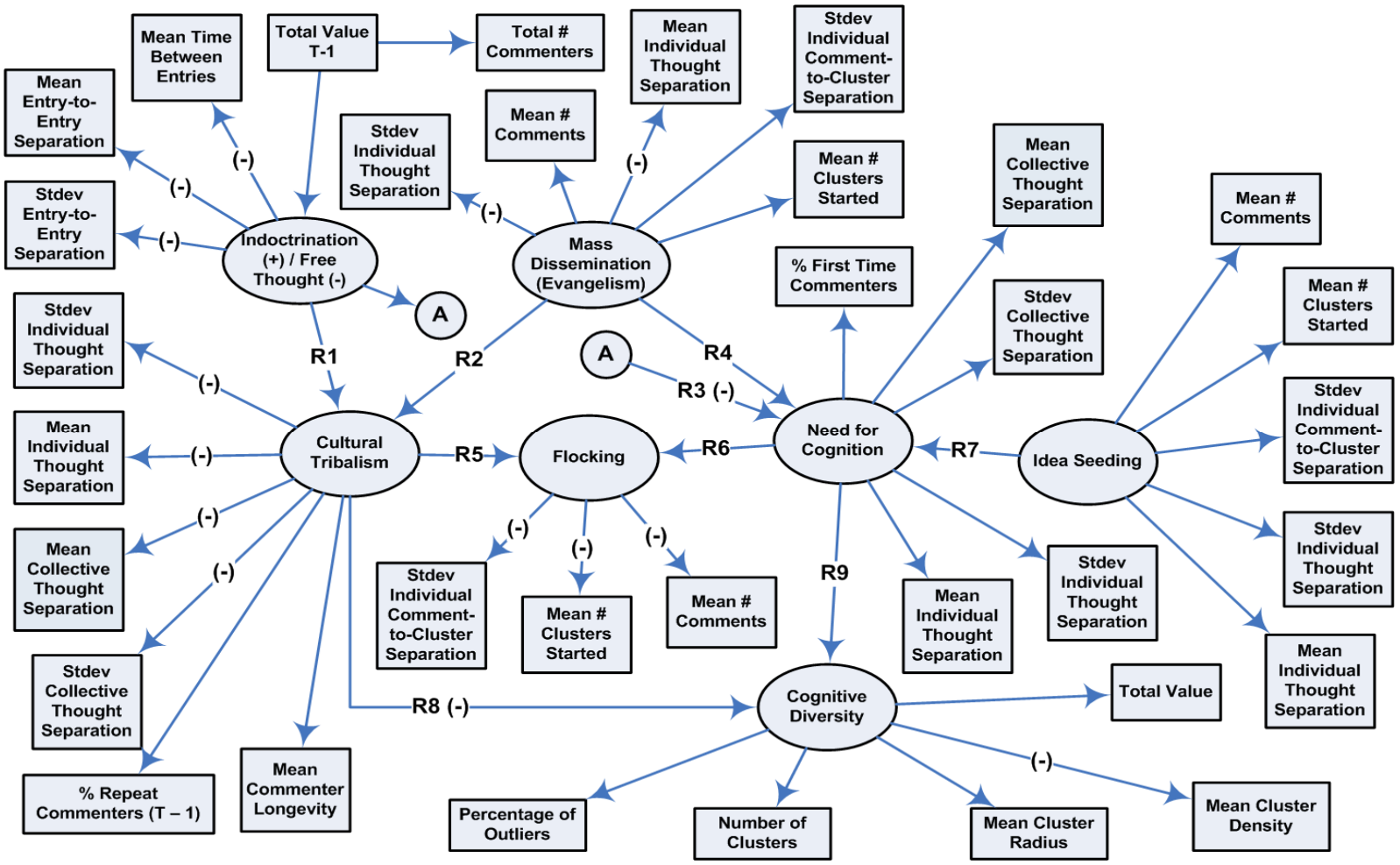


FIGURE C-3

Alternative Model of Potential Interrelationships



APPENDIX D

REFLECTIVE PARAMETERS FOR NAÏVE AND ALTERNATIVE MODELS

Naïve Model

FIGURE D-1

