

# TYING SOCIAL MEDIA TO ORGANIZATIONAL DECISION-MAKING

by

KERI McLEOD LARSON

(Under the Direction of Richard T. Watson)

## ABSTRACT

Embodying a new gestalt in firm-customer communication, social media are a nascent yet critical concern for researchers and practitioners alike. To date, very little research has accumulated in this area. The research community requires valid and reliable measures for social media effects in an organizational context, as do firms. Without such measures, firms remain unable to align their social media initiatives with organizational goals and ultimately create business value. This three-manuscript dissertation contributes a general framework for studying social media. Paper One presents a “social media ecosystem” model and focuses on the customer/firm segment entitled the “B@C Social Media Dyad” to provide a theoretical understanding of what firms and customers accomplish using social media. Paper Two further reviews the state of the art of textual analysis, a technique that can provide the deep level of qualitative analysis needed to fully ascertain important trends in firm/customer and customer/customer social media exchange, and concludes with the articulation of a set of design principles for developing a social media analytics system based on natural language processing capabilities. In Paper Three, the proposed approach is tested experimentally against sentiment analysis and manual approaches to mining knowledge from social media data and is demonstrated to provide superior support for

organizational decision-making through improved problem detection. Of particular consequence is that accuracy of problem and opportunity detection is far greater given an NLP-based approach, while sentiment analysis appears no more useful than randomly reading segments of social media data manually. These results support our recommendations for a more useful system for monitoring firm-level effects of social media. As a whole, this dissertation enlarges our meager theoretical understanding of the role social media play in an organizational context and presents the research community with a solid foundation for pursuing subsequent inquiries into a variety of social-mediated outcomes. Further, it contributes to IS research by offering an information system intended to solve the organizational dilemma of how to derive *meaningful* knowledge from social media exchanges.

INDEX WORDS: *social media; firm performance; measurement; collaboration; word of mouth; customer service; brand community; design science; experiment; natural language processing; statistical machine translation; machine learning; organizational decision-making*

TYING SOCIAL MEDIA TO ORGANIZATIONAL DECISION-MAKING

by

KERI McLEOD LARSON

BA, University of Chicago, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial  
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

© 2012

Keri McLeod Larson

All Rights Reserved



# TYING SOCIAL MEDIA TO ORGANIZATIONAL DECISION-MAKING

by

KERI McLEOD LARSON

Major Professor: Richard T. Watson

Committee: Dale L. Goodhue  
Marie-Claude Boudreau  
Nicholas Berente

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
August 2012

## ACKNOWLEDGEMENTS

I would like to express gratitude to my advisor, Dr. Rick Watson, for taking me on as a mentee and guiding me through the process of learning to become a scholar, and for sharing his perspective so that I might understand first-hand how one *should* go around thinking about the world. I'd like to thank Dr. Maric Boudreau for taking a chance on me as a co-author early in my doctoral career, giving me valuable experience with qualitative research and the pleasure of interacting with a very kind mentor. I'd also like to thank Dr. Nick Berente for co-authoring with me his first semester at UGA and steering me through the tribulations of conference publication. Opportunities to work with such outstanding faculty have been a real privilege.

I would like to acknowledge a debt of gratitude to Dr. Daniel Feldman for his role in enabling my career as an academic researcher. Without his wisdom, I would not have completed this journey.

I'd like to thank my parents for providing an environment during my formative years in which my academic pursuits were not only supported, but a given.

My husband, Tommy, deserves (and receives) my everlasting love and gratitude for caring for our family when I was not able to give my full attention to the things that really matter. I would have never contemplated this journey if not for his counsel and encouragement. Finally, thanks to my son, Soren, for nothing more than being incredibly awesome and always my top priority (even if I'm writing something important). I dedicate this dissertation to Tommy and Soren, my motivators to succeed in all things, big and small.

## TABLE OF CONTENTS

|   | Page |
|---|------|
| ACKNOWLEDGEMENTS.....   | iv   |
| LIST OF TABLES .....  | ix   |
| LIST OF FIGURES .....   | xii  |
| CHAPTER   |      |
| 1 INTRODUCTION AND LITERATURE REVIEW.....   | 1    |
| 1.1 Problem Statement and Motivation.....   | 1    |
| 1.2 Literature Review .....   | 5    |
| 1.3 Objectives and Research Questions.....  | 21   |
| 1.4 Dissertation Structure.....   | 23   |
| 1.5 References .....  | 26   |
| 2 TYING SOCIAL MEDIA STRATEGY TO FIRM PERFORMANCE: A SOCIAL<br>MEDIA ANALYTICS FRAMEWORK..... | 30   |
| 2.1 The Social Media Measurement Dilemma .....  | 32   |
| 2.2 Scope of the Problem: The Social Media Ecosystem .....                                    | 37   |
| 2.3 Focus on the Customer-Firm Social Media Dyad .....  | 42   |
| 2.3.1 Measuring layers of social media-enabled B@C activity .....                             | 46   |
| 2.3.2 Initiator goals across layers.....  | 57   |
| 2.4 The Measurement Problem.....  | 62   |
| 2.5 Discussion and Implications for Future Research.....                                      | 65   |
| 2.6 References.....   | 68   |

|       |   |     |
|-------|---|-----|
| 3     | THE ANALYSIS OF UNSTRUCTURED DATA: <i>MEANINGFUL</i>            |     |
|       | MEASUREMENT OF SOCIAL MEDIA INTERACTIONS.....                   | 73  |
| 3.1   | The Problem of Proliferation .....                              | 75  |
| 3.2   | Measuring Social Media Data.....                                | 77  |
| 3.3   | Information Processing x Decision Making.....                   | 83  |
| 3.3.1 | Information processing modes .....                              | 84  |
| 3.3.2 | Decision-making modes .....                                     | 85  |
| 3.4   | Exploiting Textual Data.....                                    | 88  |
| 3.4.1 | Inaccuracies of decontextualized, word-level analyses.....      | 89  |
| 3.4.2 | The promise of NLP.....   | 92  |
| 3.4.3 | NLP challenges .....  | 95  |
| 3.4.4 | Machine translation.....  | 96  |
| 3.5   | Social Media Analytics System .....                             | 98  |
| 3.5.1 | Why is a design theory for social media analytics needed? ..... | 101 |
| 3.5.2 | Design principles .....   | 102 |
| 3.6   | Implications and Future Research .....                          | 105 |
| 3.7   | References.....   | 106 |
| 4     | THE IMPACT OF NATURAL LANGUAGE PROCESSING-BASED TEXTUAL         |     |
|       | ANALYSIS OF SOCIAL MEDIA INTERACTIONS ON ORGANIZATIONAL         |     |
|       | DECISION MAKING.....  | 109 |
| 4.1   | The Promise of Natural Language Processing.....                 | 110 |
| 4.1.1 | Evidence-based IS .....   | 119 |
| 4.2   | Literature Review and Development of Design Principles.....     | 122 |
| 4.2.1 | Design principles and their implications .....                  | 124 |

|  |     |
|--|-----|
| 4.2.2 Exhaustiveness of design principles .....                      | 135 |
| 4.3 Propositions and Model .....                                     | 137 |
| 4.4 Research Methodology .....                                       | 143 |
| 4.4.1 General task environment .....                                 | 144 |
| 4.4.2 Experimental design.....                                       | 144 |
| 4.4.2.1 Variables and variable relationships .....                   | 147 |
| 4.4.2.2 Independent variable .....                                   | 148 |
| 4.4.2.3 Dependent variables.....                                     | 157 |
| 4.4.2.4 Control variables .....                                      | 159 |
| 4.5 Data Analysis .....  | 160 |
| 4.5.1 Background information on subjects .....                       | 160 |
| 4.5.2 Experimental results .....                                     | 160 |
| 4.6 Implications of Results.....                                     | 166 |
| 4.7 Discussion .....   | 168 |
| 4.7.1 Limitations of the study .....                                 | 171 |
| 4.7.2 Conclusions.....   | 176 |
| 4.7.3 Future research.....   | 176 |
| Appendices.....  | 182 |
| A: Regarding the Construct Validity Dilemma .....                    | 182 |
| B: Experimental Interface .....                                      | 184 |
| C: Raw Data Set Used in Experiment.....                              | 185 |
| D: Decision-Assistance Panel Contents .....                          | 187 |
| E: Description of Tweet Selection, Sentiment-Based<br>Condition..... | 193 |

|   |     |
|---|-----|
| F: Instructions Given to NLP Simulation Card Sorters .....                                  | 197 |
| G: Hypergeometric Distribution .....  | 202 |
| H: Problem/Opportunity Categorization of NLP-simulated<br>Output for Accuracy Scoring ..... | 205 |
| I: Information Letter Emailed to Prospective Participants.....                              | 213 |
| J: Instructions Read to Participants before Experiment.....                                 | 215 |
| K: Debriefing Questionnaire .....   | 217 |
| L: Detailed Analysis of Experimental Results.....   | 218 |
| I. Statistical Methods .....  | 218 |
| II. Requirements of the ANOVA Model .....   | 221 |
| III. Power Analyses .....   | 224 |
| IV. Statistical Results and Tests of Hypotheses .....                                       | 225 |
| 4.8 References.....   | 237 |
| 5 CONCLUSION.....   | 243 |
| 5.1 Discussion across Papers .....  | 243 |
| 5.2 References.....   | 245 |

## LIST OF TABLES

|   | Page |
|---|------|
| Table 2-1: Focal dyad activities according to goal.....   | 58   |
| Table 2-2: Areas of applicable literature with example citations .....  | 59   |
| Table 3-1: Cross-matrix of information-processing and decision-making modes<br>supported by textual data analysis ..... | 84   |
| Table 4-1: Examples of practice instituted prior to theory development .....  | 118  |
| Table 4-2: Theory and practice in IS research .....   | 120  |
| Table 4-3: Sources of inaccuracies across text interpretations.....   | 131  |
| Table 4-4: Operationalizations of propositions and concepts in experiment .....   | 143  |
| Table 4-5: Overlap of NLP-simulation outputs.....   | 152  |
| Table 4-6: Confidence in problem detection and opportunity detection<br>scale wording .....                             | 159  |
| Table 4-7: Means and ANOVA results for dependent variables .....  | 162  |
| Table 4-8: Summary of hypothesis testing.....   | 165  |
| Table 4-9: 20 Tweets comprising manual treatment – sample set .....   | 187  |
| Table 4-10: 20 Tweets comprising SA treatment.....  | 189  |
| Table 4-11: 20 Tweets comprising NLP treatment.....   | 191  |
| Table 4-12: Words appended to SA lexicon.....   | 195  |
| Table 4-13: Distribution of SA scores .....   | 196  |
| Table 4-14: Resultant problem and opportunity categories .....  | 208  |
| Table 4-15: Problem and opportunity Tweets presented to category sorters.....   | 208  |

|  |     |
|--|-----|
| Table 4-16: Categories of problems devised by independent raters .....                                     | 211 |
| Table 4-17: Categories of opportunities devised by independent raters .....                                | 212 |
| Table 4-18: Omnibus ANOVA F-test .....   | 220 |
| Table 4-19: Bartlett's test for homogeneity of variance .....  | 222 |
| Table 4-20: Expected power of ANOVA tests according to cell size.....                                      | 225 |
| Table 4-21: Number of problems identified: mean score, standard deviation, and cell<br>size.....           | 225 |
| Table 4-22: Number of problems analysis of variance .....  | 226 |
| Table 4-23: P-values for paired comparisons of number of problem means .....                               | 226 |
| Table 4-24: Accuracy of problem identification: mean score, standard deviation, and cell<br>size.....      | 228 |
| Table 4-25: Accuracy of problem identification analysis of variance .....                                  | 228 |
| Table 4-26: P-values for paired comparisons of accuracy of problem<br>identification.....                  | 229 |
| Table 4-27: Confidence in problem identification: mean score, standard deviation, and<br>cell size .....   | 230 |
| Table 4-28: Confidence in problem identification analysis of variance .....                                | 230 |
| Table 4-29: P-values for paired comparisons of problem confidence means .....                              | 231 |
| Table 4-30: Number of opportunities identified: mean score, standard deviation, and<br>cell size .....     | 232 |
| Table 4-31: Accuracy of opportunity identification: mean score, standard deviation, and<br>cell size ..... | 233 |
| Table 4-32: Accuracy of opportunity identification analysis of variance .....                              | 234 |



Table 4-33: P-values for paired comparisons of accuracy of opportunity

identification.....234

Table 4-34: Confidence in opportunity detection: mean scores, standard deviation, and

cell size .....235

## LIST OF FIGURES

|   | Page |
|---|------|
| Figure 2-1: Social media ecosystem.....   | 39   |
| Figure 2-2: Firm/customer (B@C) social media dyad.....  | 43   |
| Figure 2-3: Community-building-to-firm-performance chain of constructs .....                              | 48   |
| Figure 3-1: Brand-centric conversation snippet extracted from a popular social media<br>application ..... | 80   |
| Figure 4-1: Input-process-output map of design principles .....   | 136  |
| Figure 4-2: Research model highlighting propositions .....  | 142  |
| Figure 4-3: Variable relationships.....   | 147  |
| Figure 4-4: Decision-assistance panel on right, raw data feed on left.....                                | 148  |
| Figure 4-5: Graphical representation of algorithm for NLP-based simulated<br>output .....                 | 153  |
| Figure 4-6: Rationale for predictive validity of decision-making quality.....                             | 183  |
| Figure 4-7: Screen shot of browser-based experiment interface .....                                       | 184  |
| Figure 4-8: Sentiment analysis score histogram.....   | 196  |

## CHAPTER 1

### INTRODUCTION AND LITERATURE REVIEW

#### **1.1 Problem Statement and Motivation**

Social media are the wild west of information systems, especially from the firm's perspective. As evidenced by the number of organizations that have entered the social media fray, firms expect to derive some benefit from social media's capabilities (Barnes 2010). The problem is, we have no idea if that presumption is true. We do not know what areas of firm-customer interaction benefit from being mediated by social tools, or how to go about measuring all the events and processes unfolding in the social media world in a way that enables firms to judge whether they are truly gaining any value from social media.

Trying to measure social media is like trying to measure the Internet—the concept has no meaning without context and boundaries. There are far too many types of processes taking place over too wide a range of functionalities, participants, and outcomes to generalize about the social media “environment” in any sweeping way. In order to describe the effects of social media on any number of outcome variables (firm performance representing one obvious dependent variable of interest for firms), we need to work out what to measure and how to go about measuring it. Without the ability to define and gauge the consequences of their social media strategies, firms remain unable to align their social media initiatives with their organizational goals, which we

argue is a necessary precursor for creating business value from this newly pervasive phenomenon.

Concerns over measurement are especially salient given the explosive growth in the number of organizations interacting with customers through social media interfaces. A 2010 study counted 23 percent of Fortune 500 companies with public-facing blogs, 60 percent with corporate Twitter accounts, and 56 percent with corporate profiles on Facebook (Barnes 2010). Further complicating the social media scene are the diversity of additional tools available to firms for reaching customers, the varying degrees of influence that firms wield over customers depending on the tool, and the array of interests and practices supported by a wide range of technological affordances (Boyd and Ellison 2008). Commonly, companies attract “fans” to join their social media communities and consumers compile lists of “favorites” to “follow” on Facebook, popular blogging platforms, and any number of other related brand-centric communication forums. However, “social” capabilities are embedded in a wide variety of other applications as well, for example the user-generated review sections within otherwise traditional e-commerce sites, which must also be monitored or at least considered by those firms concerned with brand reputations and customer satisfaction propagating over the social IS landscape. It is suggested that firms should actively exploit all of these tools in the interest of improving internal operations; deriving value from collaboration with business partners, customers, and suppliers; and reducing support costs (Culnan et al. 2010; Demetriou and Kawalek 2010).

Correspondingly, organizations are spending increasing amounts of money on social media initiatives. Forrester expects social media marketing in the U.S. to grow at an annual rate of 34 percent from 2010 to 2014 (300 percent over five years), reaching

an estimated worth of USD 3.1 billion (VanBoskirk 2009). This rate doubles the expected growth of all other online marketing combined. Organizations ranging from the automobile industry (e.g., Volvo, Audi) to news outlets (e.g., NPR, CNN) to pharmaceutical companies (e.g., Johnson and Johnson, Pfizer) now feature Facebook, Twitter, and other social media buttons or icons directly on their home pages, television commercials, print advertisements, and even product labels.

While at first blush social media may appear to be just another channel for contacting customers, or a slight twist combining a couple of existing Internet-related technologies, the phenomenon actually conveys a huge shift in how millions of users around the world go about their daily lives compared with just a few years ago (Kane and Fichman 2009; Seo and Rietsema 2010). Looking beyond the usual suspects of Facebook, Twitter, and the like, social media are even more pervasive when we realize they include not just dedicated social networking sites; additionally, blogs, wikis, and user-contributed product reviews within traditional e-commerce sites all drastically influence how people maintain relationships with one another, consume information and services, conduct research on major and minor purchases, and interface with the world in general (Laczniak et al. 2001; Mangold and Faulds 2009; Seo and Rietsema 2010; Fournier and Avery 2011). Even when we narrow the scope of social media's reach to just the segment of activity promulgated by or directed at corporations or brands (thus excluding all the myriad personal uses of social media that preoccupy and entertain millions of humans on a daily basis (Nielson 2010)), the effects involved are substantial. The reality is that social media are not simply a new set of marketing tools in the arsenal of the firm, but entail a transformation in how firms provide customer service, garner loyalty, and any other number of traditional activities previously

achieved through non-social means (Berthon et al. 2008; Mangold and Faulds 2009; Moran and Gossieaux 2010).

Given the lack of social media related studies to date, our research represents a foray into a previously unmeasured and undissected IS phenomenon. Aside from a few scattered studies that tangentially speak to social media as simply another Internet technology (e.g., Wattal et al. 2010) and a short list of conference (e.g., Xu and Zhang 2009; Kaganer and Vaast 2010; Seo and Rietsema 2010) and practitioner (e.g., Culnan et al. 2010; Gallagher and Ransbotham 2010; Hanna et al. 2011; Kietzmann et al. 2011; Weinburg and Pehlivan 2011) papers examining certain aspects of social media or social media's impact in certain sectors of industry (e.g., Chou et al. 2009; Xu and Zhang 2009), there is very little in the way of theoretical explanation of the motivations or goals that drive customer and firm usage of social media tools. Without a rigorous examination of the range of activities that customers undertake in the social media milieu with respect to firms and their brands and services, we have no theoretical basis for deriving useful, actionable measurements that ultimately inform us as to how organizational performance is impacted by the use of social media. The present research therefore begins with a focus on social media, or more specifically the particular things that customers accomplish with social media with respect to the firm, as a basis for determining which activities firms should channel their resources into measuring, and proceeds with an exploration of particular technologies likely necessary to achieve measurement. As such, this dissertation is intended to address multitude of gaps in the current IS literature regarding the study of social media as well as the analysis of the type of unstructured text created within social media contexts.

## **1.2 Literature Review**

As per the requirements of the graduate school pertaining to three-manuscript dissertations, we include a literature review section here. These literature reviews are directly excerpted from chapters two through four and thus may be skipped in favor of reading those chapters. This introduction resumes on page 21 with a discussion of objectives and research questions.

### **From Chapter Two**

Despite extensive treatment by the popular press, social science research has yet to integrate social media extensively into its theorization (Webster 2010); within the academic literature, discussions of social media are sparse. However, a review of articles across disciplines, meager as the set is, indicates a tentative agreement on the critical characteristics of social media, if not an exact definition. Often referred to interchangeably as “consumer-generated media” (Mangold and Faulds 2009), “Web 2.0” (Wattal et al. 2010), and “user-generated information systems” (Desautels 2011), the label “social media” tends to describe those Internet-based applications predicated on the creation and exchange of user-generated content (Kaplan and Haenlein 2010) across communities of networked individuals.

As customers’ expectations of social-media-based brand-support communities compel organizations to implement these initiatives, they become the norm for both customers and organizations. Normative pressures induce companies to jump on the social media bandwagon to avoid the impression of being outdated or out of touch with innovative technologies compared to their peers and competitors (Sterne 2010). Social media represent an important enterprise-wide phenomenon that implicates multiple business functions including marketing, sales, customer service, public relations,

operations, and product development (Bernoff and Li 2008). Once customers become social media participants, instead of approaching the web as a mode of locating information and receiving marketing messages controlled and disseminated by brand managers, they now employ it as a medium for generatively co-creating a wide array of informational objects ranging from product designs to advertising campaigns to organizational processes (Etgar 2007; Berthon et al. 2008; Fournier and Avery 2011). As such, companies require mechanisms for measuring the risks and benefits accrued through interfacing with their customers across social media.

The lack of established metrics in the literature tying social media advertising (or other types of persuasive campaigning) to actual performance speaks to both the need for reliable and valid definitions and measurements of social media, and the difficulty of coming up with such measurements on the fly. In a 2010 analysis of the impact of Web 2.0 on the 2008 presidential campaign, Wattal et al turned to Gallup poll standings as a function of traditional media, Web 1.0 (traditional web sites), and Web 2.0 (YouTube, MySpace, blogs), based on mentions of the candidate via each medium the month prior. The decision to lag polling data by a month to connect Internet use to the following month's Gallup poll numbers introduces a possible disconnect between cause and effect in our social media research—this maps to the problem that IS researchers to date have had no empirically derived guidelines on which to rely, exacerbated by long feedback loops.

### **Brand Communities**

The marketing literature cites brand communities as not only a driver of loyalty and a factor that increases a consumer's likelihood of adopting a new product from the preferred brand, but also as a basis for oppositional loyalty against competitors'



products (Thompson and Sinha 2008). Organizations have been seeking the “Holy Grail” of brand loyalty through the development of communities for decades (McAlexander et al. 2002), long before the advent of social media. Defined by the commonality of its members and the relationships among them, a community is a network of social relations; a *brand* community is a specialized social group organized around a particular brand that exhibits shared consciousness, rituals and traditions, and a sense of moral responsibility (Muñiz Jr. and O’Guinn 2001). Brand community has evolved from being conceptualized as a customer-to-brand relationship (e.g., Aaker 1997) to a customer-brand-customer triad (Muñiz Jr. and O’Guinn 2001) to a network of relationships including customer-to-brand, firm-to-customer, and intra-customer interactions (McAlexander et al. 2002), a configuration that strongly resonates with the makeup of the social media ecosystem.

Reinvigorated by two-way conversations with customers, the ability to collect in-depth records of consumer preferences, and the power to “micro-target” or address customized messages to individuals—all tasks that have been simplified by the existence of social media technologies—firms are now turning to social media outlets as leverage for shaping brand-centric communities in ways previously unrealistic with traditional mass media (Fournier and Avery 2011). The existence of brand communities draws on one of the most basic human motivations, the desire to belong to a larger collection of likeminded peers, to fit in, to be accepted; with brand consumption serving as the basis for coalescence and social media facilitating the connectivity, firms have an unprecedented platform for exploiting consumers’ basic drives (e.g., to “belong”) in such a way that benefits the brand or product (Fournier and Avery 2011).

Brand communities are linked to the retention of consumers via the mechanism of increased brand loyalty, which in turn positively impacts a firm's bottom line. A valid system of measurement for firm-directed social media efforts must account for all crucial relationships determined to comprise the construct of brand community from a customer-experiential perspective: relationships between the customer and the brand, the customer and the firm, the customer and the product, and intra-group customer-to-customer relationships (McAlexander et al. 2002).

### **Social Media Marketing**

Exposure to social media advertisement is a function of consumer volition—viewers rank and rate content, disperse it to friends, and re-post or re-tweet information idiosyncratically based on their interests (Mangold and Faulds 2009). Social media marketers relinquish control over the reach, frequency, and timing of their campaigns (Fournier and Avery 2011). Once firms decide to engage customers via social media tactics, the risk becomes that customers may reject being “herded” or manipulated, simply declining to be advertised to within the social media context by brands they perceive as too calculating or that do not appear savvy in this new cultural realm.

Clever marketers have designed campaigns that clandestinely enable spoofs, identified by particularly savvy firms as desirable due to the high viral currency and ultimate cultural resonance such “hits” often indirectly effect (Ferguson 2008; Fournier and Avery 2011). Although viral tactics have been accused of merely resulting in short-term attention, it may be possible for firms to leverage such messages in building customer loyalty by launching (or covertly instigating) campaigns that ultimately beget consumer identification with other like minds and promote “sticky” dialogue (Ferguson 2008).

Viral as well as word of mouth marketers face the same problem in measuring the effects of their campaigns; researchers know that these types of advertisement build brand awareness, but they are unsure how to calculate the effect on market share (Ferguson 2008).

The unique opportunity conferred to firms by the ability to monitor customer-to-customer exchange is not limited to analyzing huge streams of data, although that is an enormous source of potential advantage. Monitoring customer-to-customer streams also imparts to firms the capacity to interject customer service into negative exchanges, thus influencing customer satisfaction and opinions and derailing potential public relations problems. Companies ranging from Comcast to Jet Blue monitor outlets like Twitter for any mention of their company, searching for opportunities to provide information to needy customers or correct misinformation subject to inadvertent propagation by members of their consumer bases (King 2008).

Studies indicate that consumers multi-task in their media consumption, simultaneously participating in online and traditional modes of information intake (Russell 2010). The literature suggests that intra-group online dialogue can yield customer insight as well market intelligence (Gallaughier and Ransbotham 2010), though we lack a foundation of measurement for deriving useful meaning from trends over hundreds of thousands or millions of these data points. While a single person or team can monitor for possible negative mentions of a brand or firm, firms invite peril when they give too much credence to extreme positive or negative feedback from a vocal but small faction of overall customers (Fournier and Avery 2011).

### **From Chapter Three**

As much as 85% of organizational data exists in unstructured (textual) format (Lindvall, Rus, and Sinha 2003). Emails, corporate documents, news articles, web pages, and voicemail transcriptions typically occur outside of the bounds of pre-defined data models yet constitute dense and voluminous bodies of data that companies must store, process, and analyze to derive business intelligence and ultimately create value (Feldman and Sanger 2007). Although information in the unstructured environment is abundant and ostensibly useful, the sophistication of techniques for the analyses of texts is meager compared to what is available in the structured environment (Inmon and Nesavich 2008: xvii).

Systems designed to address the analysis needs of organizations interested in evaluating and summarizing text communications have generally enabled visualization of metadata contained in message headers (i.e., send/reply and posting patterns), but have provided little support for the analysis of actual message body text (Abbasi and Chen 2008). Despite the potential it holds for uncovering previously unknown information from the depths of large collections of text (Hearst 1999), the IS management field's reluctance to capitalize on the advances made in this interdisciplinary territory likely stems, at least in part, from the convolution that muddles our understanding of what technologies actually constitute text data mining, and to what ends.

Analysis of the underlying interactions driving social media activity (Larson and Watson 2011) indicates three levels of measurement inherently applicable to social media-generated data, the simplest being counts of objects and actions such as users, comments, and links followed. A more revealing yet problematic mode of measurement,

sentiment analysis (or the assessment of positive and negative customer sentiment in product or brand reviews or mentions) (Pang and Lee 2008) has demonstrated potential to impart insight into customers' reactions to a given organization and its products and services. However, granularity is been lost through sorting comments and reviews according to simple negative/positive rating scales (e.g., Pavlou and Dimoka 2006).

While just-in-time quantitative business intelligence tools reduce the latency between data acquisition and analysis (Chaudhuri, Dayal, and Narasayya 2011), the congruent ability to glean deep meaning from voluminous streams of qualitative data has yet to be fully established and incorporated into organizational business intelligence-oriented monitoring.

Proliferation of qualitative data via social media applications is overwhelming, but modern technologies enable us to store and process these data and create the potential to detect patterns; as we increase the sophistication of these capabilities we leverage a valuable source of information not just for firms and brand managers but for a wide array of knowledge professionals ranging from physicians (Denecke and Nejdll 2009) and pharmacists (Agarwal and Searls 2008) to scientists (Shatkay and Feldman 2003) and manufacturers (Choudhary, Harding, and Tiwari 2009) who rely on textual data analysis to effect a variety of tangible and intellectual outcomes.

In manual analysis mode, humans interpret the message intended by the text's author by reading sentences and paragraphs as well as noting contextual features of the message or document that may convey meaning (Anderson and Pérez-Carballo 2001). This perception includes the tacit understanding of how objects relate to each other in the world, the goals people tend to seek in their daily lives, and the emotional impact of certain kinds of events or situations, refer to the concept of "common sense," or the

“obvious things people normally know and usually leave unstated” (Grassi et al. 2011). A human analyst thus consumes words and features of a text string in sum and processes them from a general perspective of shared language and common experience of the world and its everyday situations, in turn increasing the chances that sophisticated forms of speech such as verbal irony or sarcasm (Davidov, Tsur, and Rappoport 2010; González-ibáñez and Wacholder 2011), colloquialisms such as slang or pop cultural references, or even misspellings (Furnas et al. 1987) do not delimit understanding or interpretation.

Replicating this perception according to a sorting procedure, words are at a minimum sorted into two categories, the most popular schema being the “positive” versus “negative” evaluation standardly known as sentiment analysis but also called opinion mining, subjectivity analysis, appraisal extraction, and affective computing across multiple related literatures (Pang and Lee 2008). A more useful analysis would capture links between subtopics in a review and corresponding opinions, but such associations are extremely difficult to extract accurately at the word-level. It is understood that extraction algorithms perform best when the topic is known *a priori* (Yi et al. 2003), offering little benefit to organizations attempting to unearth new knowledge from open-ended text data. Statistical word-based approaches cannot reliably extract and preserve associations between multiple topics and corresponding sentiments in the same message (Yi et al. 2003).

## **Natural Language Processing**

Natural language processing (NLP) refers to a wide range of language technologies, tasks, subtasks, and related fields and is often used interchangeably with the phrase “computational linguistics” in academia. In its widest interpretation, NLP

can mean any type of computer manipulation of natural language (i.e., English, Chinese, French) used by humans to communicate. This can include simple counts of word frequencies, or extend to the automated “understanding” of human verbalizations (Bird, Klein, and Loper 2009).

NLP tools can enable the realization of fuller meaning from free text data streams through the preservation and exploitation of linguistic rules like parts of speech (i.e., nouns, verbs, adjectives, etc.) and grammatical structures (the application of sentence formation rules in a given language), advances in resolving anaphora (e.g., aligning backward-referring pronouns and phrases with the appropriate nouns) and ambiguities of language and grammatical structures, and extracting relationships among entities (Bunescu and Mooney 2007). Other common text-classification approaches treat documents or text segments as unstructured buckets of words with frequency counts but no relationship with respect to one another (Kao and Poteet 2007: 2), which does not readily support the non-trivial goals of discovering events, entities, and relationships (e.g., who likes what product, who agrees with whom and why, or how customers use products).

A relevant example for social media analysis, the accurate automatic extraction of information from biomedical texts illustrates the difficulty of semantic extraction tasks due to the misalignment between most existing natural language tools (e.g., tokenizers, parts-of-speech taggers, parsers) and the biomedical body of literature. Problems arise because such tools have traditionally been trained against news corpora, thus incurring a loss of accuracy when ported into a biomedical setting (Bunescu and Mooney 2007). Scientific publications follow a substantially different narrative type with relevant entities including proteins, genes, and cells and relations following patterns such as

subcellular location and protein-protein interaction, whereas newspaper discourse usually includes mentions of entity types such as people, organizations, and places and relation types including social relationships, positions people hold in organizations, relationships among organizations, etc. (Bunescu and Mooney 2007).

Named entity recognition (NER) is a separate NLP task whose goal is to identify within text all the names for specific types of things, typically persons, organizations, and locations (Sang and De Meulder 2003). In the case of biomedical extraction, NER faces difficulty in the seemingly straightforward task of recognizing gene, drug, and protein names (Cohen and Hersh 2005). Fundamental to more complex text mining tasks such as relationship extraction (because relationships are anchored by participating entities), the process of recognizing biological entities in order to represent them in some consistent, normalized form has met with several obstacles, notably the lack of a complete lexicon comprising all possible biological named entities which thus precludes the use of simple text-matching algorithms (Cohen and Hersh 2005). Similar challenges exist when applying NLP techniques to the domain of brands and products.

One possible approach to resolving this problem is to look to the advances made in the overlapping field of statistical machine translation (SMT). The dominant framework for modern machine translation research (Hutchins 2006), this data-driven or corpora-based, machine learning method describes the automated translation of text from source to target language through algorithms that automatically “learn” to translate by examining millions of samples of human-produced translation (Lopez 2008).

Statistical (as opposed to rules-based or example-based paradigms) translations maximize the probability that a string in the target language is the translation of a string



in the source language, although these probabilities and searches may be modelled according to numerous approaches (Brown et al. 1993). The parameters of these distribution models are derived from training data in the form of comparative analysis of bilingual corpora (Brown et al. 1993).

On a conceptual level, translation from target to source language follows the general idea of converting the source sentence into a knowledge representation via the use of a dictionary that maps words (e.g., *river*) onto concepts (e.g., river) with corresponding fact-based limitations based on world knowledge (such as, rivers cannot ride horses) (Knight 1997). Specifically, at the sentence-level, words or word sequences of the source language are aligned with corresponding sequences in the target language. Based on these alignments, translation occurs through the selection of the most probable target output for each input phrase as well as a determination of the most probable output sequence (sentence structure), based on millions of known aligned phrases (Hutchins 2006).

#### **From Chapter Four**

Social media is a relatively new mechanism for eliciting and disseminating information in such forms as consumer opinions, suggestions, and conversations (Demetriou and Kawalek 2010). It heralds both an increasing concern and an invaluable opportunity for firms whose strategies include leveraging consumer-generated qualitative data to create business value (Culnan, Mchugh, and Zubillaga 2010; Hoffman and Fodor 2010). Defined from a consumer/firm perspective as the set of connectivity-enabled applications that facilitate interaction and the co-creation, exchange, and publication of information among firms and their networked communities of customers (Larson and Watson 2011), social media engender multiple

complex layers of brand-centric text-mediated interactions. Of particular relevance to firms is the layer comprising customer-to-customer interactions such as recommendations, reviews, collaborative exchanges, and helpful suggestions or advice (Larson and Watson 2011). For firms to detect among these interchanges important cues such as adverse event mentions and consumer reactions to new products, social media analysts and managers require the ability to qualitatively mine textual data possibly symbolizing and conveying these cues. This level of measurement exceeds the simple positive-negative labeling inherent in sentiment analysis (Pang and Lee 2004, 2008) and the simpler measurement technique of gathering count data for characteristics such as number of followers, number of likes, etc.—important but incomplete methods for extracting knowledge from qualitative consumer-generated data.

Challenges to automatically processing highly unstructured text frequently reflect limitations imposed by porting unstructured text into a structured environment, a process that involves the decomposition of sentences into words that can then be easily stored, retrieved, and evaluated. While the advantages of this methodology for dealing with text include simplification of the analytical process by freeing the analyst from concerns regarding preservation or comprehension of context (Inmon and Nesavich 2008: xix), major drawbacks stem from the examination of decontextualized words, ranging from inability to resolve sarcasm or anaphora (expressions whose meanings depends on other referential elements) to the inability to decipher simple spelling errors.

Natural language processing (NLP), an alternate composite field that blends computer science, machine learning, and linguistics research, aims to extract meaning from texts by considering them in their natural language format. This field approach

encompasses a wide range of disciplines and tasks focused on extending the capabilities of text mining, or the extraction of knowledge from unstructured text (Hearst 1999), most recently by incorporating the machine-learning paradigm of language processing. NLP algorithms have met with some success in structured domains with limited lexes such as medicine and biochemistry (Tanabe et al. 1999), fields in which knowledge acquisition is ontologically bounded (Maedche and Staab 2000; Wilcox and Hripcsak 2003).

NLP-related research has recently seen progress in the technical capabilities of machines to discover new, non-trivial knowledge from free text, although the automated mining of data from unstructured text is still in its relative infancy. Emerging subfields and approaches continue to extend text mining proficiencies in the contexts of real-world data. For example, improved automation of lexicon augmentation in named entity recognition, or the accurate labeling of persons, organizations, and locations (Sang and De Meulder 2003), increases the body of task-specific lexicons available for a variety of natural language processing tasks. Thus, instead of relying on general-purpose lexicons or tediously and slowly compiling task-specific lexicons by hand, highly tailored lexicons can now be built on the fly by leveraging named entity extraction from HTML data on the Web via a search engine (McCallum and Li 2003). Similarly, incremental improvements to a wide range of specific capabilities such as parts-of-speech tagging, parsing (determining the grammatical tree of a sentence), and anaphora resolution (determining which noun or name a pronoun refers to) combine to contribute to discipline-level progress and suggest potential applicability in less-structured or unstructured text environments such as social media (Bunescu and Mooney 2007; Kao and Poteet 2007; Agichtein et al. 2008).

The mining of text encompasses a vast array of theoretical approaches and methods (Feinerer, Hornik, and Meyer 2008), including information retrieval, clustering, classification, entity-relationship and event extraction, and natural language processing (Hotho, Andreas, and Paaß 2005), each the focus of intense ongoing research. The field of NLP relates to each of the approaches and methods listed above as well as comprising a long list of additional subfields. The general goal of NLP is to create algorithms capable of “understanding” natural language through techniques ranging from the simple manipulation of strings to the automatic processing of natural language inquiries (Hotho, Andreas, and Paaß 2005). This methodology contrasts, for example, with Information Retrieval (IR), the goal of which is to return units of text matched according to pre-specified patterns. IR is essentially the confirmatory counterpart to NLP, although NLP can be incorporated into IR algorithms to increase their effectiveness through increased clarification of word ambiguity (Arazy and Woo 2007).

### **Statistical Machine Translation**

Mental-models research indicates that humans understand patterns of words locally; multiple instances of a single word situated among different surrounding words are not perceived as semantically related by most speakers of English (Fox 1986). For example, we do not consider “my soup is *cold*” to have any relation to “I have a head *cold*.” But if we extract *cold* from the rest of the sentence in which it exists, which is equivalent to what happens during sentiment analysis or other non-NLP based approaches, we then have no idea what the word actually means or whether it should be interpreted as a positive, negative, or neutral sentiment.

Collocation indexing, or the process of extracting overall syntax based on the identification of word combinations that carry specific meaning in natural text, has

proven successful at word disambiguation in large scale systems that use naturally occurring text (Arazy and Woo 2007). This statistical NLP technique has proven to reduce the gap between the way humans think of information and the way in which it is represented by machines (Arazy and Woo 2007). At the very least, a reliable sentiment analysis approach to social media analytics would require the incorporation of this NLP-predicated capacity to identify meaningful word combinations with meaning separate from that of their individual components; human communication is replete with such complex expressions.

The automated mining of text can be likened to the task of machine translation (MT) in that the goal of both is to interpret one set of words and translate them into an output of similarly-intended set of words, but in a form that is understandable to the recipient. Thus, while a language translator converts French sentences into English, a social media analytics system would interpret a Tweet or status update into an output that is meaningful to the organization deciding how it should react to the message. The output in this case may look like a phrase or sentence that conveys the latent (or even manifest) intent of the original text in terms relevant to the brand or organization.

Statistical machine translation (SMT) has emerged as the dominant, even “mainstream” machine translation approach over the last decade or so despite the competition of theoretically-driven, rules-based alternatives (Hutchins 2006). These theory-driven methods did not prove robust in practice and so subsided to a corpora-driven MT model based primarily on word frequency and word combinations derived from large volumes of real data (Hutchins 2006). Given the similarities of task, goal, and amount of data with which to begin training, we conclude that an effective approach to the “translation” of social media data into business intelligence should follow a

parallel methodology that we label statistical machine interpretation (SMI). The basis of SMT, and subsequently SMI, is machine learning, a paradigm that calls for general learning algorithms typically grounded in statistical inference. Statistical machine translation and machine learning are interrelated in their analysis of large corpora of real-world data during the training phase, from which an evaluation model is subsequently derived for new sentences (Lopez 2008).

Statistical machine learning is useful when tasks cannot be solved strictly by classical programming techniques due to the lack of an available mathematical model (Cristianini and Shawe-Taylor 2000). SMT can improve a wide range of automated processes predicated on unstructured data; for example, it has enhanced the efficacy of automated detection systems for combating fraudulent websites (Abbasi et al. 2010).

Statistical learning theory (SLT), also known as the Vapnik-Chervonenkis theory, is the underlying computational learning theory that describes the learning process from a statistical perspective. Purely theoretical until the 1990's, SLT has since bolstered the development of highly effective algorithms, in particular support vector machines (SVM) (Vapnik 1999). Support vector machines (SVM) are SLT-based learning algorithms belonging to the kernel methods class of pattern analysis, that, given a set of data, find patterns by embedding data into high dimensional feature space and looking for linear relationships in that space (Cristianini and Shawe-Taylor 2000).

Natural language processing scientists point out that progress in the field of statistical machine translation is largely driven by the availability of data (Koehn 2005). SMT thrives on the perpetuation of large quantities of parallel texts: original text paired with its translation into a target language. The process for translation, specifically between two natural languages, generally embody the following steps (Koehn 2005: 1)

Gather raw data (by crawling the web, or scraping social media sites), 2) extracting and mapping parallel chunks of text (document alignment), 3) break text into sentences (sentence splitting), 4) preprocess the corpus in preparation for SMT systems (normalization, tokenization), and 5) map original language sentences to target language sentences (sentence alignment). This general procedure for translating between languages assumes the existence of many parallel texts available for alignment. Thus, building a new SMT system for a language requires the development of parallel texts (Koehn 2005), whereby the core of the language model in the target language is the probabilistic phrase translation table learned from the parallel corpora.

### **1.3 Objectives and Research Questions**

In practice, as in the literature, there is no comprehensive set of reliable, practical, actionable measurements of social media effects that can be applied across settings. As such, it is difficult for researchers to contribute to the aggregate body of knowledge regarding social media. Without the necessary step of developing measures, the field of IS is unlikely to advance its knowledge of social media in any standardized, comparable fashion. In this three-paper dissertation, we first seek to study firm/customer-oriented social media activities in order to define and distinguish individual processes. Recognizing the importance of measures capable of tying social media efforts to business strategy, we propose a framework for understanding the novel characteristics of stakeholder interaction brought about by social media and suggest how each of these should be conceptualized. Focusing particularly on the customer-to-customer layer of social media interaction, we address the critical role of textual analysis in any worthwhile social media analytics system (SMAS) capable of tapping into the

wealth of latent information inherent in consumer “chatter.” We apply our theorization to the design of a class of context-preserving, machine-learning-based SMAS, the effects of which we subsequently examine in a laboratory experiment. We also take the opportunity with the third manuscript to consider a particular type of research setting, one in which we do not have a body of theory to guide the development of practice improvement, nor do we have a feasible working practice to observe and from which to base our research. Instead, we are positioned to draw from parallel, but distinctly separate, practice to steer our information system design.

Comprehensively, the driving question motivating this dissertation in its entirety asks:

*RQ<sub>1</sub>: How do social media impact firm performance?*

Granularly, we proceed through a set of research questions beginning with a big-picture investigation of social media in an organizational setting and converging onto a focused investigation of the specific technologies that make possible the mining social media text. The research questions underlying each essay includes the following (with RQ subscript number corresponding to chapter):

*RQ<sub>2</sub>: What stakeholder goals and activities are facilitated by social media from the perspective of the firm?*

*RQ<sub>3</sub>: How can we best measure social media effects in an organizational context?*

*RQ<sub>4-Practice</sub>: Can advanced natural-language-processing-based qualitative textual analysis techniques improve the decision-making capability of organizations?*



*RQ<sub>4</sub>—Theoretical: How should scholars advance knowledge in research areas characterized by a lack of strong guiding theory and inadequate or no observable practice?*

## **1.4 Dissertation Structure**

This dissertation comprises a three-manuscript model, and is organized as follows. Chapter Two represents Paper One and addresses Research Question Two. It lays out the problems of measuring social media effects, and presents a theoretical framework for breaking down social-mediated activity involving the firm and its customers into granular layers. By breaking down the messy, muddled sets of cross-interactions into discrete events and processes that map to simpler, measurable contexts, we enable the subsequent process of weeding out the interactions that do not ultimately matter to the firm. Armed with this knowledge, organizations can focus their time and resources on managing and measuring just those interactions critical to their decision-making capabilities. Chapter Two informs a more focused dissection of customer-to-customer interactions by the firm, which we subsequently address in Chapters Three and Four.