Naive Model Cognitive Diversity

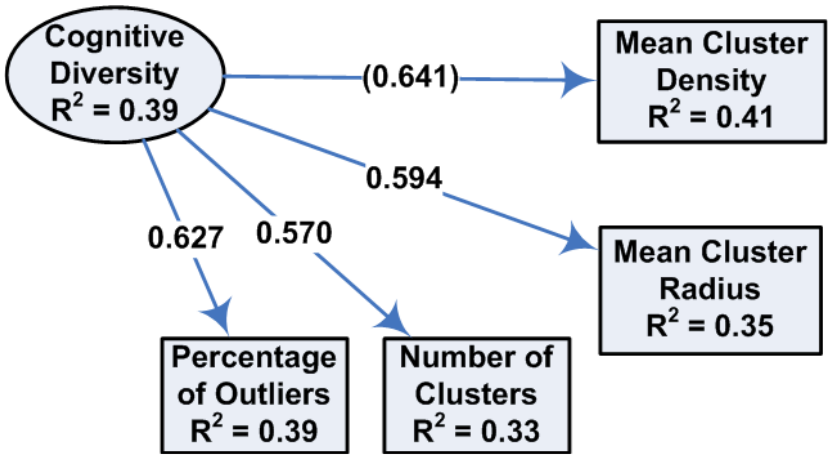


FIGURE D-2

Naive Model Indoctrination / Free Thought

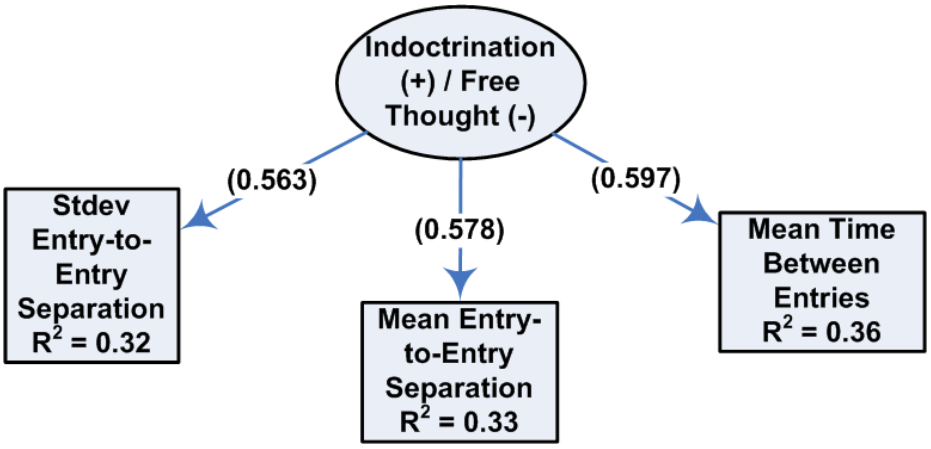


FIGURE D-3

Naive Model Cultural Tribalism