Chapter Three represents Paper Two, and addresses Research Question Three. By mapping the objective measures generated by social media use to the collaborative, persuasive, and awareness-generating activities occurring in the B@C social media dyad, we begin to codify the overall set of measures needed for assessing social media effects from the firm's perspective. We recognize three levels of measurement inherent in such a system, the simplest being count data. While overall counts of users, comments, referring links, etc., are important for grasping general trends and scope of customer use, these numbers convey a limited depth of understanding in terms of

customer reaction and sentiment. A somewhat more revealing level of measurement, analysis of positive or negative mentions, begins to impart more insight into customers' reactions to a given organization, its products and services, its special events and social media initiatives, etc. However, the majority of insight into the impact of a given product, service, campaign, etc., is going to come out of a deeper level of qualitative analysis capable of analyzing textual data on a large scale. While counts of users and actions and ratios of positive to negative mentions are fairly accessible to firms, the ability to glean deep meaning from voluminous streams of social media-generated data is a proficiency yet to be fully established.

As such, Paper Two thoroughly explicates the myriad complications associated with the qualitative analysis of what we call "highly unstructured data," or informal (e.g., social media) messaging lacking of any type of *a priori* semantic or syntactical style constraints. Following a comprehensive review of the state of the art of textual analysis, particularly that which falls under the tradition of natural language processing (NLP), we culminate our arguments with set of design principles that indicate a class of social media measurement tools expected to positively affect organizational decision-making and confer the ability to competitively manage social media initiatives in an environment characterized by extensive two-way communication and collaboration between and among firms and consumers.

Extending and testing the rationales presented in earlier chapters, Chapter Four represents Paper Three, and addresses Research Question(s) Four. We culminate our investigation into appropriate mining techniques for social media with an experiment. Focusing on organizational social media managers, we determine whether the output of an instantiation of the theoretically designed social media analytics system proposed in

Paper Two (Chapter Three) outperforms the output of traditional analytic techniques such as automated sentiment analysis or manual contextual analysis in supporting organizational decision making. Given high-volume streams of highly unstructured social media text congruent with the volume and type of data generated by consumers across social media platforms on a daily basis, we are interested in the most expedient route to the extraction of actionable, accurate, and useful knowledge from customer interactions.

Chapter Five draws conclusions across all three studies, and addresses limitations and future research directions.

## 1.5 References

- Abbasi, A., and Chen, H. (2008). CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication. *MIS Quarterly*, 32(4), 811-837.
- Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker, J. F. (2010). Detecting Fake Websites: The Contribution of Statistical Learning Theory. *MIS Quarterly*, 34(3), 1-28.
- Agarwal, P., and Searls, D. B. (2008). Literature mining in support of drug discovery. *Briefings in bioinformatics*, 9(6), 479-92.
- Agichtein, E., Castillo, C., Donato, D., Gionis, A., and Mishne, G. (2008). Finding High-Quality Content in Social Media. *Proceedings of the international conference on Web search and web data mining* (pp. 183-193). ACM.
- Anderson, J. D., and Pérez-Carballo, J. (2001). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing. *Information Processing and Management*, 37(2), 231-254.
- Arazy, O., and Woo, C. (2007). Enhancing Information Retrieval Through Statistical Natural Language Processing: A Study of Collocation Indexing. *MIS Quarterly*, 31(3), 525-546.
- Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python* (p. 479). O'Reilly Media.
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., and Mercer, R. L. (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2), 263-311.
- Bunescu, R., and Mooney, R. (2007). Extracting Relations from Text: From Word Sequences to Dependency Paths. In A. Kao and S. R. Poteet (Eds.), *Natural Language Processing and Text Mining* (pp. 29-44). Springer.
- Chaudhuri, S., Dayal, U., and Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88.
- Choudhary, A. K., Harding, J. A., and Tiwari, M. K. (2009). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501-521.
- Cohen, A. M., and Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in bioinformatics*, 6(1), 57-71.
- Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (1st ed., p. 204). Cambridge University Press.
- Culnan, M. J., Mchugh, P. J., and Zubillaga, J. I. (2010). How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value. *MIS Quarterly Executive*, 9(4), 243-260.

- Davidov, D., Tsur, O., and Rappoport, A. (2010). Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning* (pp. 107-116). Uppsala, Sweden: Association for Computational Linguistics.
- Demetriou, G., and Kawalek, P. (2010). Benefit-Driven Participation in Open Organizational Social Media Platforms: The Case of the SAP Community Network. *Issues in Information Systems*, *XI*(1), 601-611.
- Denecke, K., and Nejdl, W. (2009). How valuable is medical social media data? Content analysis of the medical web. *Information Sciences*, *179*(12), 1870-1880. Elsevier Inc.
- Feinerer, I., Hornik, K., and Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*(5), 1-54.
- Feldman, R., and Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* (p. 422). Cambridge University Press.
- Fournier, S., and Avery, J. (2011). The Uninvited Brand. *Business Horizons*, *54*(3), 193-207.
- Fox, B. A. (1986). Local patterns and general principles in cognitive processes: Anaphora in written and conversational English. *Text*, *6*(1), 25-51.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, *30*(11), 964-971.
- González-ibáñez, R., and Wacholder, N. (2011). Identifying Sarcasm in Twitter: A Closer Look. *Computational Linguistics*, (2010), 581-586.
- Grassi, M., Cambria, E., Hussain, A., and Piazza, F. (2011). Sentic Web: A New Paradigm for Managing Social Media Affective Information. *Cognitive Computation*, *3*(3), 480-489.
- Hearst, M. A. (1999). Untangling Text Data Mining. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 3-10). College Park, MD: Association for Computational Linguistics.
- Hoffman, D. L., and Fodor, M. (2010). Can You Measure the ROI of Your Social Media Marketing? *MIT Sloan Management Review*, *52*(1).
- Hotho, A., Andreas, N., and Paaß, G. (2005). A Brief Survey of Text Mining. *Machine Learning*, *20*(1), 19-62.
- Hutchins, J. (2006). Example-based machine translation: a review and commentary. *Machine Translation*, *19*(3-4), 197-211.
- Inmon, W. H., and Nesavich, A. (2008). *Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence* (p. 241). Prentice Hall.
- Kao, A., and Poteet, S. R. (2007). Natural Language Processing and Text Mining. In A. Kao and S. R. Poteet (Eds.), (p. 272). Springer.

- Knight, K. (1997). Automating Knowledge Acquisition for Machine Translation. *AI Magazine*, 18(4), 81-96.
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *MT Summit* (Vol. 11, pp. 79-86).
- Larson, K., and Watson, R. T. (2011). The Value of Social Media: Toward Measuring Social Media Strategies. *ICIS 2011 Proceedings* (pp. 1-18). Shanghai, China.
- Lindvall, M., Rus, I., and Sinha, S. S. (2003). Software systems support for knowledge management. *Journal of Knowledge Management*, 7(5), 137-150.
- Lopez, A. (2008). Statistical machine translation. *ACM Computing Surveys*, 40(3), 1-49.
- Maedche, A., and Staab, S. (2000). Mining Ontologies from Text. *Knowledge Engineering and Knowledge Management Methods, Models, and Tools: 12th international conference, EKAW* (pp. 169-189). Juan-les-Pins: Springer.
- Mangold, W. G., and Faulds, D. J. (2009). Social Media: The New Hybrid Element of the Promotion Mix. *Business Horizons*, 52(4), 357-365.
- McCallum, A., and Li, W. (2003). Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons. *Proceedings of the seventh conference on Natural language learning at HLT-NAACL* (pp. 188-191). Association for Computational Linguistics.
- Pang, B., and Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (p. 271).
- Pang, B., and Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Pavlou, P. A., and Dimoka, A. (2006). The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research*, 17(4), 392-414.
- Sang, E. F. T. K., and De Meulder, F. (2003). Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL* (pp. 142-147). Association for Computational Linguistics.
- Shatkey, H., and Feldman, R. (2003). Genomic Era: An Overview. *Journal of Computational Biology*, 10(6), 821-855.
- Tanabe, L., Scherf, U., Smith, L. H., Lee, J. K., Hunter, L., and Weinstein, J. N. (1999). MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling. *BioTechniques*, 27(6), 1210-4, 1216-7.

Vapnik, V. N. (1999). An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, 10(5), 988-99.

Wilcox, A. B., and Hripcsak, G. (2003). The Role of Domain Knowledge in Automating Medical Text Report Classification. *Journal of the American Medical Informatics Association*, 10(4), 330-339.

Yi, J., Nasukawa, T., Bunescu, R., and Niblack, W. (2003). Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques. *Third IEEE International Conference on Data Mining (ICDM'03)* (pp. 427-434). Melbourne, Florida.

## CHAPTER 2

# TYING SOCIAL MEDIA STRATEGY TO FIRM PERFORMANCE: A SOCIAL MEDIA ANALYTICS FRAMEWORK<sup>1</sup>

---

<sup>1</sup> Larson, K. and Watson, R. T. 2011. "The Value of Social Media: Toward Measuring Social Media Strategies." *ICIS 2011 Proceedings*. Paper 10. Reprinted here with permission of publisher.



## Abstract

Embodying a new gestalt in firm-customer communication, social media are a nascent yet critical concern for researchers and practitioners alike. Organizations lack valid and reliable measures for social media effects, without which they remain unable to align their social media initiatives with organizational goals and ultimately create business value. This essay presents a “social media ecosystem” framework, explicating the social-media-enabled relationships among stakeholder groups and suggesting how future researchers can address research questions based on this model. Focusing on the customer/firm segment entitled the “B@C Social Media Dyad,” the article deconstructs the phenomenon of social media into multiple layers of firm-initiated and customer-initiated actions, and provides a theoretical understanding of what firms and customers accomplish using social media. It sets the stage for developing measures of those firm/customer social media activities with a critical bearing on firm performance.

**Keywords:** *social media, firm performance, measurement, collaboration, word of mouth, customer service, brand community*

## **2.1 The Social Media Measurement Dilemma**

The measurement of social media effects is an increasing concern for organizations (Hoffman and Fodor 2010). Without the ability to define and measure the consequences of social media strategies, it is difficult for firms to align their social media initiatives with organizational goals and ultimately create business value (Culnan et al. 2010). This concern is especially salient given the explosive growth in the number of organizations that interact with customers through social media interfaces and the diversity of possible channels for reaching customers (Boyd and Ellison 2008). As information systems researchers, this quest for measurement warrants our attention because measuring a phenomenon, be it social media or otherwise, is an act of information creation that necessitates subsequent information recording and processing. The particular technological phenomenon of social media, furthermore, is recognized by top IS journals as a ubiquitous facilitator of communication and collaboration embedded in humans' lives (Aakhus et al. 2011). The resultant combination of measurement of human behavior facilitated by an underlying information technology thus brings social media analytics to the forefront of IS interest.

Within the context of social media activity, companies attract "fans" to join their social media communities, and consumers compile lists of "favorites" to "follow" on Facebook, popular blogging platforms, and any number of other related brand-centric communication forums. A 2010 study counted 23 percent of Fortune 500 companies with public-facing blogs, 60 percent with corporate Twitter accounts, and 56 percent with corporate profiles on Facebook (Barnes 2010). Correspondingly, organizations are spending increasing amounts of money on social media initiatives. Forrester expects social media marketing in the U.S. to grow at an annual rate of 34 percent from 2010 to

2014 (300 percent over five years), reaching an estimated worth of USD 3.1 billion (VanBoskirk 2009). This rate doubles the expected growth of all other online marketing combined. Organizations ranging from the automobile industry (e.g., Volvo, Audi) to news outlets (e.g., NPR, CNN) to pharmaceutical companies (e.g., Johnson and Johnson, Pfizer) now feature Facebook, Twitter, and other social media buttons or icons directly on their home pages, television commercials, print advertisements, and even product labels.

This ubiquitous appearance of social media in a multitude of settings begs the question, “what exactly are social media?” Despite extensive treatment by the popular press, social science research has yet to integrate social media extensively into its theorization (Webster 2010); within the academic literature, discussions of social media are sparse. However, a review of articles across disciplines, meager as the set is, indicates a tentative agreement on the critical characteristics of social media, if not an exact definition. Often referred to interchangeably as “consumer-generated media” (Mangold and Faulds 2009), “Web 2.0” (Wattal et al. 2010), and “user-generated information systems” (Desautels 2011), the label “social media” tends to describe those Internet-based applications predicated on the creation and exchange of user-generated content (Kaplan and Haenlein 2010) across communities of networked individuals. Sufficiently expansive to capture the spirit of the phenomenon yet able to exclude technologies or information systems not recognized as social media tools, we adopt this working description for the duration of this paper as we focus on delineating the difficulties inherent in measuring social media effects. We expect a more formal definition will emerge from our program of research as we develop our understanding of

the interactions, goals, motivations, uses, and participants of the social media landscape, both from the firm's and the customer's perspectives.

Consumers are increasingly aware of corporate social media outlets, coming to expect such forums (Cone 2008) the way they grew to expect public-facing corporate websites a generation ago and e-commerce capabilities over the past decade. As customers' expectations of such brand-support communities compel organizations to implement these initiatives, they become the norm for both customers and organizations. Normative pressures also induce companies to jump on the social media bandwagon to avoid the impression of being outdated or out of touch with innovative technologies compared to their peers and competitors (Sterne 2010). Additionally, corporations are able to purchase information about their customers from some externally hosted social media sites, thus providing a wealth of minable data. Driven by these pressures to engage in social media initiatives, organizations are investing time and money in a new phenomenon that practitioners and researchers alike know very little about, and the consequences of which they understand even less.

As an important enterprise-wide phenomenon that implicates multiple business functions including marketing, sales, customer service, public relations, operations, and product development (Bernoff and Li 2008), companies require mechanisms for measuring the risks and benefits accrued through interfacing with their customers across social media. Organizations must establish useful measures of the effects of social media strategies and investigate how these correlate with traditional measures of corporate performance, such as ROI. Companies that fail to develop effective measures of social media effects are likely to concede market share to competitors and misallocate resources.

As a precursor to the theorization of effective metrics for corporate social media use, we must identify how social media change traditional customer-firm interactions, and what new objectives these technologies introduce into the relationship. For example, once customers become social media participants, they transcend the role of mere information consumers. Instead of approaching the web as a mode of locating information and receiving marketing messages controlled and disseminated by brand managers, they now employ it as a medium for generatively co-creating a wide array of informational objects ranging from product designs to advertising campaigns to organizational processes (Etgar 2007; Berthon et al. 2008; Fournier and Avery 2011). These new processes require different measures than do traditional marketing effects because they are motivated by different goals, often aim at different outcomes, and achieve existing outcomes through different means. Likewise, instead of relying on customer-initiated complaints to trigger service solutions, firms are now empowered to patrol customer-generated content for instances where they can initiate customer service. This gives firms a new tool for meeting and even surpassing customer expectations. To illustrate the importance of filling the social media measurement gap in the organizational context, we have only to imagine a firm whose strategic focus maps to the objective of providing superior customer service. Without understanding how social media change the process of providing customer service, the wrong things are likely to be looked at and measured. And without metrics derived from a theoretical understanding of the underlying processes and objectives, this hypothetical firm has no way to substantiate (or disconfirm) the success of its efforts.

Drawing on extant literature at the nexus of net-enabled organizational and IS research streams, we find viewing major stakeholders from the firm's information-

oriented perspective (Watson and Straub 2007) to be a useful platform for contextualizing our measures of social media outcomes. By mapping onto this model (see Figure 2-1) the inter- and intra-group communications engendered by social media, we can begin the process of isolating the different layers of activities and goals that come together to produce a very complex scenario. Only after untangling the myriad objectives being accomplished by a cast of stakeholders can researchers start to develop feasible, reliable, and valid measures of social media effects that will render meaningful (and comparable) observations of social-media-enabled relationships in practice. The purpose of establishing useful measures of social media effects is two-fold; from a practitioner standpoint, it addresses the adage, “you can’t manage what you can’t measure.” Supplying organizations with pragmatic, theory-driven metrics will enable managers to evaluate the consequences of social media campaigns in relation to overall business performance and allow them to manage social media strategies from positions that are less reactionary and more grounded in established knowledge or theory. From an academic standpoint, in order for our accumulating knowledge in this emerging domain to advance from observation and description to theory development and testing for the purposes of explanation and prediction, we must have a foundation of theoretically justified measures. This paper lays the groundwork for the development of such outcomes by establishing an analytical model that dissects the phenomenon into components and then ties the conceptual underpinnings of those components to theory.

Our paper begins with a discussion of the scope of the social media landscape; we aim to convey the magnitude of complexity introduced to stakeholder interactions by social media technologies in the first two sections. We situate our current firm-centric study in the broader social media environment, specifying the stakeholders and

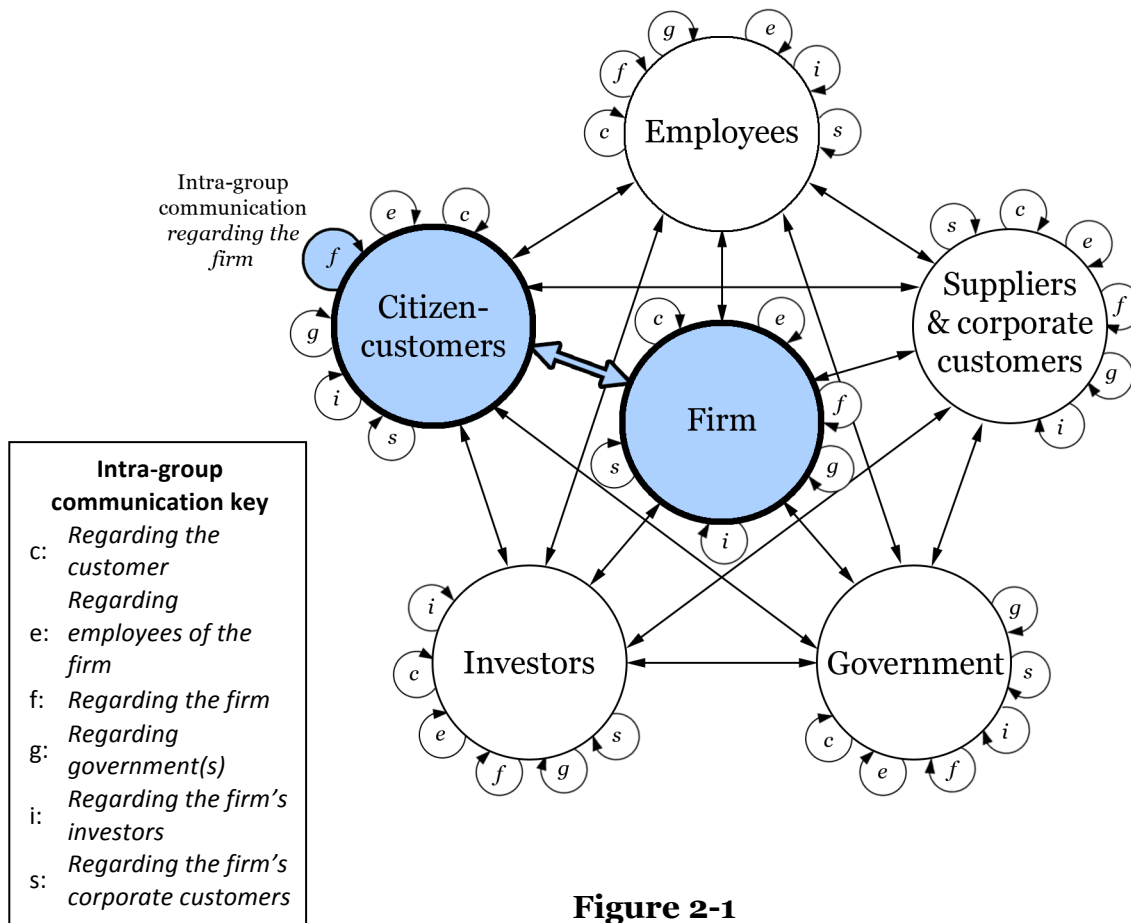
interactions relevant to our current research questions of interest. Based on this discussion, we propose how researchers might go about developing a set of measures for social media effects, and conclude with a discussion of how such measures might apply to other relationships within the larger social media milieu.

## **2.2 Scope of the Problem: the Social Media Ecosystem**

Theorizing about social media effects is an important, albeit nascent, concern for IS research. This technology-enabled phenomenon changes the nature of traditional relationships in an organizational context, a transformation that organizations must address in order to fully compete with rivals in an era of widespread social media communication. Historically, enterprises have achieved certain goals regarding their customers through unilateral, one-to-many channels such as print, radio, television, and more recently the Internet, broadcasting carefully-controlled messages of persuasion with limited opportunities for reciprocity (Berthon et al. 2008). However, the advent of social media technologies has altered this dynamic by enabling a high degree of two-way dialogue between the organization and its customers, as well as by providing a mechanism for customers to collaborate amongst themselves. Consumers can suddenly participate in the efforts to create and share knowledge about a company's products and services, a process that simultaneously conveys to the company potential risks such as negative word of mouth marketing (Fournier and Avery 2011) as well as opportunities such as gaining competitive advantages (Cook 2008) in the forms of collaboration-based productivity (Soriano et al. 2007) and customer-driven innovation (Tapscott and Williams 2008).

Our current study focuses on the stakeholder dyad of citizen-customers (which we will shorten to “customers” for the purposes of this paper) and the firm, seeking to understand, in order to measure, the effects of social media within a business-to-consumer (B2C) framework. While this portion of the social media ecosystem is the most relevant to us as scholars concerned with business systems, it is nonetheless important for us to call attention to the magnitude of the social media landscape as a whole (see Figure 2-1). Figure 2-1 includes a map of all stakeholders (represented by large circles) from a firm’s perspective that might interact via social media. In addition to inter-stakeholder communications (e.g., government-to-corporate supplier, employee-to-investor), members of each stakeholder group can also communicate with one another in what we call intra-group communication. While inter-group exchanges are easily understood as being initiated by one group and directed toward a recipient group, distinguishing among various possible intra-group exchanges is more difficult. As such, we opt to differentiate intra-group exchanges based on the subject of the exchange. For example, employee-to-employee discussions of a government mandate (i.e., intra-group employee communication regarding the government) are conceptually distinct from employee-to-employee discussions regarding investor relations (i.e., intra-group employee communication regarding the investor) according to our model.





Given these distinctions, among the six stakeholder entities specified in the model there are fifteen possible two-way inter-group interactions and thirty-six possible intra-group interactions (each focused on a different stakeholder object) that could combine to form at least 540 different communication configurations. While some of these configurations may not make practical sense, the framework is still available for a given researcher to determine which entities, and in what combination, might be worthy of investigation. Other researchers might derive useful knowledge regarding the role of social media by carving out other portions of the ecosystem to scrutinize. For example,

political scientists might examine the function of social media in the 2011 Egyptian uprising by measuring citizen-to-citizen communication regarding the government. Various configurations of inter- and intra-group communications imply different ramifications for a range of policies—in the current study, our concern is firm-level social media strategy, but other reasonable outcomes might include social media campaign strategies for politicians (e.g., Wattal et al. 2010), health communication strategies for public health organizations (e.g., Chou et al. 2009), or management strategies for disclosing financial information to investors (e.g., Xu and Zhang 2009).

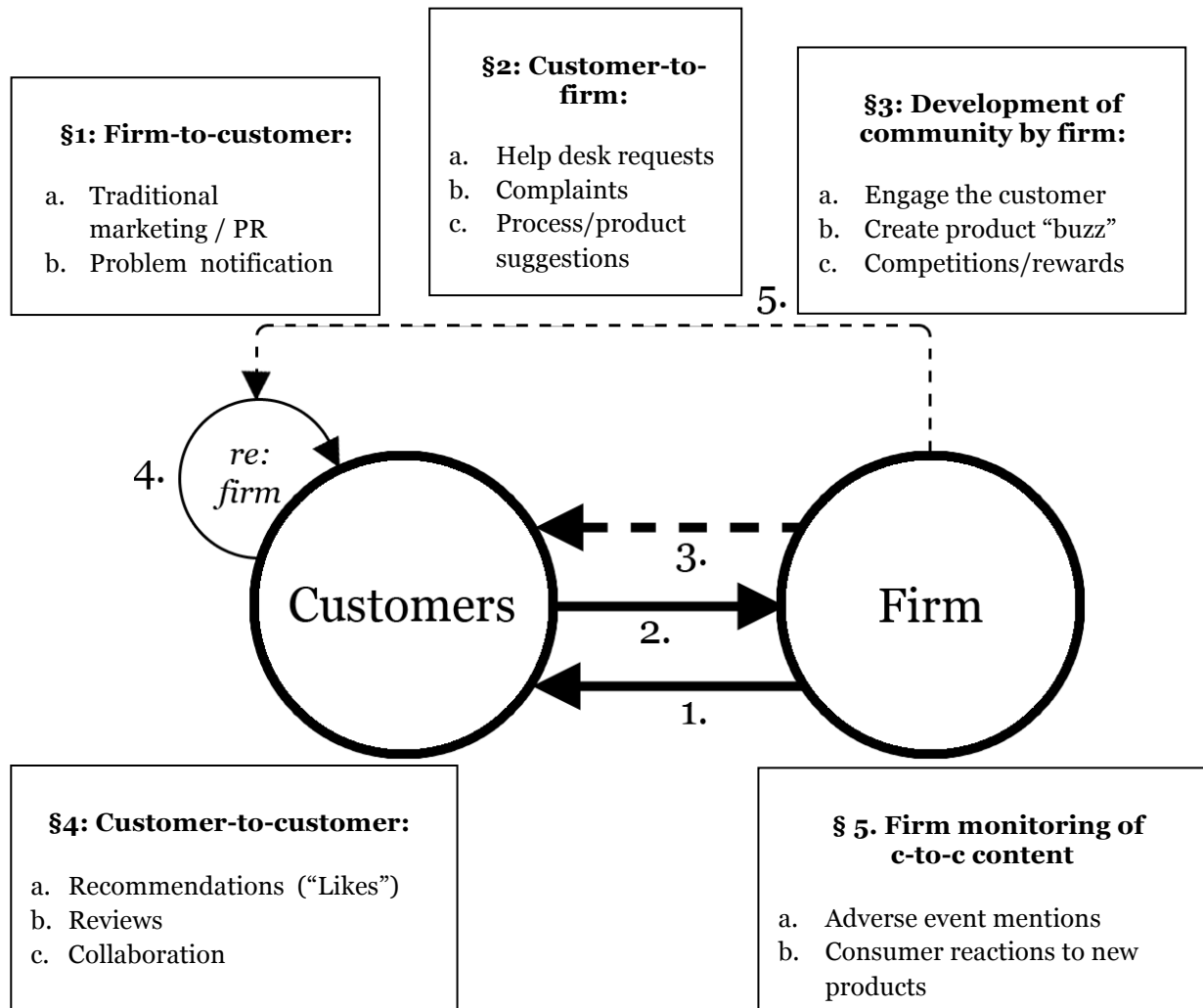
In the current study, our research interests pertain to how firms can measure the success of their social media efforts with respect to their customers. In addition to reciprocal exchanges between the firm and its customers, we also examine customer-to-customer interactions, restricting our focus to those communications pertaining to the firm in order to hone our model to a parsimonious yet predictive set of measurements. Defining this construct to include general mentions of the focal firm and its brands, products, services, and competition, we conjecture that this set of intra-group customer exchanges will shed far more light on the ultimate dependent variable we seek to understand (firm performance) than intra-group customer exchanges regarding the firm's employees, investors, corporate suppliers/customers, or governing bodies. We do not exclude the possibility that intra-group customer conversations regarding these additional stakeholders could help predict some variance in firm performance in some situations, for example in measuring employee behavior as customers complain or complement interactions with particular employees; we simply believe the relative importance is low compared to communications pertaining to the firm.

Conversely, we also exclude firm-to-firm intra-group interaction regarding the customer as not germane to our study because it implies some form of inter-organizational relationship (IOR) among individual firms; while this may yield an interesting level of analysis for future related studies, our immediate concern is to understand the social media interactions between a focal organization and its human customers (as opposed to organizational customers) and figure out how best to relate those to firm performance. Although we exclude external firms from our focal dyad, we do include customer mentions of external firms—intra-group customer-to-customer communication regarding competitive firms—in our measurement schema for the logical reason that criticism or praise of a competitors' products or services is likely to inform a firm's competitor analysis, which in turn suggests probable ramifications for firm performance.

Given the definitional rationalizations presented here, we offer the caveat that future studies seeking to examine additional effects of social media interactions among other stakeholders in the ecosystem should carefully specify definitions of each group of interest, particularly when including the organizational-level entities of firm, supplier, or government. Restricting "firm" to represent a single firm versus allowing it to vary as a network of firms, deciding whether "government" will embody a singular governing body (e.g., local, state, national, or corporate) or multiple nested or networked administrations (e.g., national governments of all countries in which a multinational firm operates), and defining "suppliers and corporate customers" as a single partner, a specific industry, or all possible suppliers, are definitional decisions that will affect the external validity of results.

### **2.3 Focus on the Customer-Firm Social Media Dyad**

Isolating our model of interest (shaded sub-model, Figure 2-1) from the overarching social media ecosystem allows us to unpack the range of social-media-enabled activities transpiring between customers and the firm (to which will we refer henceforth as the customer/firm social media dyad, or “B@C” dyad, to denote the representation of B2C, C2B, and C2C interactions, for short). Expanded in Figure 2-2, each layer of stakeholder-initiated activity within the B@C dyad is driven by a different set of goals, and is thus potentially ascribable to different theoretical bases for the purpose of measurement. By decomposing the complex configuration of social media interactions into its constituent relationships, we are able to simplify it into a stratified system of measurable, manageable processes. It should be noted that arrows 1 - 3 in Figure 2-2, while appearing to visually denote unidirectional messages from one stakeholder to the other, each imply an initial message (cause) as well as some type of response (effect), be it a literal response such as a message back to the initiator or a set of behavioral reactions such as a visit to a web site, a refund or product replacement, a blog posting, a purchase, etc.



**Figure 2-2**  
**Firm/customer (B@C) social media dyad**

Examining the layer of firm-initiated interactions with the customer (Figure 2-2, §1), further isolation of the exchange into two unidirectional paths (meaning we do not look at reciprocity but focus strictly on messages in one direction and then the other) results in pathways that suggest a traditional Internet marketing model, whereby firms efficiently channel advertisements and persuasive promotional messages to their customers via the Web (Hoffman and Novak 1996), and customers respond by following the firm’s suggestion to visit an e-commerce site (or brick-and-mortar location, when

appropriate). The types of tasks initiated in this firm-to-customer layer include advertising new products to customers, flash-promoting time-sensitive discounts or limited-availability goods, and otherwise targeting customers with specific messages of tailored interest, a practice shown to increase profits by increasing differentiation in the market and eliminating extraneous advertisement to unsuited consumers (Iyer et al. 2005). As such, we are able to map some “firm-to-customer” initiatives (§1.a) to traditional web-based marketing and PR activities, allowing us to refer to the existing literature in these traditions for suitable measurements for assessing this component of the social media landscape. The role of web-mediated advertising is well established in the IS and marketing literatures, with strong theoretical foundations and time-tested measurements (e.g., Berthon et al. 1996).

Encapsulated by this same layer (§1.b), firms may also initiate pursuits toward the customer intended to achieve some aspects of customer service. The type of activity serving this objective concerns customer notification—e.g., notifying patrons about potential problems (as in urgent safety-related recalls) or impending service interruptions.

A complementary layer of the B@C dyad that further harkens to the province of customer service is the reverse-oriented set of customer-initiated service requests directed toward the firm (Figure 2-2, §2). Firms have long employed a variety of digital systems supplemented by human service to enable customers to seek product or service support (Ba et al. 2010); such structures include call centers, web-based self-service systems, and email correspondence (Featherman et al. 2006). Whether initiated by the firm or by the consumer, customer support facilitated by social media IS can be appropriately described and measured by looking to scales within the comprehensive

body of customer service and quality work. While the focal IS medium of interest may present a novel mode for communicating and certainly implicates a complex mesh of goals, activities, and participants, once the customer-service oriented tasks are isolated from the overall phenomenon we find a set of interchanges that can be understood from the conventional perspective of IS-enabled customer service.

The combined set of firm-initiated and customer-initiated service requests depicted in Figure 2-2 can be understood by returning to the stream of literature launched decades ago in which customer service has been acknowledged as a strategic imperative for most firms (Parasuraman et al. 1985), the most critical factor in the quest for customer satisfaction (Ray et al. 2005), and a fundamental driver of IS priorities (El Sawy and Bowles 1997). Customer service measurements have been established in a variety of contexts, a recent study linking IS-driven customer service to improved firm performance applying especially neatly to the context at hand (Ray et al. 2005). The Ray et al. manuscript observes social complexity of IS capability as a critical explanatory factor of performance, informing our notion of social media-enabled customer service as a mechanism that firms may exploit in the quest for improving their bottom line.

Having categorized the first two segments of B@C activity as sets of tasks well understood and measured in the IS and marketing literatures, we turn our attention to three additional layers that embody social media's novel contribution to the B@C relationship. Our goal in the following section is to shed light on the implications and opportunities that these contributions convey to both firms and consumers. The ultimate goal of deriving useful theory-based measurements of social media effects that predict changes in firm performance hinges on thoroughly understanding the novel

modes of interaction that social media propagates and comprehending the variety of natures and drivers for these interactions.

### **2.3.1 Measuring layers of social media-enabled B@C activity**

The granular layers of B@C social media activity include actions that can be understood either as events or processes, the latter of which have intermediary as well as ultimate effects that should be monitored (and thus measured) by the firm. For example, firm-initiated community building through which the firm attempts to influence customer-to-customer exchange is actually a series of events and outcomes (see Figure 2-3). As customers interact with one another, the firm is able to monitor and derive from these exchanges useful knowledge as the process unfolds, while the end result of the series of exchanges is also another measurable outcome. The crux of the social media measurement problem as we see it is deciding what aspects of customer/firm social media interaction ultimately relate to firm performance, and which have no bearing on firm performance and so do not need to be monitored. The following subsections lay out the activities that should be considered for measurement but that are not conveniently described or operationalized in extant literature.

#### **I. Firm-to-customer community building**

Due to the collaborative functionalities enabled by the multitude of social media applications and technologies embraced by consumers, the ability now exists for firms to influence consumer behavior in unheralded ways. By engaging customers in a “social” experience revolving around the brand, firms are able to develop brand-centric communities in such a way that ties customers to their products (Figure 2-2, §3). The marketing literature cites brand communities as not only a driver of loyalty and a factor



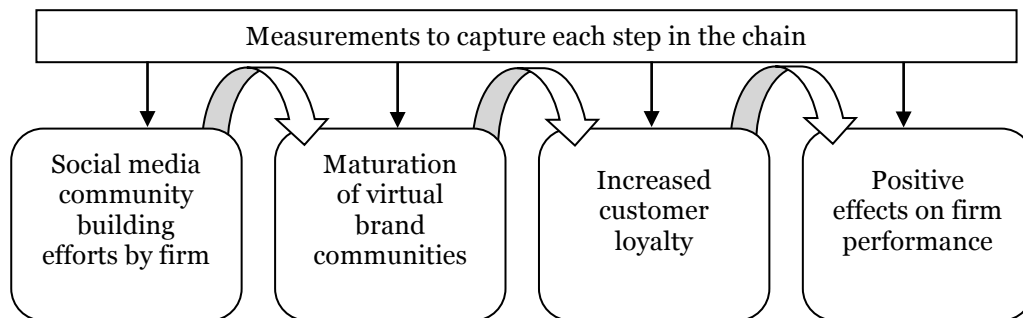
that increases a consumer's likelihood of adopting a new product from the preferred brand, but also as a basis for oppositional loyalty against competitors' products (Thompson and Sinha 2008). The array of firm innovation in this domain is expansive; companies are continually inventing novel approaches to creating "buzz" about events and services, conducting competitions, and facilitating reward systems.

Of course, organizations have been seeking the "Holy Grail" of brand loyalty through the development of communities for decades (McAlexander et al. 2002), long before the advent of social media. Defined by the commonality of its members and the relationships among them, a community is a network of social relations while a brand community is a specialized social group organized around a particular brand that exhibits shared consciousness, rituals and traditions, and a sense of moral responsibility (Muñiz Jr. and O'Guinn 2001). Brand community has evolved from being conceptualized as a customer-to-brand relationship (e.g., Aaker 1997) to a customer-brand-customer triad (Muñiz Jr. and O'Guinn 2001) to a network of relationships including customer-to-brand, firm-to-customer, and intra-customer interactions (McAlexander et al. 2002), a configuration that strongly resonates with the makeup of the social media ecosystem.

Reinvigorated by two-way conversations with customers, the ability to collect in-depth records of consumer preferences, and the power to "micro-target" or address customized messages to individuals—all tasks that have been simplified by the existence of social media technologies—firms are now turning to social media outlets as leverage for shaping brand-centric communities in ways previously unrealistic with traditional mass media (Fournier and Avery 2011). The existence of brand communities draws on one of the most basic human motivations, the desire to belong to a larger collection of

likeminded peers, to fit in, to be accepted; with brand consumption serving as the basis for coalescence and social media facilitating the connectivity, firms have an unprecedented platform for exploiting consumers' basic drives (e.g., to "belong") in such a way that benefits the brand or product (Fournier and Avery 2011).

Considering the chain of influence linking brand communities in the retention of consumers via the mechanism of increased brand loyalty, which in turn positively impacts a firm's bottom line (see Figure 2-3), firms need measurements to help them monitor their community-building efforts. A valid system of measurement for firm-directed social media efforts must account for all crucial relationships determined to comprise the construct of brand community from a customer-experiential perspective: relationships between the customer and the brand, the customer and the firm, the customer and the product, and intra-group customer-to-customer relationships (McAlexander et al. 2002).



**Figure 2-3**  
**Community-building-to-firm-performance**  
**chain of constructs**

Given that building brand communities is a valuable facet of deriving value from social media, firms must be heedful of choosing a community-building strategy that

resonates with their social media capabilities and overall goals. Depending whether it is a laissez-faire approach in which the firm preemptively renounces designs on steering the social media behavior of its consumer base out of respect for its autonomy, an appropriative approach whereby the firm waits to take its cues from its consumer base and then jumps in to take advantage of the content created by the participants, or a more dominating approach meaning that from the outset, the firm actively attempts to mold the social collective of its consumer base by orchestrating calculated campaigns, a range of consequences may ensue. The literature points to the pros and cons of each of these routes, ranging from the benefit of preserving authenticity of fan-created content by remaining hands-off according to the first approach, to the risk of inviting caustic parodies from hyper-critical consumers despite attempts to heavily guard against such possibilities, according to the last approach.

On one end of the spectrum it has been suggested that the successful firms (at least in terms of reaping the benefits of social media) are the ones that cede jurisdiction to consumers despite the difficulties inherent in relinquishing control. This may be attributable in part to the respect this relinquishment signals to consumers' regarding their autonomy and influence over user-generated content. On the other hand, extremely clever marketers have managed to design campaigns that clandestinely enable spoofs, identified by particularly savvy firms as desirable due to the high viral currency and ultimate cultural resonance such "hits" often indirectly effect (Ferguson 2008; Fournier and Avery 2011). Although viral tactics have been accused of merely resulting in short-term attention, it may be possible for firms to leverage such messages in building customer loyalty by launching (or covertly instigating) campaigns that

ultimately beget consumer identification with other like minds and promote “sticky” dialogue (Ferguson 2008).

However successful companies may be at instigating such marketing campaigns, viral as well as word-of-mouth marketers face the same problem in measuring the effects of their campaigns. Researchers know that these types of advertisement build brand awareness, but they are unsure how to calculate the effect on market share (Ferguson 2008). Expanding awareness into loyalty via the development of brand/product/service communities is a critical driver for this segment of the B@C Social media dyad, given the ultimate ties of this activity to firm performance. Beyond the measurement of community building efforts by firms, accounting for additional variations introduced by the existence of viral messages is especially difficult, especially given the lack of understanding by practitioners and academics alike as to how one might successfully foment an effective, positive viral campaign. Until such elusive antecedents are more thoroughly accounted for, it is likely that community-building efforts will be measured in terms of more conventional components. This is not to say that viral or word of mouth effects cannot or should not be ascertained; we simply conclude that efforts to produce such effects should not be included in community building measures.

## **II. Customer-to-customer exchange**

Once a product or service enters the marketplace, it is ripe for inclusion in customer-to-customer social media interaction. This may take the form of a consumer commenting on or reviewing a product, service, or event within the comments section of a blog for the perusal and reaction of other consumers, clicking the “recommend” button on product’s page within any number of e-commerce sites with integrated social media

functionality, “tweeting” about an experience to a network of “followers,” or staking a claim on a product within Facebook or similar social media networking application by “liking” it, thus joining the ranks of that product’s “fans.” Customers may focus messages directly at one another, contribute to collaborative social media sites such as Wikipedia (a collective online encyclopedia), Kaboodle (a forum for compiling public shopping lists), or IMDB (an actor/movie information database), or broadcast helpful information through a variety of online product/business review (e.g., Epinions, Yelp) or news recommendation (Fark, Yahoo! Buzz) sites.

In fact, it is irrelevant whether a firm actively sponsors a social media community or not; once a product is accessible to consumers to purchase or experience, it in turn becomes a viable candidate for customer-to-customer discussion. Firms may become implicated in online word-of-mouth “advertising” whether or not they have designed a corresponding strategy or ever intended to enter that realm in the first place. As a corollary, firms do not have the luxury of opting out of the customer-to-customer information market; the choice becomes whether to actively plan to influence how and where some of the “conversations” occur by building social media communities to supplement existing outlets, or to completely relinquish control and let customers fully determine the context in which the firm’s product and services are critiqued or recommended. Even if a firm opts for the former and creates a blog or competition site to attract customers, all the usual suspects of third-party social media outlets remain, for the most part, outside of the firm’s control. As such, the “portfolio” of social media outlets pertaining to a given product or service will include a wide range of non-firm-controlled entities supplemented with whatever internally-directed channels the firm

opts to host, suggesting that the overall set of customer-to-customer interactions will remain outside the control of the firm.

Although firms largely lack power to regulate the customer-driven content within social media applications, they gain an enormous wealth of public, monitorable, analyzable data. We propose the capability of firms to monitor intra-group customer exchanges to be one of the biggest sources of benefit to firms introduced by social media, and a driver of the need for measurements of customer-to-customer exchanges. While a desirable system of measurement of such exchanges would certainly account for simpler characteristics like counts of awareness (e.g., number of “likes” and “recommendations” of a product or service promulgated throughout the network of social media instantiations), more complex analytical capabilities should also be incorporated. Some type of semantic differentiation mechanism—i.e., analysis of positive comments versus negative comments—and, ostensibly, some form of deeper interpretation, capable, for example, of detecting sarcasm, spoofing, or other types of behavior likely indiscernible by more simplistic modes of analysis should also contribute to measurements adopted by firms.

In keeping with our model’s scope, including the specification that customer-to-customer communication regarding the firm should include mentions of competitive firms and their products when appropriate, a useful system of measurement should account for as many of these factors in relation to competitive firms or products as possible. For the most part, all of the data available about the monitoring firm should also be harvestable about competing firms, since the bulk of consumer-generated content is available freely on social media sites across the Internet; the only data possibly obscured from collection would be comments and collaborations facilitated

within the sphere of a competitor's internally hosted social media site. For example, a blog or virtual community that is password protected and mediated by a human approver may contain content that is unobtainable by scraping scripts or other mechanisms; such locked-down data would simply be excluded from measurement (though, a firm's own internally-hosted data would likewise be unavailable to its competitors who might attempt to gain the same form of competitive intelligence).