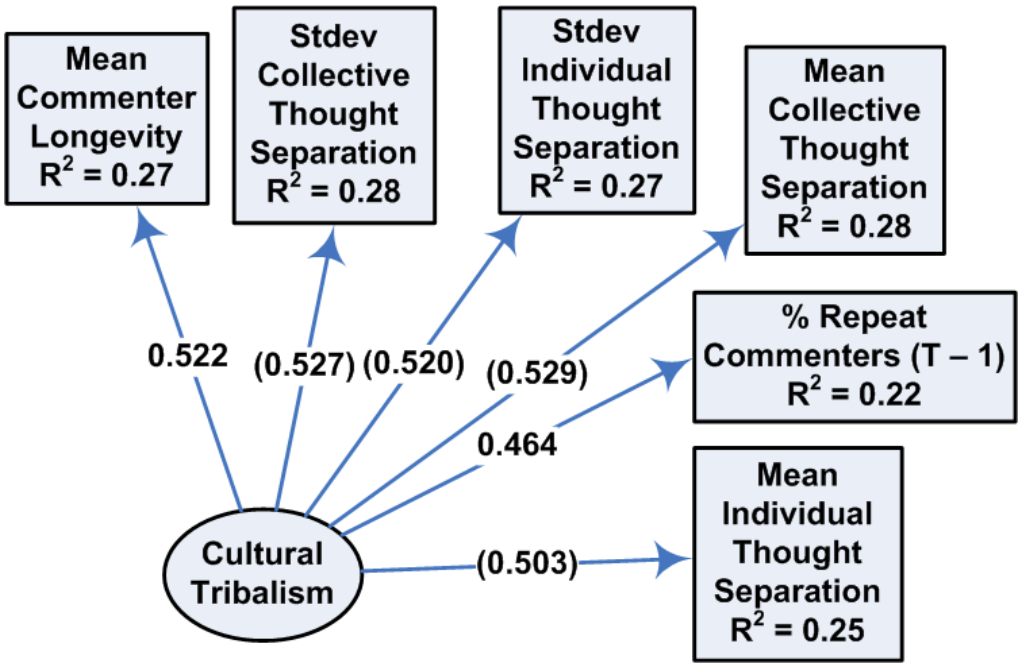


FIGURE D-4

Naive Model Idea Seeding

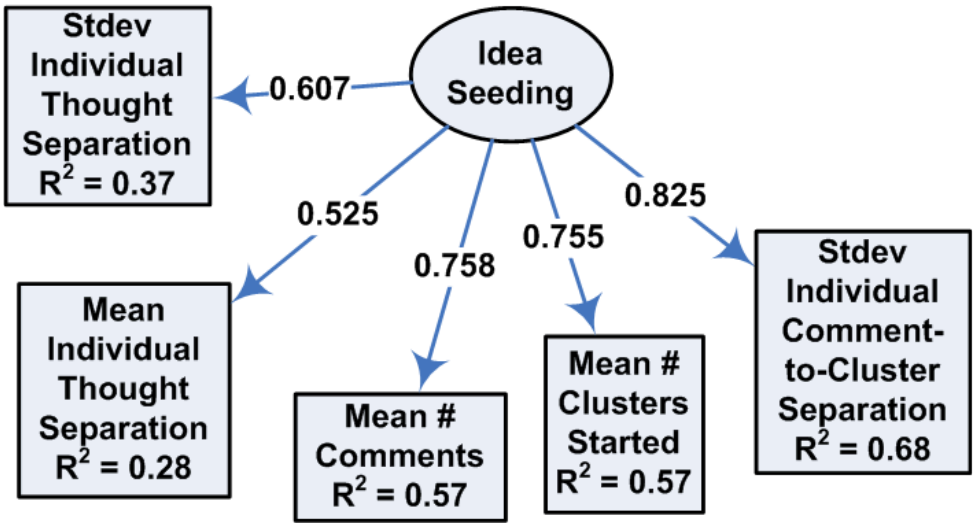


FIGURE D-5

Naive Model Flocking

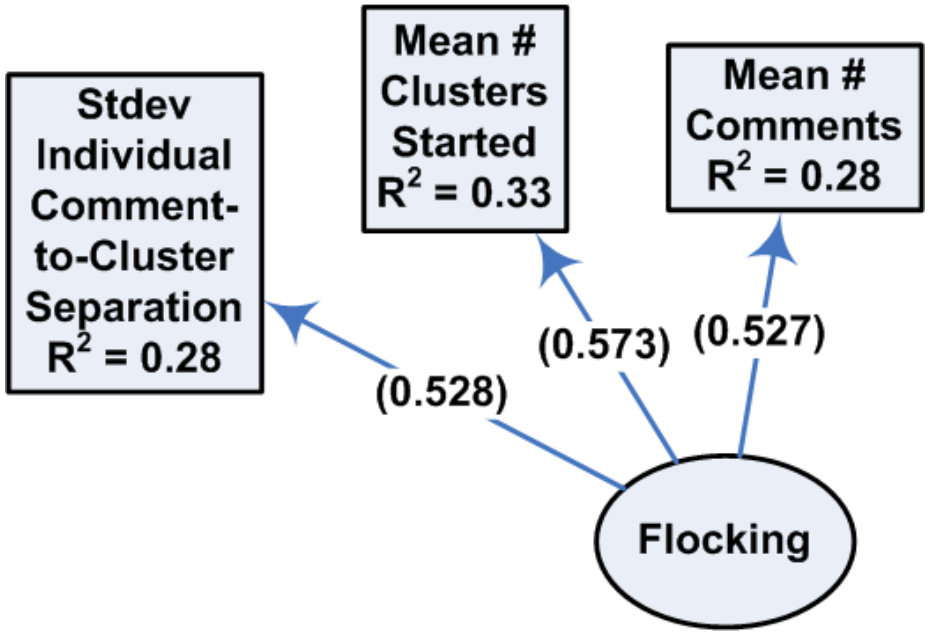