### **III. Firm monitoring of customer-to-customer exchange**

The unique opportunity conferred on firms by the ability to monitor customer-to-customer exchange is not limited to analyzing huge streams of data, although that is an enormous source of potential advantage. Monitoring customer-to-customer streams also imparts to firms the capacity to interject customer service into negative exchanges, thus influencing customer satisfaction and opinions and derailing potential public relations problems. Companies ranging from Comcast to Jet Blue monitor outlets like Twitter for any mention of their company, searching for opportunities to provide information to needy customers or correct misinformation subject to inadvertent propagation by members of their consumer bases (King 2008).

An interesting risk factor arises in the B@C social media dyad in the form of non-social media. Traditional or "legacy" media outlets that existed prior to the advent of the Internet, including public broadcasting, newspapers, magazines, and network newscasts, perform a unique function in the social media landscape. Specifically, traditional media outlets serve as an amplification mechanism, especially (but not strictly) within the customer-to-customer segment. It is not uncommon for a news outlet to become aware of customer-firm discord unfolding in a social media setting, often available for general consumption when dissatisfied customers broadcast their

service problems to other consumers in the pursuit of a) making peers aware of potential problems with certain brands or services and b) garnering peer support in the fight against whatever the focal complaint may be. Whereas such a complaint may or may not attract mass attention within the social media context, once it is detected and amplified outside of the social media arena, it becomes available for true mass consumption. Studies indicate that consumers multi-task in their media consumption, simultaneously participating in online and traditional modes of information intake (Russell 2010); whereas customer service complaints may not reach viral mass within social media, once it becomes supplementarily available to through traditional news, an amplificatory effect is likely. Participants can turn to their social media outlets to expand. Conversely, while firms must guard against the risk of a negative message becoming amplified, this mechanism can always work in a firm's favor when the message being amplified favors the firm, essentially serving as free PR.

However, probably the most important characteristic of the B@C social media dyad that lends itself to firm exploitation is the colossal stream of real-time customer-to-customer interchanges that are publicly facilitated by the myriad social media applications in operation daily. These data, which firms can ostensibly interpret to acquire clues about customer likes and dislikes, trends in the marketplace, changes in technology use—the list of derivable intelligence is constrained only by firms' imaginations—is out there in the ether to be analyzed. The literature suggests that this intra-group dialogue can yield customer insight as well market intelligence (Gallaughier and Ransbotham 2010); what academics and researchers alike lack is the foundation of measurements that can derive useful meaning from trends over hundreds of thousands or millions of these data points. While a single person or team can monitor for possible



negative mentions of a brand or firm, firms invite peril when they give too much credence to extreme positive or negative feedback from a vocal but small faction of overall customers (Fournier and Avery 2011). Being swayed by extremes does not entail the reliability inherent in detecting patterns across the comprehensive base of customers communicating via social media. This ultimate objective remains to be established, and motivates our proposal to suggest how available data can be analyzed in its entirety.

#### **IV. The role of digital data generation**

A unique antecedent to the processes that contribute to social media's value to the firm, the ability to monitor and measure is facilitated by the generation of digital data. As social media interactions are computer-mediated and occur within the infrastructure of the Internet, firms are able to compile stores of all interactions for the purposes profiting from customer data (Piccoli and Watson 2008), in this case in the form of customer-to-customer and firm/customer interactions with the ultimate goal of increasing firm performance. Comparable to the capture of customer transactions, firms are able to record six critical details about each interaction: when the interaction occurs (i.e., time/date stamp), where (i.e., within which particular social media application), the nature of the interaction (i.e., is it a persuasive customer-to-firm message?), how it was executed (i.e., Facebook "like" button click, wall posting, or personal message?), who initiated the interchange and to whom it was directed (i.e., firm-initiated toward the customer, customer-initiated toward the firm or to other customers), and the outcome (i.e., strengthened brand community, alerting customers about a potential problem with a particular product) (Piccoli and Watson 2008).

Whereas manual versions of these processes are laborious, time-consuming, and thus expensive to execute or analyze (e.g., transcribing customer complaint calls, printing and distributing paper-based advertisements), when these processes are accomplished digitally they become immediately available, cheap to analyze, and abundant sources of intelligence. The low cost and high analyzability of data makes it available and valuable to customers and firms, factors particularly germane to the context of social media. For example, digital video recorders and editing software are very inexpensive and accessible to a wide range of amateurs who would have found creating, editing, and broadcasting video prohibitively costly and cumbersome just a few decades ago; further, channels over which to broadcast video were not accessible to the masses the way the Internet is today. But the ease with which anyone can record and publicly circulate video messages (or any other type of electronic signal) via social media today means that the flow of digital information is enormous and ripe for the picking. Once social media interactions are recorded as data, firms are then in a position to turn streams of these data into information through measurement techniques; the trick is to determine which aspects of these data should be analyzed and compared, and how that might be accomplished. We approach this decision by examining stakeholder goals driving social media activities; from this understanding we can a) know which areas of research to look to for theoretically-justifiable measures, and b) start to ascertain which activities are important for firms to monitor and which are irrelevant to the ultimate end of firm performance.

### **2.3.2 Initiator goals across layers**

A complementary angle for approaching the task of fully explicating the B@C social media dyad is to examine the underlying goals motivating the events and processes within each layer. Returning to the advertising/marketing/PR literatures, two very important effects include the traditional factors of 1) increased awareness and 2) subsequent persuasion (e.g., Keller 1993), both of which map to the uses of social media. Firms have been able to jump right into the social media scene to achieve these objectives because financial and technical barriers are low. However, while awareness and persuasion are important antecedents of market share and certainly pertain to a variety of both firms' and consumers' social media uses, they do not tell the whole story. The additional function of "collaboration" is a third distinguishing characteristic of a large percentage of consumer-to-consumer and consumer/firm interactions that do not fulfill the purposes of persuasion or simply increasing awareness. Falling outside the scope of most traditional marketing models, "collaboration" introduces a whole new set of considerations that must be factored into the development of an accurate and useful social media measurement system.

We briefly discuss these three pervasive objectives to which we refer as "Initiator Goals" and present in tabular format (see Table 2-1) these three goals crossed with the five layers of social media activity previously mapped out in Figure 2-2 (i.e., firm-to-customer, customer-to-firm, and customer-to-customer interactions, plus the additional firm pursuits of community building and customer-to-customer monitoring).

We populate the table with descriptions of the activities occurring at each intersection of goal  $\times$  initiator, then map to each cell relevant areas of literature (see Table 2-2) in order to frame each activity in terms of academic conversations that can

inform our understanding of each segment of activity. Drawing on extant bodies of established work in these theoretical realms serves us twofold; first, it allows us to capitalize on decades of accumulated scholarly knowledge in our attempt to understand, explain, and measure aspects of the social media phenomenon; reciprocally, it allows us to contribute to organizational science by expanding the reach of established theory to the novel, yet pervasive and evidently irrevocable, IS environment of social media.

**Table 2-1**  
**Focal dyad activities according to goal**

| Initiator Goal |               | Activity Initiator   |   |   |   |  |
|----------------|---------------|--|---|---|---|--|
|                |               | Firm (community building)  | Customer (to customer)  | Firm (to customer)  | Customer (to firm)  | Firm (monitors customer-to-customer)   |
|                | Awareness     | Ω Firms promote virtual brand communities in order to increase customer knowledge about the brand.   | □ Customers make one another aware of products and services, an important antecedent of market share complicated by the moderating role of sentiment. | + Firms alert customers about new products and services.<br>‡ Firms alert customers about potential problems. | ‡ Customers make firms aware of product flaws.  | ‡ Firms become aware of consumer dissatisfaction and adverse events by monitoring customer-to-customer interactions.             |
|                | Persuasion    | Ω Brand communities facilitate “sense of belonging;” this supports brand loyalty that in turn advances customers’ propensity to repeat purchase. | □ User-generated reviews of products and services sway subsequent consumer purchases. Negative persuasion is a possibility here.                      | + Firms market products and services to customers persuasively.   | ‡ Customers directly campaign the firm for product updates or changes.  | ‡ Customer-to-customer interactions change firm behavior when firms intercept conversations conveying the need for intervention. |
| Initiator Goal | Collaboration | Ω Firms collaborate with customers in an effort to advance identification with a particular brand community.                                     | ◇ Social media participants collaborate with one another to achieve entertainment and bonding.  | □ Firms engage customers in collaborative projects from which both parties gain value.                        | □ Customers initiate collaboration with firm to create knowledge, for the satisfaction of co-creating products. | □ Firms monitor customer-to-customer collaborations for opportunities to derive value.   |

**Table 2-2**  
**Areas of applicable literature with example citations**

| Online word of mouth advertising (□)               | Online marketing and PR (+)                 | Web-delivered customer service (‡)        | Customer loyalty, brand community (Ω)                                   | Entertainment, bonding (◇)         | Co-creation of value (□)                           |
|--|---|---|---|------------------------------------|--|
| Godes and Mayzlin 2004; 2009; Kozinets et al. 2010 | Chatterjee et al. 2003; Stewart et al. 2001 | Shankar et al. 2003; Zeithaml et al. 2001 | Algesheimer et al. 2005; McAlexander et al. 2002; Muñiz Jr. et al. 2001 | Ko et al. 2009; Whitty et al. 2007 | Berthon et al. 2008; Etgar 2007; Lewis et al. 2010 |

Our set of activities, which we consider comprehensive if not exhaustive, derives from a review of social media literature augmented by informal discussions with social media marketing practitioners. We suggest that as a whole, the objectives underlying social media activities can be understood within the contexts of online word of mouth advertising (Godes and Mayzlin 2004, 2009; Kozinets et al. 2010), online marketing and PR (Stewart et al. 2001; Chatterjee et al. 2003), web-delivered customer service (Zeithaml et al. 2001; Shankar et al. 2003), customer loyalty/ brand community (Muñiz Jr. et al. 2001; McAlexander et al. 2002; Algesheimer et al. 2005), entertainment and bonding (Whitty et al. 2007; Ko et al. 2009), and co-creation of value (Etgar 2007; Berthon et al. 2008; Lewis et al. 2010). The following sub-sections review these conversations in the academic literature and argue applicability of each to the corresponding layers as proposed.

## **I. Awareness**

A variety of social-media-based activities achieve the goal of increasing awareness, whether of the firm (by the customer), the customer (by the firm), or peers (by other customers). Stakeholders may become aware of a new product, service, or event or of an existing or potential problem, Awareness may be accomplished directly

(by express contact with one stakeholder by another) or indirectly (via monitoring customer activity).

Within the process of community building, awareness is a first step for firms to take toward developing customer loyalty. Within the context of customers making one another aware of products, services, etc., we see online word-of-mouth effects occur. When firms use social media to make customers aware of new products or services, we can understand this as traditional online marketing. From the customer's perspective, it is useful to employ social media as an expedient route for making firms aware of product flaws; this accomplishes customer service, especially benefitting the firm (and other customers) in cases where it is necessary to act quickly to diffuse a potential large-scale problem. Through the mechanism of monitoring customer-to-customer interactions, firms are also able to make themselves aware of consumer dissatisfaction and adverse events, which enables them to take appropriate action proactively.

## **II. Persuasion**

Traditionally the main purview of advertising and marketing efforts, attempts to persuade customers to purchase a given product or service are undoubtedly augmented by social media campaigns. Given the caveat that social media is far more than just another conduit for broadcasting one-way messages at consumers, the functionality enabled by social media allows firms to persuade customers, customers to influence firms, and customers to sway one another's opinions and behaviors. The bi-directional communication that characterizes social media interaction is a crucial aspect of the type of relationship-based marketing expected to be a necessary component of future marketing strategies (Andersen 2005).

In addition to traditional conceptualizations of persuasion inherent in the notion of marketing or advertising products and services, customers also now wield a substantial leverage in the relationship back to the firm. Consumers are more easily able to engage firms in conversations as they lobby for changes in products or supplemental services. This can benefit the organization due to the additional value that suggested improvements generate for the firm (Nambisan and Baron 2009), although a firm may not always immediately embrace a customer's desire for product change implementation. As such, in some cases social media further enables persuasion when it facilitates the assembly of groups of customers who can then wield their strength in numbers (Fournier and Avery 2011).

Customers may also unwittingly influence a firm's decisions to implement changes; as firms monitor customer-to-customer conversations across social media applications, the firm may unilaterally decide to make certain adjustments based on the intelligence gleaned from such monitoring. This is a particular facet of the social media world instigating a clear need for measurements to ensure strategies born of analyzable data; the threat to firms making decisions based on gathered intelligence is that their sample may be biased, incomplete, or unreliable. Whether targeted directly by consumers or induced into change due to assessments of customer-to-customer exchanges, firms must ensure that they are not simply giving in to what they mistakenly perceive to be the collective's desires, especially considering the self-interest inherent in such a collective that may be completely unaligned with the best interests of the firm or its brands (Fournier and Avery 2011).

### **III. Collaboration**

Although historically conceptualized as external to the firm, the evolving view of customers as co-creators has brought the customer directly into consideration as firm value generators (Nambisan and Baron 2009); such value might include the benefits of augmented innovation processes and competitive strategies, or direct product or marketing development (Schau et al. 2009). Customers also derive value from collaborating with the firm, although the benefits are of a different, individual nature—it is suggested that customers co-create value with firms in order to derive personal enjoyment, self-promote, and as an outlet for activism (Berthon et al. 2008).

While a social media measurement system needs to account for collaborative activity, this is arguably one of the more complex aspects to capture due to the fact that it is infeasible to break this objective down into asynchronous “cause and effect” paths the way we are able to understand awareness- and persuasion-based activities transpiring at the firm-to-customer and customer-to-firm levels. Furthermore, it involves one of the least explored areas of research listed in Table 2-1 as our major bases for describing social media activity. The co-creation of value is one of the most difficult for which to suggest measures, although it is arguably an extremely important aspect of B@C social media activity for which firms must plan to account.

#### **2.4 The Measurement Problem**

The lack of established metrics in the literature tying social media advertising (or other types of persuasive campaigning) to actual performance speaks to both the need for reliable and valid definitions and measurements of social media, and the difficulty of coming up with such measurements on the fly. In their 2010 analysis of the impact of



Web 2.0 on the 2008 presidential campaign, Wattal et al. turned to Gallup poll standings as a function of traditional media, Web 1.0 (traditional web sites), and Web 2.0 (YouTube, MySpace, blogs), based on mentions of the candidate via each medium the month prior. The decision to lag polling data by a month to connect Internet use to the following month's Gallup poll numbers introduces a possible disconnect between cause and effect in our social media research—this maps to the problem that IS researchers to date have had no empirically derived guidelines on which to rely, exacerbated by long feedback loops. Without basing such measurement decisions on theory or precedent, internal validity may be difficult to prove. It may be questionable, for example, whether polls a month in the future will accurately convey opinions developed today, especially when opinions are formed in response to information mediated by dynamic social systems that provide instantaneous as well as interactive communication.

Additional questions arise regarding the attribution of performance to, in this case, the number of monitored blogs that mention the focal product or person; this particular operationalization may not reveal a great deal of variance across candidates or about the relationship between particular social media strategies and the resulting dependent variable of choice. We reason that counting the number of blogs in a finite set that mention a particular product (or candidate) over a given period of time, especially a duration as great as a month, is unlikely to convey the insight we might glean from some alternate choices. For example, percentage breakdown of total blog coverage per candidate, absolute counts of the number of individual mentions (or discussions) of each candidate across all blogs, and gauges of sentiment of mentions may all represent more fruitful avenues for assessing impact of social media. The number of blogs that

mention product X in a single month may be equivalent to the number of blogs that mention product Y in a single month, while the number of conversations about product X could far exceed the number of conversations about product Y, indicating that the level of analysis must be considered carefully in terms of measurement. This essentially equates to the difference between the amount of useful information we can derive from a multivariate, over a univariate, analysis.

As implied above, we contend that any attempts to measure social media use for purposes of predicting performance should factor in the valence of mentions or discussions regarding a particular product or service. Negative publicity has been found detrimental to a wide range of outcomes including product and brand evaluation (Tybout et al. 1981), consumer preference and purchase activity (Sullivan 1990; Charlett et al. 1995), and net present value at both the individual and the network level (Goldenberg et al. 2007). The impression formation literature is clear on the point that people place more weight on negative than positive information in forming overall evaluations of both people and products (Eagly and Chaiken 1993; Ahluwalia et al. 2000) and that dissatisfied customers discuss their experiences with a greater number of individuals than satisfied customers and thus yield more influence on fellow consumers as a whole (Herr et al. 1991; Charlett et al. 1995; Laczniak et al. 2001). Consequently, measuring the number of customer conversations about a product or service facilitated by a social medium without regard to content may lead researchers to draw erroneous conclusions about the relationships under scrutiny. In an organizational context, we cannot assume a positive linear relationship between the frequency of customers' social media use and the firm's desired outcomes with regard to its

customers such as satisfaction because use may comprise negative or positive sentiments, implying opposing effects of asymmetrical magnitude.

A critical step in defining a social media analytics framework is to decipher, ultimately, which things actually matter to the firm, meaning which activities are worth a firm's time, efforts, and financial resources to bother monitoring.

## **2.5 Discussion and Implications for Future Research**

Moving forward with this stream of research, our next imperative is to establish a system capable of accommodating a large scale analysis of data in order to identify overall trends and patterns. Desirable characteristics of a social media measurement system, we suggest, should include 1) accuracy, 2) actionability, meaning that firms can change their goals based on the information extracted from the measures, 3) ability to accommodate multivariate data, meaning it enables complex analyses of multiple variables in order to identify relationships, 4) economic feasibility, in that the cost of measurement is less than resulting benefits, and 5) high orthogonality of measures, meaning we want to avoid multiple measures that capture very similar information.

In Table 2-1, we have identified the activities transpiring across the B@C dyad, mapping them to broad domains of research. The next step in the subsequent series of research efforts requires that we look at the activities in each cell and, referring to the respective research domains, decide how to best measure the particular exchanges. It should be noted that often, the activity captured in a given cell is part of a linear sequence of steps (see Figure 2-3), so a system of measurement should also record such linkages.

A firm is subject to multiple social media streams that can be described by multiple measures, and in turn, a given measure is likely to pertain to multiple social media. This second factor enables us to exploit economies of scale in our measurement efforts, meaning that a particular measurement may describe a component that factors into several different types of social media types, and would thus pertain to the entire category instead of just a single technology, like in a factor analysis. We propose the measurement of change in performance will take something of the following form:

$$\delta P = \sum (a_{ij}x_{ij}) + \varepsilon$$

which sums the effects of all media  $i$  per all measures  $j$  for a change in performance,  $\Delta P$ .

The problems we face in measuring the effects of social media stem, in part, from the need for clearly defined objectives. To this end, it is critical that we craft a measurement system with respect to the context that defines our study. Without context, measurements of various effects lack meaning; without being tied to specific goals, metrics are likewise futile. Ultimately, we envision firm performance as the dependent variable of concern, since the focal stakeholder of our context is the firm.

As noted in a variety of studies intended to predict firm performance, a fundamental problem in researching the effects of such variables as advertising is isolating them from competing or supplemental effects (Berthon et al. 1996). Distinguishing between the direct and indirect effects of such factors is difficult. This same concern pertains to any attempt to tie social media-facilitated interaction to firm performance. A decade and a half ago Berthon et al. (1996) stressed the importance to firms of establishing specific communication objectives for their Web initiatives and identifying measurable means for establishing the success of such ventures; we co-opt

that advice today as completely applicable to the realm of social media enterprise. Their observations regarding the ease with which Web-mediated efforts, or in this case social media effects, are measured combined with a far shorter feedback loop than many other non-digital efforts, encourage our expectations for deriving valid, actionable, reliable measurements of social media that we can ultimately connect to firm performance.

## 2.6 References

- Aaker, J. L. 1997. "Dimensions of Brand Personality," *Journal of Marketing Research* (34:3), pp. 347-356.
- Aakhus, M., Ågerfalk, P. J., Lyytinen, K., and Teeeni, D. 2011. "Call for Papers: MISQ Special Issue on Information Systems for Symbolic Action: Social Media and Beyond," *MIS Quarterly*.
- Agarwal, R., and Karahanna, E. 2000. "Time Flies when You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage," *MIS Quarterly* (24:4), pp. 665-694.
- Ahluwalia, R., Burnkrant, R. E., and Unnava, H. R. 2000. "Consumer Response to Negative Publicity: The Moderating Role of Commitment," *Journal of Marketing* (37:2), pp. 203-214.
- Algesheimer, R., Dholakia, U. M., and Herrmann, A. 2005. "The Social Influence of Brand Community: from Evidence European Car Clubs," *Journal of Marketing* (69:3), pp. 19-34.
- Andersen, P. H. 2005. "Relationship Marketing and Brand Involvement of Professionals through Web-enhanced Brand Communities: the Case of Coloplast," *Industrial Marketing Management* (34:3), pp. 285-297.
- Ba, S., Stallaert, J., and Zhang, Z. 2010. "Balancing IT with the Human Touch: Optimal Investment in IT-Based Customer Service," *Information Systems Research* (21:3), pp. 423-442.
- Barnes, N. G. 2010. "The Fortune 500 and Social Media: A Longitudinal Study of Blogging, Twitter and Facebook Usage by America's Largest Companies Marketing Research," University of Massachusetts Dartmouth Center for Marketing Research.
- Bernoff, J., and Li, C. 2008. "Harnessing the Power of the Oh-So-Social Web," *MIT Sloan Management Review* (49:3), pp. 36-42.
- Berthon, P., Pitt, L. F., and Campbell, C. 2008. "Ad lib: When Customers Create the Ad," *California Management Review* (50:4), pp. 6-30.
- Berthon, P., Pitt, L. F., and Watson, R. T. 1996. "The World Wide Web as an Advertising Medium: Toward an Understanding of Conversion Efficiency," *Journal of Advertising Research* (36), pp. 43-54.
- Boyd, D. M., and Ellison, N. B. 2008. "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication* (13:1), pp. 210-230.
- Charlett, D., Garland, R., and Marr, N. 1995. "How Damaging is Negative Word of Mouth?" *Marketing Bulletin* (6), pp. 42-50.
- Chatterjee, P., Hoffman, D. L., and Novak, T. P. 2003. "Modeling the Clickstream: Implications for Web-based Advertising Efforts," *Marketing Science* (22:4), pp. 520-541.
- Chou, W. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., and Hesse, B. W. 2009. "Social Media Use in the United States: Implications for Health Communication," *Journal of Medical Internet Research* (11:4), pp. e48.

Cone. 2008. *Business in Social Media Study Media*.

Cook, N. 2008. *Enterprise 2.0: How Social Software Will Change the Future of Work*, Gower Publishing, Ltd., pp. 164.

Culnan, M. J., Mchugh, P. J., and Zubillaga, J. I. 2010. "How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value," *MIS Quarterly Executive* (9:4), pp. 243-260.

Desautels, P. 2011. "UGIS: Understanding the Nature of User Generated Information Systems," *Business Horizons* (54:3), pp. 185-192.

Duncan, T., and Moriarty, S. E. 1998. "Marketing for Model Managing Relationships," *Journal of Marketing* (62:2), pp. 1-13.

Eagly, A. H., and Chaiken, S. 1993. *The Psychology of Attitudes*, Fort Worth, TX: Harcourt Brace Jovanovich.

El Sawy, O. A., and Bowles, G. 1997. "Redesigning the Customer Support Process for the Electronic Economy: Insights From Storage Dimensions," *MIS Quarterly* (21:4), pp. 457-483.

Etgar, M. 2007. "A Descriptive Model of the Consumer Co-production Process," *Journal of the Academy of Marketing Science* (36:1), pp. 97-108.

Featherman, M. S., Valacich, J. S., and Wells, J. D. 2006. "Is that Authentic or Artificial? Understanding Consumer Perceptions of Risk in E-service Encounters," *Information Systems Journal* (16:2), pp. 107-134.

Ferguson, R. 2008. "Word of Mouth and Viral Marketing: Taking the Temperature of the Hottest Trends in Marketing," *Journal of Consumer Marketing* (25:3), pp. 179-182.

Fournier, S., and Avery, J. 2011. "The Uninvited Brand," *Business Horizons* (54:3), pp. 193-207.

Gallaugh, J., and Ransbotham, S. 2010. "Social Media and Customer Dialog Management at Starbucks," *MIS Quarterly Executive* (9:4), pp. 1389-1404.

Godes, D., and Mayzlin, D. 2004. "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science* (23:4), pp. 545-560.

Godes, D., and Mayzlin, D. 2009. "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science* (28:4), pp. 721-739.

Goldenberg, J., Libai, B., Moldovan, S., and Muller, E. 2007. "The NPV of Bad News," *International Journal of Research in Marketing*, pp. 186-200.

Herr, P. M., Kardes, F. R., and Kim, J. 1991. "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," *Journal of Consumer Research* (17:4), pp. 454-462.

- Hoffman, D. L., and Fodor, M. 2010. "Can You Measure the ROI of Your Social Media Marketing?," *MIT Sloan Management Review* (52:1).
- Hoffman, D. L., and Novak, T. P. 1996. "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing* (60:3), pp. 50-68.
- Iyer, G., Soberman, D., and Villas-Boas, J. M. 2005. "The Targeting of Advertising," *Marketing Science* (24:3), pp. 461-476.
- Kaplan, A. M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons* (53:1), pp. 59-68.
- Keller, K. L. 1993. "Conceptualizing, Measuring, and Managing Customer-based Brand Equity," *Journal of Marketing* (57:1), pp. 1-22.
- King, R. 2008. "How Companies Use Twitter to Bolster Their Brands," BusinessWeek, <http://www.businessweek.com/stories/2008-09-06/how-companies-use-twitter-to-bolster-their-brandsbusinessweek-business-news-stock-market-and-financial-advice>.
- Ko, H.-C., and Kuo, F.-Y. 2009. "Can Blogging Enhance Subjective Well-being Through Self-disclosure?" *CyberPsychology and Behavior* (12:1), pp. 75-9.
- Kozinets, R. V., Valck, K. de, Wojnicki, A. C., and Wilner, S. J. S. 2010. "Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities," *Journal of Marketing* (74:3), pp. 71-89.
- Laczniak, R. N., Decarlo, T. E., and Ramaswami, S. N. 2001. "Consumers' Responses to Negative Word-of-Mouth Communication: An Attribution Theory Perspective," *Journal of Consumer Psychology* (11:1), pp. 57-73.
- Lewis, S., Pea, R., and Rosen, J. 2010. "Beyond Participation to Co-creation of Meaning: Mobile Social Media in Generative Learning Communities," *Social Science Information* (49:3), pp. 351-369.
- Mangold, W. G., and Faulds, D. J. 2009. "Social Media: The New Hybrid Element of the Promotion Mix," *Business Horizons* (52:4), pp. 357-365.
- McAlexander, J. H., Schouten, J. W., and Koenig, H. F. 2002. "Building Brand Community," *Journal of Marketing* (66:1), pp. 38-54.
- Muñiz Jr., A. M., and O'Guinn, T. C. 2001. "Brand Community," *Journal of Consumer Research* (27:4), pp. 412-432.
- Nambisan, S., and Baron, R. A. 2009. "Different Roles, Different Strokes: Organizing Virtual Customer Environments to Promote Two Types of Customer Contributions," *Organization Science* (21:2), pp. 554-572.
- Parasuraman, A., Zeithaml, V. A., and Berry, L. L. 1985. "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing* (49:4), pp. 41-50.



- Piccoli, G., and Watson, R. T. 2008. "Profit from Customer Data by Identifying Strategic Opportunities and Adopting the 'Born Digital' Approach," *MIS Quarterly Executive* (7:3), pp. 113-122.
- Ray, G., Muhanna, W. A., and Barney, J. B. 2005. "Information Technology and the Performance of the Customer-Service Process: A Resource-Based Analysis," *MIS Quarterly* (29:4), pp. 625-652.
- Russell, M. G. 2010. "A Call for Creativity in New Metrics for Liquid Media," *Journal of Interactive Advertising* (9:2), pp. 44-61.
- Schau, H. J., Muñiz, A. M., and Arnould, E. J. 2009. "How Brand Community Practices Create Value," *Journal of Marketing* (73:5), pp. 30-51.
- Shankar, V., Smith, A. K., and Rangaswamy, A. 2003. "Customer Satisfaction and Loyalty in Online and Offline Environments," *International Journal of Research in Marketing* (20:2), pp. 153 - 175.
- Soriano, J., Lizcano, D., Cañas, M. A., Reyes, M., and Hierro, J. J. 2007. "Fostering Innovation in a Mashup-oriented Enterprise 2.0 Collaboration Environment," *System and Information Sciences Notes* (1:1), pp. 62-68.
- Sterne, J. 2010. *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*, Wiley, pp. 272.
- Stewart, D. W., and Pavlou, P. A. 2001. "From Consumer Response to Active Consumer: Measuring the Effectiveness of Interactive Media," *Journal of the Academy of Marketing Science* (30:4), pp. 376-396.
- Sullivan, M. 1990. "Measuring Image Spillovers in Umbrella-branded Products," *Journal of Business* (63:3), pp. 309-329.
- Tapscott, D., and Williams, A. D. 2008. *Wikinomics: How Mass Collaboration Changes Everything*, Portfolio, pp. 351.
- Thompson, S. A., and Sinha, R. K. 2008. "Brand Communities and New Product Adoption: The Influence and Limits of Oppositional Loyalty," *Journal of Marketing* (72:6), pp. 65-80.
- Tybout, A. M., Calder, B. J., and Sternthal, B. 1981. "Using Information Processing Theory to Design Marketing Strategies," *Journal of Marketing Research* (18:1), pp. 73-79.
- VanBoskirk, S. 2009. "US Interactive Marketing Forecast, 2009 To 2014," Cambridge, MA, pp. 1-20.
- Watson, R. T., and Straub, D. W. 2007. "Future IS Research in Net-enabled Organizations," *ACM SIGMIS Database* (38:3), pp. 8.
- Wattal, S., Schuff, D., Mandiviwalla, M., and Williams, C. B. 2010. "Web 2.0 and Politics: The 2008 U.S. Presidential Election and an E-politics Research Agenda," *MIS Quarterly* (34:4), pp. 669-688.

Webster, J. G. 2010. "User Information Regimes: How Social Media Shape Patterns of Consumption," *Northwestern University Law Review* (104:2) *Northwestern University Law Review*, pp. 593–787.

Whitty, M. T., and McLaughlin, D. 2007. "Online Recreation: The Relationship between Loneliness, Internet Self-efficacy and the Use of the Internet for Entertainment Purposes," *Computers in Human Behavior* (23:3), pp. 1435-1446.

Xu, S. X., and Zhang, X. (Michael). 2009. "How Do Social Media Shape the Information Environment in the Financial Market?" in *ICIS 2009 Proceedings*.

Zeithaml, V. A., Parasuraman, A., and Malhotra, A. 2001. "Service Quality Delivery Through Web Sites: A Critical Review of Extant Knowledge," *Journal of the Academy of Marketing Science* (30:4), pp. 362-375.

CHAPTER 3

THE ANALYSIS OF UNSTRUCTURED DATA: MEANINGFUL MEASUREMENT OF  
SOCIAL MEDIA INTERACTIONS<sup>2</sup>

---

<sup>2</sup> Larson, K. and Watson, R. T., to be submitted to *MIS Quarterly*.

## Abstract

Motivated by the organizational need to exploit valuable product- and brand-oriented consumer-generated text as it flows across social media settings, this paper explicates the difficulties posed by the automated (machine) analysis of highly unstructured data. The authors appraise the state of text mining and knowledge discovery capabilities through a review of existing approaches, particularly those falling under the rubric of natural language processing. Based on modern advances in statistical machine translation, this manuscript proposes a set of design principles to guide the development of an automated, corpora-based analytics system able to extract knowledge from social media text and thus, potentially, to improve the quality of organizational decision making.

**Keywords:** *text mining; natural language processing; social media; organizational decision-making*

### **3.1 The Problem of Proliferation**

While most companies deal with structured data on a daily basis in such forms as financial transactions, time stamps, and sensor and RFID data streams, the unstructured written word is one of the most important ways in which humans have communicated for centuries. A significant proportion of organizational data exists in unstructured (textual) format, perhaps as much as 85% (Lindvall, Rus, and Sinha 2003). Emails, corporate documents, news articles, web pages, and voicemail transcriptions typically occur outside of the bounds of pre-defined data models yet constitute dense and voluminous bodies of data that companies must store, process, and analyze to derive business intelligence and ultimately create value (Feldman and Sanger 2007). Due to the lack of repeatability, predictability, and definition at the atomic level, unstructured data pose a substantial challenge to organizations interested in exploiting the profusion of text available to potentially guide analytic efforts and, ultimately, decision-making. Although information in the unstructured environment is abundant and ostensibly useful, the sophistication of techniques for the analyses of texts is meager compared to what is available in the structured environment (Inmon and Nesavich 2008: xvii). For the most part, a few search engines constitute the majority of available mechanisms for uncovering information from streams of text.

The proliferation of unstructured data and the attendant problems of absorbing and processing large quantities of textual data are exacerbated by the emergence and increasing popularity of social computing services. Within the purview of “social media,” a wide range of platforms and technologies galvanize collaborative phenomena; in addition to traditional social networking via Facebook, LinkedIn, etc., Foursquare and Twitter-type tools enable environment-interaction and awareness-creation while sites

such as Wikipedia and IMDB facilitate crowd-level knowledge sharing. Further potential sources of intelligence, platforms such as Dropbox and Mendeley serve as robust repositories for text-based collaboration. The communal production and exchange of various digital commodities define a new type of collective service that bears promise for a variety of stakeholders ranging from organizations, investors, and customers to governments, citizens, and global society as a whole. However, in addition to the wealth of intelligence they render, the enormity of data emerging from these shared processes also poses a serious measurement conundrum for information managers: how might a company wade through the more than 140 million Tweets<sup>3</sup> and 2 billion “likes” and comments<sup>4</sup> created daily through Twitter and Facebook alone to derive a meaningful understanding of a variety of firm/customer processes such that organizational decision-making is enhanced? Thus motivated by the pragmatic organizational requirement of a useful and effective mechanism for resolving this information-processing dilemma, this conceptual paper addresses the following research question:

*RQ: How can we best measure social media effects in an organizational context?*

As organizational researchers interested in discovering and deriving new information from consumer-generated data, we look to the field of text data mining to understand the state of the art of finding patterns across datasets and separating signal from noise. Whereas this relatively young research area has occupied a great deal of attention on the parts of computer scientists and computational linguists over the past

---

<sup>3</sup> <http://blog.twitter.com/2011/03/numbers.html>

<sup>4</sup> <https://www.facebook.com/press/info.php?statistics>

couple of decades, it has yet to generally infiltrate the collective vision of IS and organizational scholars. Systems designed to address the analytic needs of organizations interested in evaluating and summarizing text communications have generally enabled visualization of metadata contained in message headers (i.e., send/reply and posting patterns), but have provided little support for the analysis of actual message body text (Abbasi and Chen 2008). Despite the potential it holds for uncovering previously unknown information from the depths of large collections of text (Hearst 1999), the IS management field's reluctance to capitalize on the advances made in this interdisciplinary territory likely stems, at least in part, from the convolution that muddles our understanding of what technologies actually constitute text data mining, and to what ends. Thus, the aims of this manuscript are manifold. Our primary goal is to address a practical organizational need. We develop a set of theoretical design guidelines for a class of analytics systems capable of promoting organizational decision-making through the analysis of social-media-generated texts. Secondly, this paper presents the IS community with a concise yet insightful explication of the difficulties posed by the automated analysis of text and appraises the state of text data mining capabilities by reviewing existing text data analysis tools and methodologies relevant to organizational research. Our hope is to bring relevant knowledge accumulated in other fields into IS conversations about the analysis of very large datasets, such as those extracted from application-mediated social interactions.

### **3.2 Measuring Social Media Data**

Analysis of the underlying interactions driving social media activity (Larson and Watson 2011) indicates three levels of measurement inherently applicable to social

media-generated data, the simplest being counts of objects and actions such as users, comments, and links followed. While such tallies are useful for tracking certain high-level trends (e.g., purchase conversion, a decidedly critical metric), these numbers provide a limited depth of understanding in terms of customer reactions and opinions, an important source of additional intelligence for organizations. A possibly more revealing – but simultaneously problematic – mode of measurement, sentiment analysis, or the assessment of positive and negative customer sentiment in product or brand reviews or mentions (Pang and Lee 2008), has the potential to impart more insight into customers’ reactions to a given organization and its products and services. However, we are somewhat dubious of the practical accuracy of this methodology on an automated and large scale due to the role of assessment at the word level that ignores the meaning of whole comments and is especially susceptible to misclassifying sentiment conveyed by irony or sarcasm. We also refer to recent demonstrations of the range of granularity lost through sorting comments and reviews according to simple negative/positive rating scales (e.g., Pavlou and Dimoka 2006).

Our assessment of the capabilities conveyed by count and sentiment analyses leads us to conclude that deeper insight into the impact of a given product or service requires a correspondingly deeper level of qualitative analysis. Toward the goal of deriving valuable business intelligence from consumer-generated unstructured data, organizations must become capable of qualitatively analyzing textual data on a large scale in near-real time, similar to the operational capabilities of just-in-time BI that reduce the latency between data acquisition and analysis (Chaudhuri, Dayal, and Narasayya 2011). However, a survey of current technologies indicates that while ratios of positive to negative mentions and counts of users and actions are accessible to firms, the



ability to glean deep meaning from voluminous streams of social media-generated data is an expertise yet to be fully established and incorporated into organizational business intelligence-oriented monitoring.

To illustrate the organizational need for deep textual/situational scrutiny above and beyond counts and current sentiment analysis (SA) capabilities, the brand-centric customer-to-customer exchange in Figure 3-1 portrays a common situation requiring a more complex mode of measurement to extract the tenor of the intra-group customer conversation as it pertains to the participants' intellectual and affective engagement with the product. Particularly lacking, yet critical for comprehension of the conversation, is a means for preserving the sequencing and interconnectedness of each participant's verbal (in the form of comments) or active (in the form of "liking," or clicking the "Like" button) contributions as the conversation unfolds. To our knowledge, there is no automated (high-volume) tool available to organizations that can capture temporal ordering and convey through analysis how comments and actions incrementally build on one another.



**Figure 3-1**  
**Brand-centric conversation snippet**  
**extracted from a popular social media application**

When read with an understanding of modern speech patterns and popular cultural references, the conversation above represents an overall highly positive reaction toward the Keurig brand single cup coffee brewer, or at least it does to the participants of the conversation and should be interpreted as such by an organization attempting to understand consumers' attitudes toward the Keurig brand. Time-based analysis of sub-events occurring as the conversation unfolds provides evidence of this claim: Elaine asks a general question about a type of coffee maker, Meredith offers a counter suggestion that a non-commenting participant subsequently "likes," and then Michele returns to the original type of product and answers the original question more directly. Elaine shows appreciation for the direct answer by "liking" it. Keri textually concurs with Michele's opinion, for which Michele subsequently shows appreciation by "liking."

Simple count and sentiment analysis is unlikely to accurately ascertain the gist of this exchange. A count of the conversation may capture somewhat limited information—one mention of the Keurig brand, or possibly that four participants contribute to a

conversation about the Keurig brand—but sentiment analysis is unlikely to detect the essence of the participants’ reactions because of the lack of synonyms for “good,” “bad,” “like,” “dislike,” within the text. In fact, it is possible that sentiment analysis would only associate a positive reaction with “single-serve French press,” potentially reckoning negative sentiment toward the Keurig brand since the comment expresses preferential words in connection with a competing product. Faced with the task of trying to glean useful intelligence from thousands of such “cryptic” conversations daily, the ability of firms to derive valuable knowledge relies on the capability to discern patterns of speech based on the semantics of *interdependent* clauses and responses. For example, the final comment (“What Michelle...said”) in Figure 3-1 fully depends on its relationship to the preceding comment (“Keurig changed my life...”) for comprehension; an analytic tool capable of extracting reliable intelligence from such an interchange would require the capability of “understanding” this interconnectivity.

Further, to account for scale, firms require a tool that is able to perform such temporal semantic analyses on *high volume* data streams. While a single person or even a team of analysts might be able to make some segmented assumptions about brand trends based on manually monitoring a limited set of social media exchanges that transpire within a bounded time period, humans’ ability to absorb and process information is incommensurate with the hordes of data generated hourly by the entirety of consumers covering across the social media landscape. And in the same way a company would not compute key performance indicators based on 5 or 10 percent of its sales data, it is likewise unreasonable to expect a company to make decisions based on the analysis of a limited portion of customer interchanges in the form of random

selection, especially in an age when storage and processing capabilities preclude the need for sampling.

Expanding the fundamental logic of this example, we propose that the unstructured textual data generated in mass quantities within the context of social media necessitate analysis techniques able to move beyond simple extraction and make inferences from the contents of communications. Organizational interest is not limited to whether consumers like or dislike products or services; a wide range of additional value-adding information is potentially derivable (and thus co-optable) from social-media-facilitated conversations, including, for example, novel uses of products devised by consumers (or product “hacks,” popular with and heavily discussed among particular interest groups such as Ikea enthusiasts), multimedia productions created by company fans (see YouTube for any number of viral videos created by product devotees from which companies enjoy free publicity), and suggestions for products or improvements that might arise through collaborative discussion amongst social media participants but are not communicated through direct channels to the company. Proliferation of such data via the realm of social media applications is overwhelming, but modern technologies enable us to store and process these data and create the potential to detect patterns; as we increase the sophistication of these capabilities we leverage a valuable source of information not just for firms and brand managers but for a wide array of knowledge professionals ranging from physicians (Denecke and Nejd1 2009) and pharmacists (Agarwal and Searls 2008) to scientists (Shatkay and Feldman 2003) and manufacturers (Choudhary, Harding, and Tiwari 2009) who rely on textual data analysis to effect a variety of tangible and intellectual outcomes.

### **3.3 Information Processing x Decision Making**

When the analysis of unstructured data becomes part of the decision-making process, organizations have the opportunity to make timelier, more accurate, and better-informed decisions. In this sense, text analytics has the capacity to determine important key performance indicators (KPIs) for driving business decisions and added functionalities such as personalization of offers and services for customers, much the way BI tools have enabled these goals for businesses based on numerical data over the past few decades (Chaudhuri, Dayal, and Narasayya 2011).

Given that the ultimate purpose of textual data analysis is to process non-numerical information in order to support decision-making, we propose in Table 3-1 a classification scheme of textual analyses as a function of how the information is sorted (Mode of Information Processing) in conjunction with the type of decision making the information supports (Mode of Decision Making). Specifically, a text data stream can be read manually by a human agent or processed automatically by a machine algorithm, while the data analysis can support either a hypothesis verification-type of decision (wherein the text is classified according to a pre-set classification scheme) or can inform efforts to uncover previously unknown knowledge in information-discovery mode.

**Table 3-1**  
**Cross-matrix of information-processing and decision-making modes**  
**supported by textual data analysis**

|                                       |   | <i>Mode of Decision-making</i>   |   |
|---------------------------------------|---|--|---|
|                                       |   | <b>Verification</b>  | <b>Discovery</b>  |
| <b>Mode of Information Processing</b> | <b>Manual</b><br>(effective, inefficient)             | Manual verification<br>(e.g., contextual analysis by human raters to sort words or messages into categories theoretically determined a priori) | Manual discovery<br>(e.g., qualitative text coding by human raters to develop grounded grasp of themes underlying cohesion of texts)      |
|                                       | <b>Computer-automated</b><br>(efficient, ineffective) | Automated verification<br>(e.g., machine sorting of each word in a review as positive or negative to determine overall sentiment of a review)  | Automated discovery<br>(e.g., machine processing of dynamic data stream to uncover critical underlying trends and discover new knowledge) |

### 3.3.1 Information processing modes

Manual text interpretation methods such as contextual analysis result in effective understanding of unstructured data due to the entirety with which text data is naturally processed by human readers. In manual mode, humans interpret the message intended by the text's author by reading sentences and paragraphs as well as noting contextual features of the message or document that may convey meaning (Anderson and Pérez-Carballo 2001). It is typically through holistic reading that humans are able to accurately categorize the information underlying a particular message, comment, or other unit of text data according to a conceptual scheme, drawing on some degree of similar background knowledge of typical human experiences. These taken-for-granted

perceptions—including the tacit understanding of how objects relate to each other in the world, the goals people tend to seek in their daily lives, and the emotional impact of certain kinds of events or situations—refer to the concept of “common sense,” or the “obvious things people normally know and usually leave unstated” (Grassi et al. 2011). Primed by this common-sense state, the human analyst thus consumes words and features of the text string in sum and processes them from a general perspective of shared language and common experience of the world and its everyday situations, in turn increasing the chances that sophisticated forms of speech such as verbal irony or sarcasm ( Davidov, Tsur, and Rappoport 2010; González-ibáñez and Wacholder 2011), colloquialisms such as slang or pop cultural references, or even misspellings (Furnas et al. 1987) do not delimit understanding or interpretation.