FIGURE D-6

Naive Model Mass Dissemination

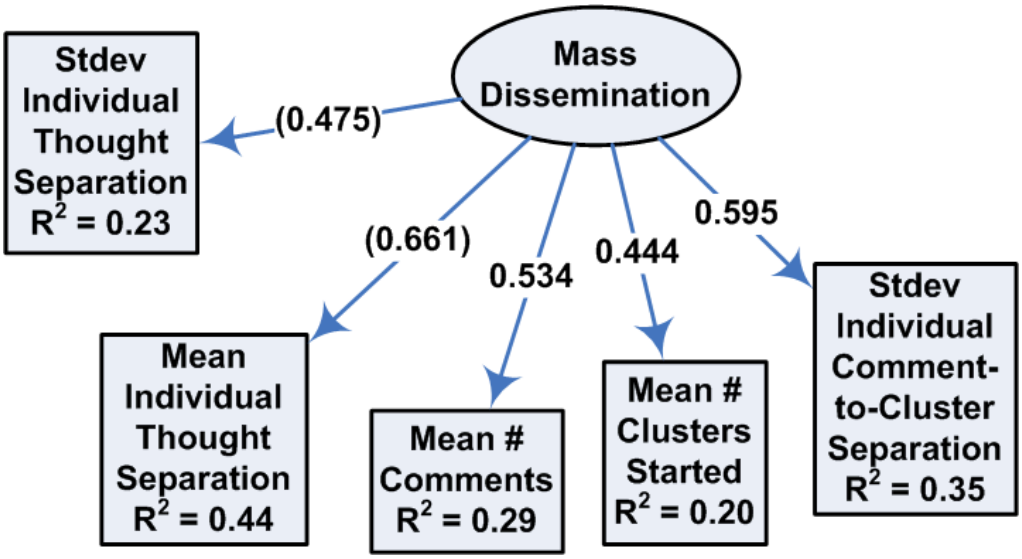


FIGURE D-7

Naive Model Reciprocation

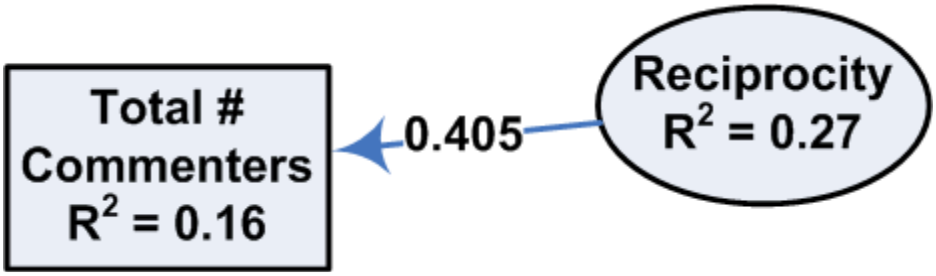
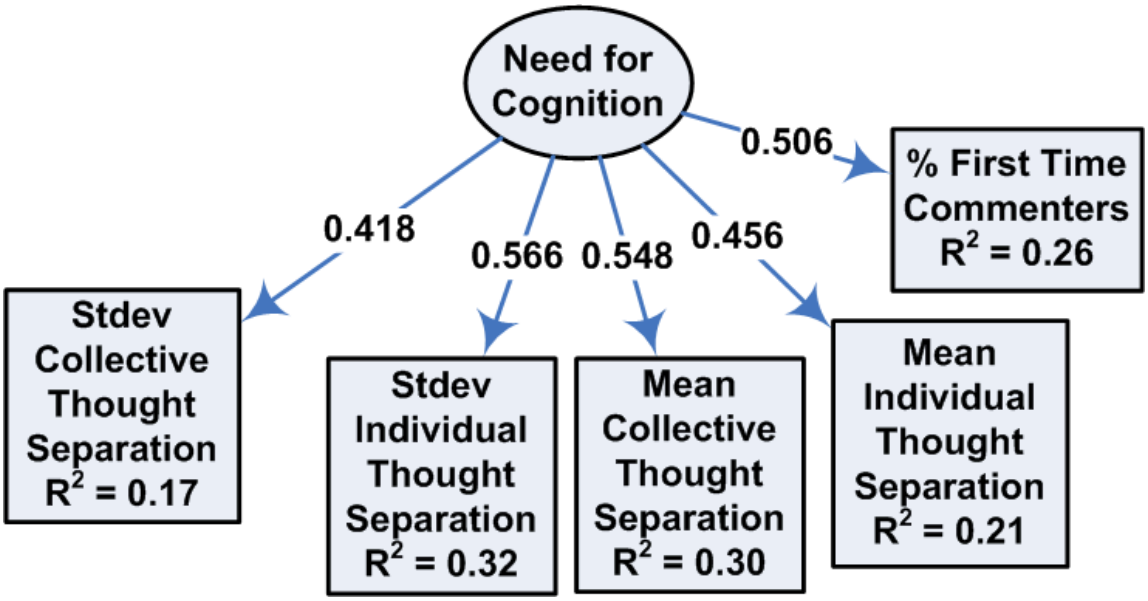