### **3.3.2 Decision-making modes**

Many of the challenges faced by linguists and scientists in their quest to create computer algorithms that can automatically understand written language stem from the absence of this state of human common sense. However, the efficiency gains realized by the automatic analysis of text are essential to any realistic analytic approach that aims to synthesize and extract knowledge from the text generated across social media platforms; the volume of text publically issued on a daily basis far outstrip the reading and processing capacity of even a large team of human analysts. Given that the number of daily Tweets alone ranges above the 100 million mark, at best, a manual approach to comprehensive social media text mining would require random sampling of a very small ratio of overall data, meaning that the overall inferences and trends assumed by a human analyst team would be based on a random and small proportion of customer

feedback. The probability of a high rate of accuracy in this case would be unlikely, meaning that critical action items (e.g., a dangerous product failure, extreme dissatisfaction by a highly influential customer) could easily, and detrimentally, go undetected and result in a waste of expenditure due to the inability to leverage the effectiveness advantage of human common sense previously assumed.

Regardless of the approach, whether manual or automated, we identify two modes of decision making effected through the positivist analysis of textual data: verification and discovery. Choice of technique, and thus mode of decision-making employed, depends on the research question of interest (UGA Working Paper 2011). Specifically, a researcher would employ a verification methodology to find patterns of information in a corpus of text to support a hypothesis or theoretically derived research question. From this perspective, an example of this deductive approach is the use of content analysis, a popular research technique intended to objectively and systematically quantify the meaning of textual communications (Berelson 1952). Typically, a human agent accomplishes this technique by reading each sample of text (e.g., feedback review comment, a CEO's letter to the shareholders, etc.,) and making a judgment as to which preconceived category the text's intended meaning most accurately supports. A recent analysis of eBay seller feedback comments, for example, hypothesized that prior feedback indicating outstanding benevolence and credibility increase buyers' beliefs in sellers' benevolence and credibility which in turn positively influences price premiums, while comments indicating abysmal benevolence and credibility negatively influence buyers' beliefs in sellers' benevolence and credibility (Pavlou and Dimoka 2006). In order to derive a numeric data set to build and test a statistical model from 11,000 textual feedback comments, each comment was manually



sorted into the categories, “outstanding benevolence,” “abysmal benevolence,” “outstanding credibility,” “abysmal credibility,” and a catch-all for “ordinary comments.”

Thus, we consider the verification mode of textual analysis to be a hypothesis-testing mechanism of sorts. Similar to information retrieval methods, hypothesis-testing researchers typically seek out information that fits with a predetermined schema and then judge the success of their retrieval attempts. For example, in the case of sorting eBay feedback content into categories of a predetermined classification system, each time the comment fits, or is fitted into, a classification, the researchers’ hypothesis that their set of categories explains variation in the phenomenon is essentially supported.

Alternately, the discovery mode of text data mining, which encompasses a variety of goals such as discovering new knowledge, uncovering critical trends in consumer behaviour, and grasping themes that are driving the collaborative interactions of customers within social media platforms, implies far more complexity than most verification-oriented techniques. While pre-processing prior to sorting or classification techniques often disassembles text into a “bag of words” without preservation of relationships or context, discovering new knowledge and trends and uncovering unexpected concepts almost certainly indicates the need for understanding how a text’s words are related. Not only must relationships between subjects and objects be retained for interpretation, so should the context of multiple users’ comments when embodying *interaction*, thus requiring that sequential relationships among texts also be captured. We offer a more detailed discussion of the implications of text mining and classification in the following section.

### 3.4 Exploiting Textual Data

While text mining encompasses a vast array of theoretical approaches and methods (Feinerer, Hornik, and Meyer 2008), efforts in the quest to extract useful and important information from unstructured data align with two possible major architectural schemas. Porting textual data into the structured environment to leverage existing analytical (prediction, forecasting) tools, software, and infrastructures is the approach underlying many social media and reputation monitoring applications available to organizations at this time. Within the framework of this methodology for dealing with textual data streams, examining words (as opposed to sentences, phrases, messages, etc.) as the relevant unit of data simplifies the analytical process by freeing the analyst from concerns regarding context (Inmon and Nesavich 2008: xix). This approach is particularly suited to those types of research questions that can be answered through word frequency counts, the detection of certain pre-specified words or phrases, or measurement of distance between words. For example, Google's Ngram Viewer<sup>5</sup> allows users to query a structured database of words from more than 5 million digitized books in a variety of languages published over five centuries, enabling answers to a variety of such quantitative research topics as counting the number of words that exist in the English language, calculating the duration of time it takes a verb to regularize, and tracking the impact of censorship on a person's cultural influence (Michel et al. 2011). Specialists of this unstructured-to-structured approach, however, acknowledge that the expediency and relative ease with which decontextualized words can be stored, retrieved, and evaluated come at the loss of context-based precision, a major

---

<sup>5</sup> <http://books.google.com/ngrams/>

consequence of which is the possibility that substantial effort and budget are expended in return for very little gained knowledge.

### **3.4.1 Inaccuracies of decontextualized, word-level analyses**

While word-level analytical approaches are especially convenient in that they obviate attempts to understand sarcasm, slang, innuendo, pop cultural references, or colloquial spellings, either by a piece of software or a sentient human agent, we argue that the amount of *actionable* intelligence organizations can expect to glean from such evaluations is limited. At a minimum, words of a text segment or document are sorted into two categories, the most popular schema being the “positive” versus “negative” evaluation standardly known as sentiment analysis but also called opinion mining, subjectivity analysis, appraisal extraction, and affective computing across multiple related literatures (Pang and Lee 2008). Beyond a heuristic for a quick assessment of whether a customer is overall more pleased than displeased, logic dictates that a simple calculus of negative versus positive words within a review sheds limited light for organizational decision makers (despite the utility it may convey to a potential consumer). A more useful analysis would capture links between subtopics in a review and corresponding opinions, but such associations are extremely difficult to extract accurately at the word-based level. It is understood that extraction algorithms perform best when the topic is known *a priori* (Yi et al. 2003), offering little benefit to organizations attempting to unearth new knowledge from open-ended text data. For example, a consumer may create an overall favourable review of his new refrigerator but be particularly dissatisfied with the noise of the icemaker. In this case, the individual weaknesses of the product may be of more import to the manufacturer than the overall

review, but statistical word-based approaches cannot reliably extract and preserve associations between multiple topics and corresponding sentiments in the same message (Yi et al. 2003).

Considering the extreme simplicity of a bifurcated classification scheme, one might reasonably question the extended usefulness of traditional sentiment analysis even in the case of a human agent able to understand the most veiled of verbal intricacies processing the text at the sentence level. A recent study of text comments in the domain of online reviews (Pavlou and Dimoka 2006) demonstrates that important nuances in the information conveyed by consumers' text comments cannot be accurately represented by a crude positive/negative rating system. As such, tracking polarity trends in consumer reactions to a given product or service in an attempt to ascertain whether overall reception is more "good" than "bad" fails to leverage a wide range of discernible information falling along the continuum of positive to negative poles. As example, a consumer complaint about a slight delivery delay constitutes a far less critical problem for an organization to address immediately than a complaint detailing severe underpackaging resulting in systematic product damage, although both complaints would likely be categorized as negative in a sentiment analysis-type of approach.

Congruent with the logic of Pavlou and Dimoka, we assert that fine-grained differentials of meaning are lost in the course of typical "on/off" analyses that attempt to catalog individual words into one of two buckets. In addition to the loss of information gradation, attempts to classify words as bad or good are often simply inaccurate. Human vocabulary is fraught with ambiguity; entire manuscripts have been dedicated to explicating the positive or negative nature of a particular word (e.g., "whatever," Benus et al. 2007). Thus, even complex approaches for identifying sentiment at the sentence

and subsentence levels based on standard lexicons risk critical imprecision during the categorization process. To illustrate, classifiers developed within a recent cascading binary classification approach fell apart when applied to the sentiment analysis of political blogs due to the high frequency of sarcasm and domain-specific subjective terminology (Fink et al. 2011). The authors conclude that the poor performance of their system in the case of political text analysis is due to the generality of the reference lexicon, pointing to the need for specialized feature sets including domain-specific subjective terms.

Intensifying the inaccuracy of a decontextualized, word-based approach is the evaluation of word polarity *without regard to the words around it*, a mechanism common to many tools that classify tone and sentiment based on lookup tables of words. A recent exploration of word-based sentiment analyses of financial documents, for example, discovered a 73.8 percent misalignment rate between words considered negative according to the *Harvard Psychosociological Dictionary* and those not typically considered negative in a financial context (Loughran and McDonald 2011). In more prosaic terms, classification of a review describing a product as “hard” is likely to connote a negative assessment if pertaining to a loaf of Italian bread, yet signifies a desirable trait in diamonds, and probably does not indicate an emotional judgment either way in the phrase “hard-boiled detective novel.” Described as a word’s *contextual polarity*, the phrase in which a word appears may evoke a different polarity than the word’s *prior polarity*, or decontextualized sentiment (Wilson, Wiebe, and Hoffmann 2005). Thus, while “enthusiasm” is *a priori* a positively-oriented noun, it can conversely indicate negative sentiments as evidenced by the example phrases, “he’s lost his enthusiasm for life” and “she has an enthusiasm for breaking the law.”

Additional lexicographic complications to consider in the effort to analyze text at the word level include the fact that the number of English words consistently conveying a strictly negative or strictly positive meaning is limited; a human agent would be hard-pressed to generate a list of such words void of variable meanings across generations, contexts, or communities of speakers that use words in particular, idiosyncratic ways. Traditionally affirmative words may be used sardonically to convey actual dissatisfaction (e.g., “this was a great waste of time,” or in the case of a book review, “the movie was great”), while those terms historically denoting the vilest of affronts have sometimes been reappropriated to disempower the language and, in turn, convey approbation. Because words fluctuate in meaning, the power to understand them resides in considering them relative to context, a functionality that is precluded when unstructured textual data is ported on a word-by-word basis into the structured environment for analysis.

### **3.4.2 The promise of natural language processing (NLP)**

An alternate approach to dealing with unstructured text is to attempt to analyze it within its native form, processing it as natural language with context instead of trying to reduce it to easily taggable, searchable, database-storable keywords prior to classification.

Whereas the former tactic of decontextualization lends itself suitably to the discovery of patterns of words and concepts that can in turn inspire explorations into a wide variety of interesting problems, we contend a key characteristic shared by these efforts is that such inquiries are contextualized from the outset by the datasets upon which the analyses are conducted. To illustrate, research questions answerable through

an understanding of certain historical trends as reflected by keyword appearances among letters to shareholders (UGA Working Paper 2011) across time are necessarily shaped—and limited—by the corpus on which they are predicated; letters crafted by CEOs to communicate with their shareholders present a finite context of unilateral communication characterized by concepts such as company growth, financial performance, operational concerns, and the like (Kohut and Segars 1992).

On the other hand, the corpus of participant-generated social media data is not delimited by sender, recipient, message context, level of formality, or professional authorship the way a database of CEO letters to shareholders is, but instead represents an unknown range of interchanges created through multiple sources by users of all ilks, bents, and agendas regarding any number of people, places, things, or situations (i.e., all Tweets, Facebook statuses/likes/comments, YouTube videos, etc. generated by millions of people daily). From these myriad combinations of subjects, objects, and mediating platforms, organizations hope to perceive and accurately assess useful brand- or product-level customer opinions, reactions, suggestions, and collaborative projects, but first must detect such instances from amongst the rest of the “noise,” including the now-customary use of a wide array of spelling, punctuation, and special character shortcuts, innuendos, and reference pointers that further complicate the intent to analyze customers’ social media conversations. As such, the data are not neatly packaged into preconceived subject parcels but must be heavily pre-processed in order for an organization to viably mine meaning from them. However, while considerably more complex than decontextualized word-based approaches, we argue that the synthesis of a far greater range of unconstrained organizational intelligence is at least theoretically possible from a contextual approach.

Because natural language processing (NLP) refers to a wide range of language technologies, tasks, subtasks, and related fields and is often used interchangeably with the phrase “computational linguistics” in academia, we provide definitional boundaries here to guide our inquiry. In its widest interpretation, NLP can mean any type of computer manipulation of natural language (i.e., English, Chinese, French) used by humans to communicate. This can include simple counts of word frequencies, or extend to the automated “understanding” of human verbalizations (Bird, Klein, and Loper 2009). We look to the latter, more sophisticated extreme in our quest to extract meaning from unstructured data, approaching NLP tools as means for realizing fuller meaning from free text data streams through the preservation and exploitation of linguistic rules like parts of speech (i.e., nouns, verbs, adjectives, etc.) and grammatical structures (the application of sentence formation rules in a given language), and advances in resolving anaphora (e.g., aligning backward-referring pronouns and phrases with the appropriate nouns) and ambiguities of language and grammatical structures (Kao and Poteet 2007: 1). Following our distinction between verification and discovery modes of decision making, the natural language approach of extracting relationships among entities (Bunescu and Mooney 2007), as opposed to other common text-classification types of mining approaches that treat documents or text segments as unstructured buckets of words with frequency counts but no relationship with respect to one another (Kao and Poteet 2007: 2), aligns with the non-trivial goals of discovering events, entities, and relationships (e.g., who likes what product, who agrees with whom and why, or how customers use products).



### **3.4.3 NLP challenges**

Despite the information discovery opportunities potentially heralded by NLP text mining approaches, the achievement of enabling a machine to extract or infer a meaningful level of intelligence from consumers' social media text interactions is by no means trivial, or even assured. Automated analysis of text messages, blog posts and comments, Tweets, and status updates and comments pose a challenge beyond the general difficulties encountered in the quest to extract semantic relationships among entities mentioned in text documents or segments due to the lack of formalized writing style generally inherent in these particular types of written language.

Researchers involved in significant current efforts to develop accurate automatic extraction of information from biomedical texts point to the difficulty of their semantic extraction tasks introduced by the misalignment between most existing natural language tools (e.g., tokenizers, parts-of-speech taggers, parsers) and the biomedical body of literature; as these tools have traditionally been trained against news corpora, they incur a loss of accuracy when ported into a biomedical setting (Bunescu and Mooney 2007). Whereas newspaper discourse usually includes mentions of entity types such as people, organizations, and places and relation types including social relationships, positions people hold in organizations, relationships among organizations, etc., scientific publications follow a substantially different narrative type with relevant entities including proteins, genes, and cells and relations following patterns such as subcellular location and protein-protein interaction (Bunescu and Mooney 2007).

In addition to these obstacles to extracting relationships, biomedical extraction is further complicated by the seemingly straightforward task of named entity recognition

(NER), a separate NLP task whose goal is to identify within text all the names for specific types of things, typically persons, organizations, and locations (Sang and De Meulder 2003) but in this case, gene, drug, and protein names (Cohen and Hersh 2005). Fundamental to more complex text mining tasks such as relationship extraction (because relationships are anchored by participating entities), the process of recognizing biological entities in order to represent them in some consistent, normalized form has met with several obstacles, notably the lack of a complete lexicon comprising all possible biological named entities which thus precludes the use of simple text-matching algorithms (Cohen and Hersh 2005).

By extrapolation, considering the complexities involved in porting algorithms from one fairly-well-defined, formalized narrative to another similarly formalized one, the prospect of applying existing tools to an unstructured, ad hoc, informal text collection unrestricted to any domain whatsoever and replete with misspellings, slang, and emoticons appears, at the very least, daunting. When resolving ambiguities such as multiple names that refer to the same gene adds substantial complexity to the task of automatic analysis of biomedical text, the far higher degree of open-endedness inherent in millions of possible conversations regarding any topic is without doubt a major concern to the efforts to build a useful social media text analytics system.

#### **3.4.4 Machine translation**

One possible approach to resolving these challenges is to look to the advances made in the overlapping field of statistical machine translation (SMT). The dominant framework for modern machine translation research (Hutchins 2006), this data-driven or corpora-based, machine learning method describes the automated translation of text

from source to target language through algorithms that automatically “learn” to translate by examining millions of samples of human-produced translation (Lopez 2008).

Statistical (as opposed to rules-based or example-based paradigms) translations maximize the probability that a string in the target language is the translation of a string in the source language, although these probabilities and searches may be modelled according to numerous approaches (Brown et al. 1993). The parameters of these distribution models are derived from training data in the form of comparative analysis of bilingual corpora (Brown et al. 1993).

On a conceptual level, translation from target to source language follows the general idea of converting the source sentence into a knowledge representation via the use of a dictionary that maps words (e.g., *river*) onto concepts (e.g., river) with corresponding fact-based limitations based on world knowledge (such as, rivers cannot ride horses) (Knight 1997). This step gives additional context to a phrase like “Rick saw the Colorado River riding a horse” to resolve some possible ambiguities and improve translation accuracy. Finally, the conceptual structure is expressed in the target language. Specifically, at the sentence-level, words or word sequences of the source language are aligned with corresponding sequences in the target language. Based on these alignments, translation occurs through the selection of the most probable target output for each input phrase as well as a determination of the most probable output sequence (sentence structure), based on millions of known aligned phrases (Hutchins 2006).

An important feature shared by machine translation and an ideal social media text mining process is the unrestricted nature of the source text; just as we must

consider technical means for coping with non-domain-specific written language in our quest to extract information, machine translation also translates text featuring non-specific entities and relationships. While statistical machine translation is a very active realm of ongoing research and is by no means a perfected science, the processes through which accuracy has increased can motivate the parallel pursuit of accurate, or meaningful, social media text mining expertise.

### **3.5 Social Media Analytics System**

Machine translation has evolved from a theoretical to a practical undertaking, with early translation systems driven by computer science, linguistics, and artificial intelligence theories yet consistently producing bad translations (Knight 1997). “Translation 1.0” executed word-by-word translations, and though was theoretically informed, often produced meaningless or nonsensical results because it looked at each word in isolation without sensitivity to context. Humans use theories to solve problems, but sometimes our theories are insufficient and we move to a practical mode of problem solving—in the case of machine translation, the field has shifted to a statistics-based approach because linguistics theories were not hearty enough to sustain practical results.

Similarly, the optimal approach to the design of a social media analytics information system may not be rooted in a particular theory, per se, but perhaps is best served by looking to the realm of design science for its theoretical underpinnings. Information Systems Design Theories (ISDTs) are prescriptive theories offering theory-based principles as guidance for the development of effective information systems (Walls, Widmeyer, and El Sawy 1992). Although traditionally guided by kernel theories

from the natural or social sciences as governance mechanisms for the design process, we look instead to the practical elements we have identified as necessary for a meaningful, useful system given the constraints and challenges posed by consumer-generated texts. We incorporate an historical and empirical understanding of recent advances in machine learning, machine translation, textual data mining, and natural language processing to ground our articulation of a set of design principles for a social media text mining system sufficiently robust to reliably inform organizational decision making.

The main purposes of an ISDT are to grow human and organizational capabilities through the creation of new and innovative systems while simultaneously reducing the uncertainty inherent in developing such novel, and therefore untested, classes of information systems (Brohman et al. 2009). In addition to the practical contribution of addressing a serious management problem, in our case the problem of how to reliably and automatically analyze social media data result in a timely enough manner to support organizational decision making, the design and building phases of the proposed system will also result in knowledge of the new problem domain (Brohman et al. 2009) and, iteratively, a better understanding of appropriate solution(s). In our case, for example, we propose IS design guidelines for a new class of social media analytics in an effort to develop an understanding of text mining and natural language processing in a highly unstructured context.

Development of an IS design theory means addressing a novel problem not solvable using existing design theories (Markus, Majchrzak, and Gasser 2002) and grounded in a ‘kernel’ theory (Walls, Widmeyer, and El Sawy 1992; Hevner et al. 2004). Our approach to this particular facet of the design science methodology is to incorporate a type of “kernel theory” akin to practitioner theory-in-use (Brohman et al. 2009),

contextualizing the area of applicability of our system through the natural language processing capabilities we have identified as essential to the meaningful machine analysis of organizational social media data and the organizational requirements we identify as driving this need at the practical level.

In a similarly practical IS domain of business process redesign (BPR), although prior researchers characterized extant BPR literature as “atheoretical,” subsequent design scientists differentiated between scientific theory and the representation of a form of knowledge pertaining to everyday people (managers, executives, practitioners) as theory-in-use (Sarker and Lee 2002). Theories-in-use contrast with the type of theory that informs social-scientific inquiry. Although a theory-in-use incorporates commonplace concepts derived from practice instead of formal theoretical concepts that characterize scientific theory (Sarker and Lee 2002) and do not make claims about objective truth but about effective action (Argyris and Schön 1978), the design science literature argues similarity between the two in that they are “both theories that can and should be tested and, when refuted by the facts, be discarded, making room for a better theory” (Sarker and Lee 2002: 5). Claims about effective operation must also undergo rigorous testing to ensure that we, as researchers, provide reliable results to practitioners and, by extension, those who rely on these practitioners for services (Argyris and Schön 1978). We are able to test these applied theories-in-action by using the systems they engender.

In the spirit of the accumulating body of design science literature, the practical theories-in-use (or theories-in-practice) on which we draw will thus contextualize the evaluation of the final product. We further adhere to the design science methodology according to the following prescription of:

- gathering data directly (in our case, in the form of both secondary data and through the review of multiple relevant literatures),
- developing a set of principles for our proposed system, and subsequently,
- producing and evaluating a viable prototype instantiation (Hevner et al. 2004).

In the vein of recent work (e.g., Gregor and Jones 2007) positing the appropriateness of separating design products (including features and/or instantiation of the system) from the design process (such as developer guidelines) (Walls, Widmeyer, and El Sawy 1992), we focus on the product, entailing the contribution of a proposed architecture and design and instantiation of its components (Gregor and Jones 2007). This choice is particularly apt given the complexity and novelty of the domain we address (Brohman et al. 2009).