FIGURE D-8

Naive Model Need for Cognition



Alternative Model

FIGURE D-9

Alternative Model Cognitive Diversity

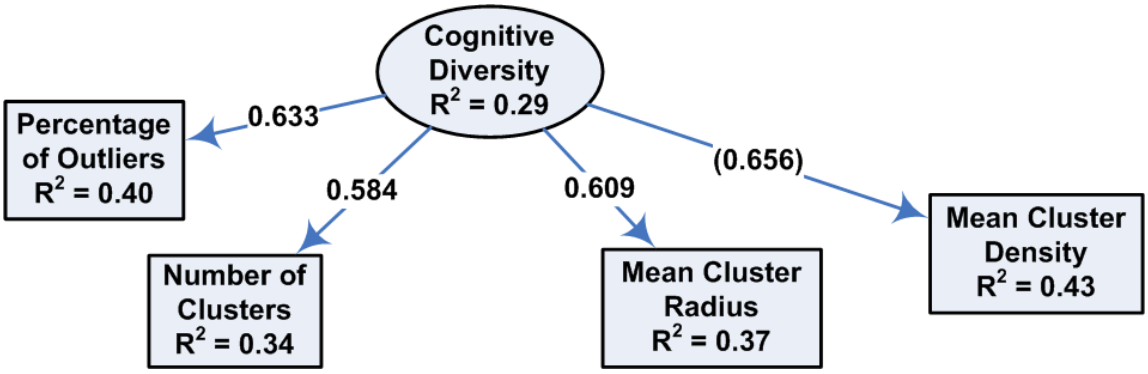


FIGURE D-10

Alternative Model Indoctrination / Free Thought

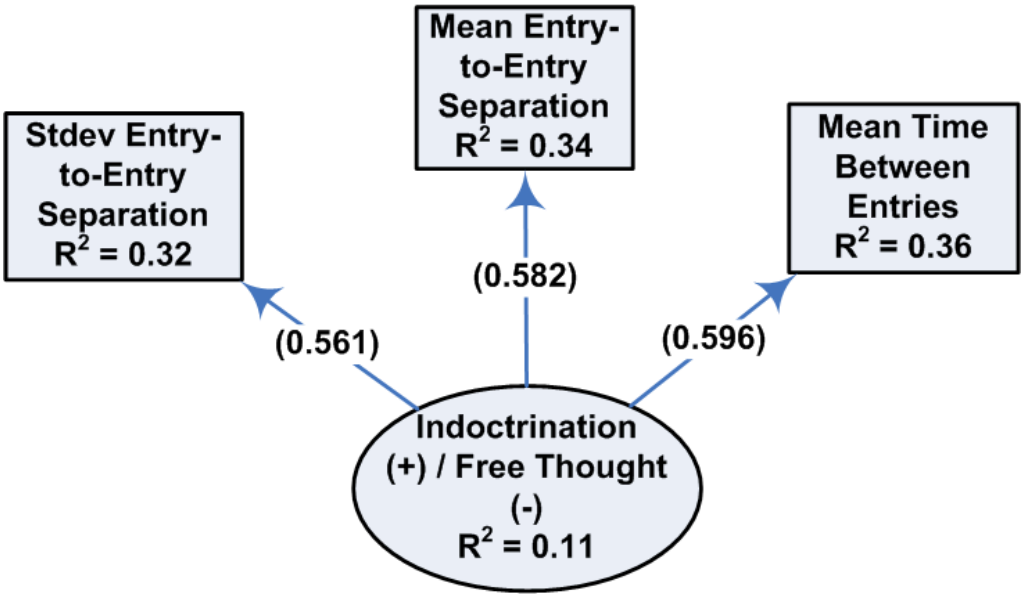


FIGURE D-11

Alternative Model Cultural Tribalism

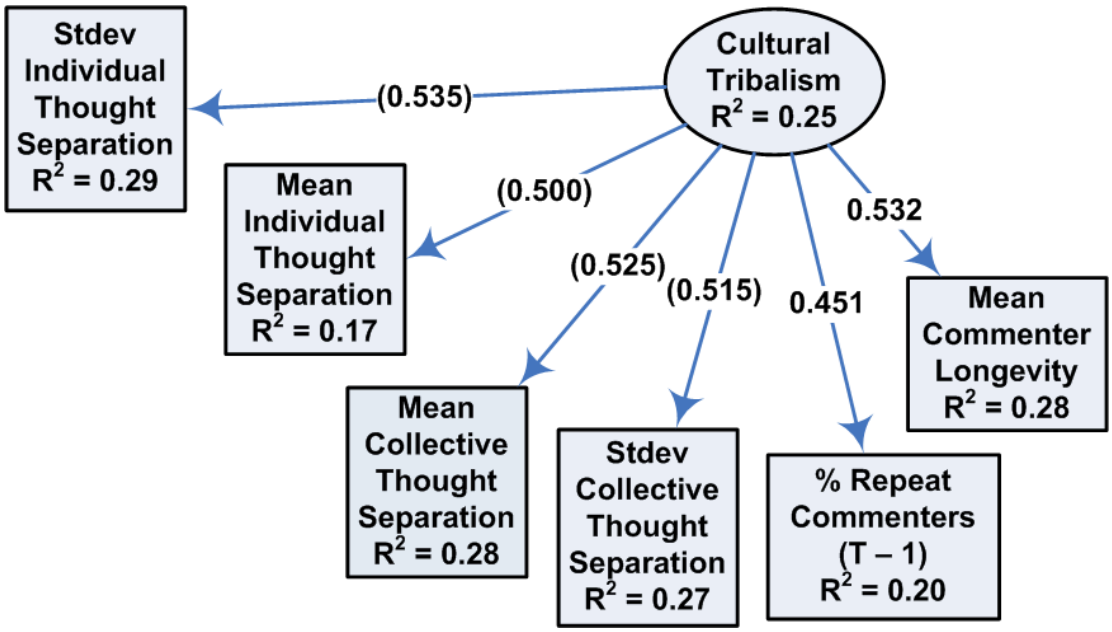


FIGURE D-12

Alternative Model Idea Seeding

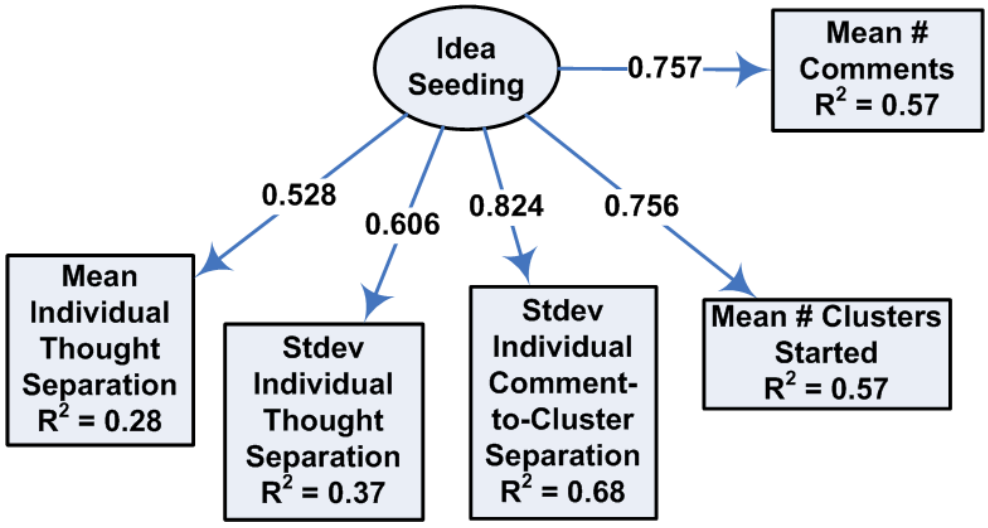


FIGURE D-13

Alternative Model Flocking

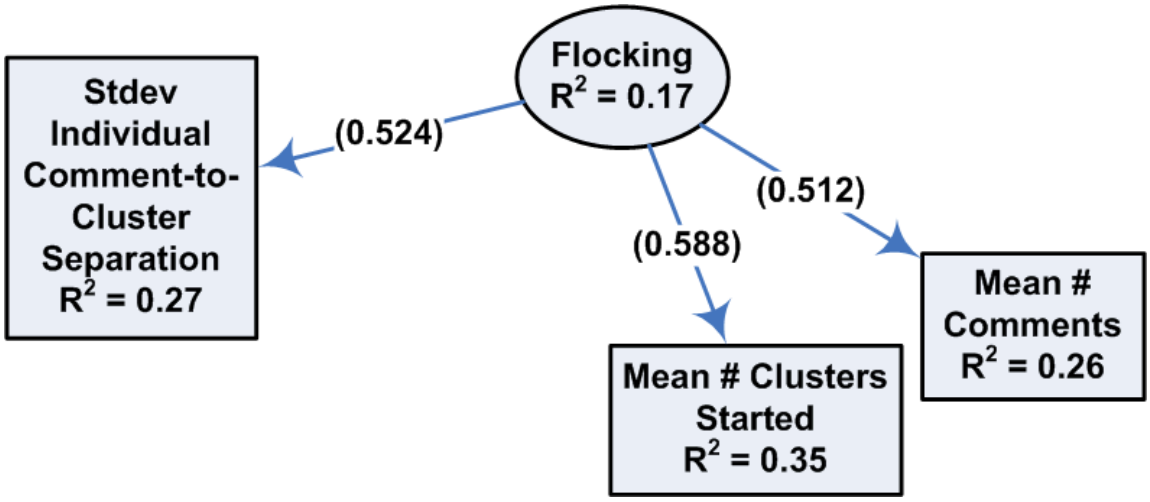


FIGURE D-14

Alternative Model Mass Dissemination

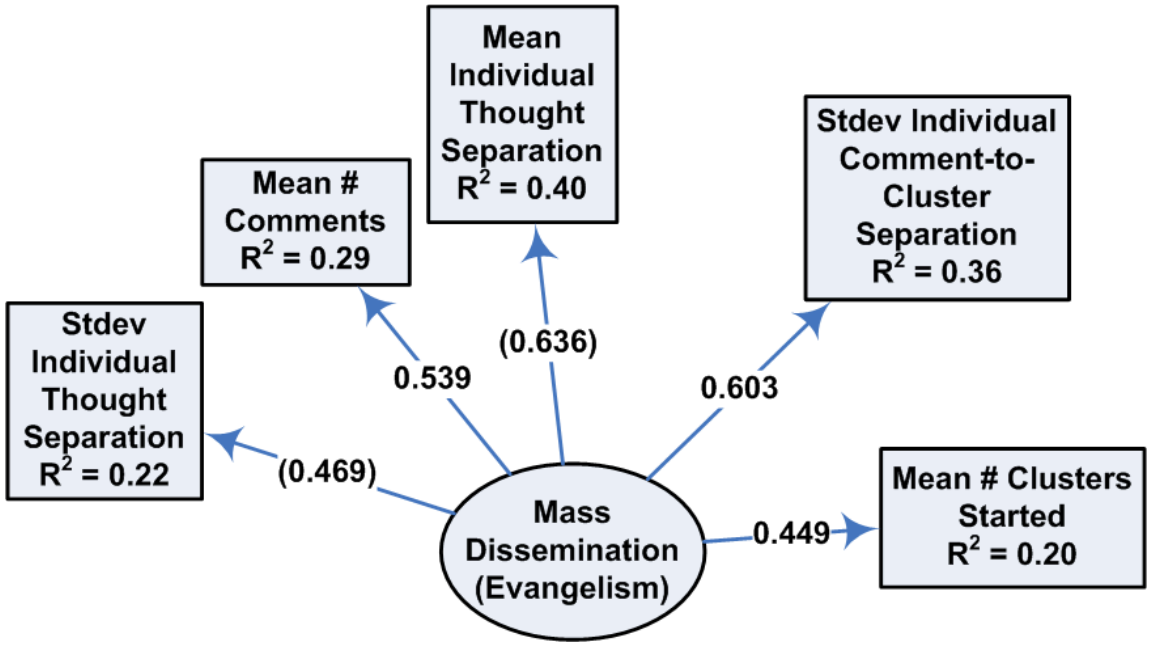
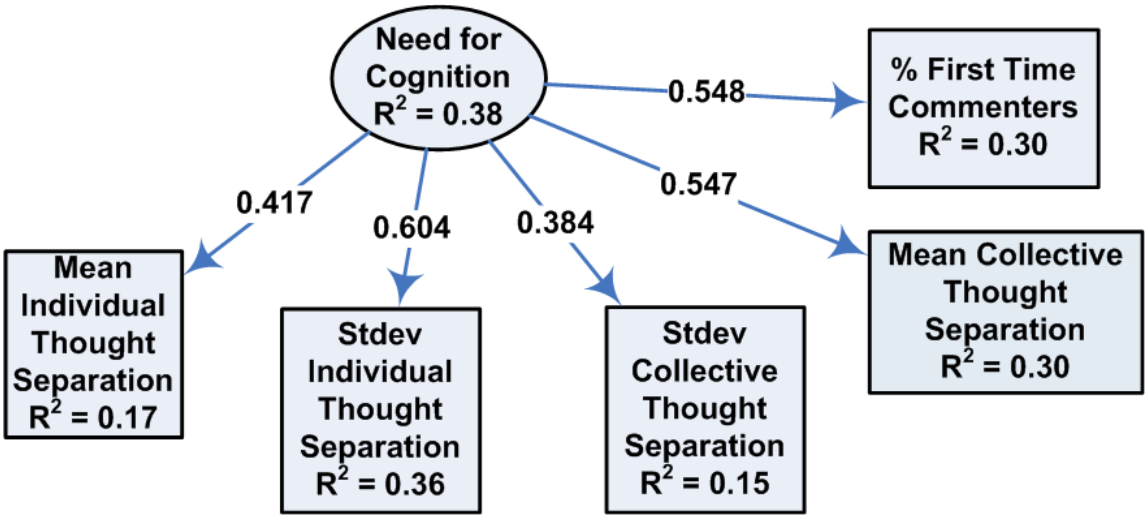


FIGURE D-15

Alternative Model Need for Cognition



VITA

Paul Dwyer received a Bachelor of Science in Computer Engineering from the University of South Florida in 1994, and a Master of Business Administration degree from Texas A&M University in 2004. Prior to entering Texas A&M, he was a senior software engineer at an Internet startup company.

His research generally focuses on consumer behavior as revealed in social networks, word-of-mouth, participation in virtual communities, and the managerial implications of that behavior. His work in this research stream has been published in the Marketing Science Institute's Working Paper Series and the Journal of Interactive Marketing, where he serves as an ad hoc referee. He has also presented his research at American Marketing Association conferences, the International Network Science Conference and the International Conference on Weblogs and Social Media. He was the recipient of the Emory Marketing Institute's Doctoral Dissertation Award.

Additionally, he is interested in cognitive modeling, the emergence of complex behavior in groups of individuals, computational modeling, data visualization and agent-based simulation. In this vein, he was a research scholar at the Santa Fe Institute's Complex Systems Summer School (2007).

Paul may be reached at Willamette University, Atkinson Graduate School of Management, 900 State Street, Salem, OR 97301. His email is pdwyer@willamette.edu.