### **3.5.1 Why is a design theory for social media analytics needed?**

Design science research often proposes new design theory in order to rectify limitations to existing theory as identified by the researchers and manifested by unsuccessful prior enactments. For example, a recent article addressing pervasive implementation failure of customer relationship management (CRM) systems articulates an IS design theory for service systems intended to rectify a variety of shortcomings of traditional CRM approaches, including overly IT-centric and firm-centric approaches that fail to recognize the role of customer (Brohman et al. 2009). In the domain of natural language processing of unstructured text, we see from the machine translation literature that prior gap between theorizing and a successful implementation. As such, missing from the literature is any type of IS design theory for decision-making-oriented textual analytics. Specifically, we identify a need for new design theory that can help firms leverage a system predicated on the achievements

made in the realm of statistical machine translation and the practical requirements we previously identified for the extraction of relationships from unstructured text. Based on the tenets of such a design theory, firms would find themselves in position to exploit social-media-generated text in a way that is appropriate and sufficient for driving improved decision making in an era of widespread social media engagement by customers.

In our case, we propose incorporating knowledge gleaned from the history of machine translation to guide our theorization and framing our work according to the design science methodology in order to leverage the rigor and legitimacy it has attained in IS research (Gregor and Jones 2007) in recent years. By choosing a formal framework for grounding our efforts to determine critical elements of a text mining system capable of producing *actionable* results, we integrate our research into a tradition of cumulative knowledge building, as opposed to risking an inadvertent contribution to the re-invention of design systems and methods under differing rubrics, identified by scholars as a potential hazard in IS research (Gregor and Jones 2007).

### **3.5.2 Design principles**

In prior sections, we describe critical characteristics of the type of highly unstructured text flowing across social media settings and recent pertinent advances in natural language processing research from which we draw the design principles that guide our social media text analytics system design and prototype. As we discuss above, a substantial portion of consumer-generated social media text streams contains potentially-useful product and brand intelligence. However, as enticing as these data are from an organizational decision-making standpoint, firms currently lack the ability to



dissect more than a meager and doubtfully representative sample of them beyond word counts and simple, decontextualized sentiment analyses. The result of this gap between available data and analytics capabilities is substantial missed opportunity to tap into collaborative reactions, opinions, and activities of a firm's most important stakeholder, the customer. In the remainder of this section, we recap the properties driving the design principles for our system.

***Contextual sensitivity.*** Prior work points to a higher level of qualitative measurement beyond word counts, sentiment analysis, and even contextual analysis when conducted at the word level, to accurately capture and measure the interactions comprising product- and brand-oriented social media content. Based on evidence, we argue for the necessity of preserving context in a meaningful analysis of unstructured, messy, highly irregular text. Meaning is derived from context, and a sentence, the smallest unit of contextual analysis, is contained within a sequence of sentences in a single document or, significantly, a conversation among multiple participants. In general, the more context afforded, the greater the level of comprehension, which leads to our first design principle:

**Context-sensitivity principle:** A design must create output that is contextually situated by supporting sentence-level analysis, sentence-sequence analysis, and multi-party conversation analysis.

***The irregularity of informal text.*** Formal and consistently-written structures of entities and relationships, such as interactions between proteins, can be extracted based on a relatively small set of manually-developed rules (i.e., searching for instantiations of "<proteins> (0-5 words) <verbs> (0-5 words) <proteins>" within a biomedical treatise). However, the irregularity of informal text does not lend itself to

machine deconstruction according to predetermined patterns; rather, a machine learning capability is required for reliable extraction. Based on vast reference corpora containing millions of possible “translations” between informal segments and their equivalent organizationally-relevant transformations, a system should learn to “decode” additional, unknown text blurbs according to statistically-derived parameters. As such, we propose the following as our second and third design principles:

**Machine learning principle:** A design must allow a machine to learn from textual elements and associated codification (e.g., tags, classification, formalization) created by humans.

**Socio-technical principle:** A design must create a socio-technical system supporting human input (e.g., tagging, classification) and evaluation during the original corpus-building phase and ongoing machine learning.

***Ability to trigger timely resolutions.*** Business Intelligence (BI) software enables executives, managers, and analysts to make better and faster business decisions based on structured operational data that is analyzed in near-real time (Chaudhuri, Dayal, and Narasayya 2011). While the examination of historical social media text may serve an organization in certain capacities, a useful social media analytics tool intended to improve organizational decision making must similarly evaluate unstructured data, if not in real time, then quickly. Critical issues must be extracted at a rate sufficient to enable responsive problem solving by executives and managers. Based on these arguments, we propose as our final principle the following:

**Actionability principle:** A design must produce results in a timely output suitable for decision making such that organizations can make well-timed assessments and take rapid action.

### **3.6 Implications and Future Research**

The state of the art of natural language processing and text mining of highly unstructured data contextualizes this design science article as a work-in-progress of sorts due to the technical limitations constraining the instantiation of our design principles. We can only incorporate capabilities as they are honed to reliability in practice; as such, the implications of this manuscript offer an expansive agenda for IS research in conjunction with, and informed by, a variety of scholarly fields.

In parallel, if we can work out the ideal environment for the use of this type system we can employ a prototype to investigate whether efforts expended in further development are even worthwhile. This inquiry lends itself well to experimentation in order to determine if the theorized output of a system developed according to these identified principles can indeed improve decision quality. Thus, the next step in this stream of research is to model the expected output of the described social media text analytics system and, ultimately, compare the qualities of decision supported by this output versus those outputs yielded by currently-available tools used in practice including word frequency counts, context-free word-based assessments of sentiment trends, and even raw data in the forms of untransformed Twitter and Facebook feeds.

The realm of natural-language-based text mining faces many interesting and difficult technical challenges as computer science and linguistics scholars work to advance the reliability of this important capability. However, this area of research also promises rich opportunities to reshape the landscape of business intelligence as our knowledge of how to approach informal written language and extract information from it continues to thrive and experience new breakthroughs.

### 3.7 References

- Abbasi, A., and Chen, H. 2008. "CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication," *MIS Quarterly* (32:4), pp. 811-837.
- Agarwal, P., and Searls, D. B. 2008. "Literature Mining in Support of Drug Discovery," *Briefings in bioinformatics* (9:6), pp. 479-92.
- Anderson, J. D., and Pérez-Carballo, J. 2001. "The nature of Indexing: How Humans and Machines Analyze Messages and Texts for Retrieval. Part I: Research, and the nature of human indexing," *Information Processing & Management* (37:2), pp. 231-254.
- Argyris, C., and Schön, D. 1978. *Organizational Learning: A Theory of Action Perspective*, Reading, MA: Addison Wesley.
- Benus, S., Gravano, A., and Hirschberg, J. 2007. "Prosody, Emotions, and... 'Whatever'," in *Proc. of Interspeech* Antwerp, Belgium: Citeseer, pp. 2629-2632.
- Berelson, B. 1952. *Content Analysis in Communication Research*, New York, New York, USA: Free Press, pp. 220.
- Bird, S., Klein, E., and Loper, E. 2009. *Natural Language Processing with Python*, O'Reilly Media, pp. 479.
- Brohman, M. K., Piccoli, G., Martin, P., and Watson, R. T. 2009. "A Design Theory Approach to Building Strategic Network-Based Customer Service Systems," *Decision Sciences* (40:3), pp. 403-430.
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., and Mercer, R. L. 1993. "The Mathematics of Statistical Machine Translation: Parameter Estimation," *Computational Linguistics* (19:2), pp. 263-311.
- Bunescu, R., and Mooney, R. 2007. "Extracting Relations from Text: From Word Sequences to Dependency Paths," in *Natural Language Processing and Text Mining*, A. Kao and S. R. Poteet (eds.), Springer, pp. 29-44.
- Chaudhuri, S., Dayal, U., and Narasayya, V. 2011. "An Overview of Business Intelligence Technology," *Communications of the ACM* (54:8), pp. 88.
- Choudhary, A. K., Harding, J. A., and Tiwari, M. K. 2009. "Data Mining in Manufacturing: A Review Based on the Kind of Knowledge," *Journal of Intelligent Manufacturing* (20:5), pp. 501-521.
- Cohen, A. M., and Hersh, W. R. 2005. "A Survey of Current Work in Biomedical Text Mining," *Briefings in bioinformatics* (6:1), pp. 57-71.
- Davidov, D., Tsur, O., and Rappoport, A. 2010. "Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon," in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning* Uppsala, Sweden: Association for Computational Linguistics, pp. 107-116.

- Denecke, K., and Nejdl, W. 2009. "How Valuable is Medical Social Media Data? Content Analysis of the Medical Web," *Information Sciences* (179:12) Elsevier Inc., pp. 1870-1880.
- Feinerer, I., Hornik, K., and Meyer, D. 2008. "Text Mining Infrastructure in R," *Journal of Statistical Software* (25:5), pp. 1-54.
- Feldman, R., and Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, pp. 422.
- Fink, C. R., Chou, D. S., Kopecky, J. J., and Llorens, A. J. 2011. "Coarse- and Fine-Grained Sentiment Analysis of Social Media Text," *Johns Hopkins APL Technical Digest* (30:1), pp. 22-30.
- Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. 1987. "The Vocabulary Problem in Human-System Communication," *Communications of the ACM* (30:11), pp. 964-971.
- González-ibáñez, R., and Wacholder, N. 2011. "Identifying Sarcasm in Twitter : A Closer Look," *Computational Linguistics* (2010), pp. 581-586.
- Grassi, M., Cambria, E., Hussain, A., and Piazza, F. 2011. "Sentic Web: A New Paradigm for Managing Social Media Affective Information," *Cognitive Computation* (3:3), pp. 480-489.
- Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the Association for Information Systems* (8:5), pp. 312-335.
- Hearst, M. A. 1999. "Untangling Text Data Mining," in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* College Park, MD: Association for Computational Linguistics, pp. 3-10.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *Management Information Systems* (28:1), pp. 75-105.
- Hutchins, J. 2006. "Example-based Machine Translation: A Review and Commentary," *Machine Translation* (19:3-4), pp. 197-211.
- Inmon, W. H., and Nesavich, A. 2008. *Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence*, Prentice Hall, pp. 241.
- Kao, A., and Poteet, S. R. 2007. "Natural Language Processing and Text Mining," A. Kao and S. R. Poteet (eds.), Springer, pp. 272.
- Knight, K. 1997. "Automating Knowledge Acquisition for Machine Translation," *AI Magazine* (18:4), pp. 81-96.
- Kohut, G. F., and Segars, A. H. 1992. "The President's Letter to Stockholders: An Examination of Corporate Communication Strategy," *Journal of Business Communication* (29:1), pp. 7-21.
- Larson, K., and Watson, R. T. 2011. "The Value of Social Media: Toward Measuring Social Media Strategies," in *ICIS 2011 Proceedings* Shanghai, China, pp. 1-18.

- Lindvall, M., Rus, I., and Sinha, S. S. 2003. "Software Systems Support for Knowledge Management," *Journal of Knowledge Management* (7:5), pp. 137-150.
- Lopez, A. 2008. "Statistical Machine Translation," *ACM Computing Surveys* (40:3), pp. 1-49.
- Loughran, T., and McDonald, B. 2011. "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance* (66:1) Wiley Online Library, pp. 35-65.
- Markus, M. L., Majchrzak, A., and Gasser, L. 2002. "A Design Theory for Systems that Support Emergent Knowledge Processes," *MIS Quarterly* (26:3) JSTOR, pp. 179-212.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. a, and Aiden, E. L. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science (New York, N.Y.)* (331:6014), pp. 176-82.
- Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval* (2:1-2), pp. 1-135.
- Pavlou, P. A., and Dimoka, A. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," *Information Systems Research* (17:4), pp. 392-414.
- Sang, E. F. T. K., and Meulder, F. De. 2003. "Introduction to the CoNLL-2003 Shared Task : Language-Independent Named Entity Recognition," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL Association for Computational Linguistics*, pp. 142-147.
- Sarker, S., and Lee, A. S. 2002. "Using a Positivist Case Research Methodology to Test Three Competing Theories-in-Use of Business Process," *Journal of the Association for Information Systems* (2:7), pp. 1-72.
- Shatkay, H., and Feldman, R. 2003. "Genomic Era : An Overview," *Journal of Computational Biology* (10:6), pp. 821-855.
- UGA Working Paper: Larson, K., Watson, R. T., Piccoli, G., Osbon, K., and Green, L. 2011. "Building an Infrastructure to Support Textual Data Mining: The Case of the Letter to the Shareholder," *UGA Working Paper*, pp. 11.
- Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36-59.
- Wilson, T., Wiebe, J., and Hoffmann, P. 2005. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)* Vancouver, pp. 347-354.
- Yi, J., Nasukawa, T., Bunescu, R., and Niblack, W. 2003. "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques," in *Third IEEE International Conference on Data Mining (ICDM'03)* Melbourne, Florida, pp. 427-434.

## CHAPTER 4

# THE IMPACT OF NATURAL LANGUAGE PROCESSING-BASED TEXTUAL ANALYSIS OF SOCIAL MEDIA INTERACTIONS ON ORGANIZATIONAL DECISION MAKING<sup>6</sup>

---

<sup>6</sup> Larson, K. and Watson, R. T., to be submitted to *MIS Quarterly*.

#### **4.1 The Promise of Natural Language Processing**

Fundamental to the success of an organization is its ability to process information to reduce uncertainty (Galbraith 1974; Daft and Lengel 1986). Recently, a new mechanism for eliciting and disseminating information in such forms as consumer opinions, suggestions, and conversations (Demetriou and Kawalek 2010) has emerged, challenging organizations to develop and apply novel methods for unearthing potentially-valuable intelligence from these data. Broadly termed “social media,” this mechanism heralds both an increasing concern and an invaluable opportunity for firms whose strategies include leveraging consumer-generated qualitative data to create business value (Culnan, Mchugh, and Zubillaga 2010; Hoffman and Fodor 2010). Of particular import, social media necessitate new tools for real-time mining of consequential information underlying ever-increasing volumes of continually generated textual data. Given the state of existing social media monitoring tools, in particular the gap between actual and desired capabilities for extracting latent information, a central question for social media researchers is whether a theoretically-informed, natural language processing (NLP) approach to text-data analytics can confer an informational advantage to organizations beyond prevalent approaches currently available.

Defined from a consumer/firm perspective as the set of connectivity-enabled applications that facilitate interaction and the co-creation, exchange, and publication of information among firms and their networked communities of customers, social media engender multiple complex layers of brand-centric text-mediated interactions. Of particular relevance to firms is the layer comprising customer-to-customer interactions such as recommendations, reviews, collaborative exchanges, and helpful suggestions or advice (Larson and Watson 2011). For firms to detect among these interchanges



important cues such as adverse event mentions and consumer reactions to new products, social media analysts and managers require the ability to qualitatively mine textual data possibly symbolizing and conveying these cues. This level of measurement exceeds the simple positive-negative labeling inherent in sentiment analysis (Pang and Lee 2004, 2008) and the simpler measurement technique of gathering count data for characteristics such as number of followers, number of likes, etc.—important but incomplete methods for extracting knowledge from qualitative consumer-generated data.

Despite the prevalence of sentiment analysis as the basis of many social media brand reputation monitoring tools, we point out the large degree of meaning and knowledge potentially lost by simply sorting suggestions, comments, and complaints into negative and positive categories, regardless of the number of intervals into which the continuum may be subdivided. For example, a comment about the problematic sticking of keys on a keyboard may be scored as highly negative (depending on the individual words chosen to express the problem), thus signifying a potentially important problem for the brand or firm. However, to extract the subject of the customer's negativity, further processing must occur because sentiment analysis does not provide a mechanism for isolating the topic of concern or its context. Either an individual reader must further manually examine the negative comment to determine its significance (teams of which organizations employ at great cost), or some type of machine-based algorithm must be further applied for qualitative analysis, which leads back to the original requirement of a tool capable of *contextual* text data mining.

Currently available social media tracking or monitoring tools also heavily rely on keyword searches and alerts to the firm based on the appearance of pre-specified

keywords and filters. While the ability to detect comments or suggestions regarding key topics is a necessary social media monitoring capability, it is insufficient for *meaningful* analyses of brand-oriented social media content. Keyword detection, a confirmatory as opposed to exploratory device, does not provide a mechanism for the unearthing of new, possibly critical issues not yet identified as critical by the firm and thus not pre-specifiable as search terms. Keyword searching can only continue to confirm a company's hypotheses about what is important to its customers, not readily facilitate the discovery of new or latent knowledge about them and their preferences. An informal examination of widely available free and for-fee tools confirms that keyword searching is largely a manual activity, and not highly automated beyond the pre-specification of alerts by humans for machine execution. For example, the automated aggregation of an evolving set of critical search terms based on machine learning feedback loops, while a useful possibility, is not yet a commercially-available functionality as far as our research indicates; this type of intelligence accumulation still requires human agency. Many tools simply serve as an interface between a human reader and social media texts: *Twitter Advanced Search*, for example, enables the user to "look for keywords, search by location, date, or with other filters."<sup>7,8</sup> *Track* enables users to "track keywords and have them sent directly to"<sup>2,9</sup> a mobile phone. *Monitter* allows managers to "monitor Twitter for key words, phrases and topics being discussed online at a glance."<sup>2</sup> In each of these and most other cases, it is important to note that a person must manually process the

---

<sup>7</sup> <http://www.brandmill.com/featured/100-social-media-listening-tools/>

<sup>8</sup> <http://www.socialbrite.org/2011/01/11/guide-to-free-social-media-monitoring-tools/>

<sup>9</sup> <http://socialmediatrader.com/tracking-the-buzz-tools-to-monitor-your-brand-effectively/>

results of an automated keyword or other search in order to read the keyword within its context and derive meaning from the search.

Foundational to the lack of automation available to social media managers, a recent review of current capabilities available for mining textual data delineates myriad problems complicating efforts to extract previously unknown knowledge from “highly unstructured text.” This freeform text, proliferating by the second across social media applications, characteristically lacks boundaries of subject, grammar, structure, and even spelling. Challenges to automatically processing this type of highly unstructured text frequently reflect limitations imposed by the tactical choice of porting unstructured text into a structured environment, a process that involves the decomposition of sentences into words that can then be easily stored, retrieved, and evaluated. While the advantages of this methodology for dealing with text include simplification of the analytical process by freeing the analyst from concerns regarding preservation or comprehension of context (Inmon and Nesavich 2008: xix), this purported benefit also gives rise to major drawbacks stemming from the examination of decontextualized words instead of sentences, phrases, or chunks of messages as the relevant unit of data. Such weaknesses range from difficulties in resolving sarcasm or anaphora (expressions whose meanings depends on other referential elements) to the inability to decipher simple spelling errors.

In direct contrast with this structured approach to text mining, an alternate composite field blending computer science, machine learning, and linguistics research aims to extract meaning from texts by considering them in their natural language format. This field, natural language processing (NLP), encompasses a wide range of disciplines and tasks focused on extending the capabilities of text mining, or the

extraction of knowledge from unstructured text (Hearst 1999), most recently by incorporating the machine-learning paradigm of language processing. NLP algorithms have met with some success in structured domains with limited lexes such as medicine and biochemistry (Tanabe et al. 1999), fields in which knowledge acquisition is ontologically bounded (Maedche and Staab 2000; Wilcox and Hripcsak 2003).

Recent reinvigoration of NLP-related research has shepherded progress in the technical capabilities of machines to discover new, non-trivial knowledge from free text, although the automated mining of data from unstructured text is still in its relative infancy. Emerging subfields and approaches continue to extend text mining proficiencies in the contexts of real-world data. For example, improved automation of lexicon augmentation in named entity recognition, or the accurate labeling of persons, organizations, and locations (Sang and De Meulder 2003), increases the body of task-specific lexicons available for a variety of natural language processing tasks. Thus, instead of relying on general-purpose lexicons or tediously and slowly compiling task-specific lexicons by hand, highly tailored lexicons can now be built on the fly by leveraging named entity extraction from HTML data on the Web via a search engine (McCallum and Li 2003). Similarly, incremental improvements to a wide range of specific capabilities such as parts-of-speech tagging, parsing (determining the grammatical tree of a sentence), and anaphora resolution (determining which noun or name a pronoun refers to) combine to contribute to discipline-level progress and suggest potential applicability in less-structured or unstructured text environments such as social media (Bunescu and Mooney 2007; Kao and Poteet 2007; Agichtein et al. 2008).

Now virtually ubiquitous, the social media environment facilitates a type of collective *textual* interaction among organizations, communities, and individuals. We recognize a natural alignment between the knowledge discovery goal of NLP-based automated text mining and the organizational goal of extracting knowledge to create business value from highly unstructured text interactions (e.g., comments, opinions, and suggestions) propagated by customers across social media platforms. Given this apparent alignment, we are interested in determining whether NLP-based approaches may indeed prove more useful to firms for information extraction or whether existing manual or basic sentiment-based techniques are sufficient for leveraging consumer-generated text in ultimate support of organizational decision-making. It may be the case that current techniques commonly used—namely, sentiment analysis, or the sorting of texts into piles of positives, negatives, and sometimes neutrals based on word-level calculations—are “good enough”: adequate for detecting critical problems and opportunities from highly unstructured Tweets, updates, and comments. Were this the case, we would be in a position to inform both research and practice regarding the development of NLP-based social media analytics tools for the purposes of knowledge discovery, as well as contribute to theoretical understanding of text mining in an unstructured environment. In such a scenario, the more practical assumption would conclude that efforts expended by computational linguists, artificial intelligence programmers, and computer scientists to develop machine understanding of unstructured text would be more usefully channeled into domains characterized by constrained forms of text.

Intuition, however, leads us to suspect that organizational text data mining tools such as NLP-based social media analytic systems will in fact prove critical to the

decision-making processes of executives and managers of firms in this day of pervasive, application-mediated exchange of ideas, opinions, and suggestions. Assuming that the detection of significant problems and opportunities promulgated across various social media platforms by consumers can improve downstream decisions made by a firm, it then follows that the capabilities conveyed by an NLP-based social media analytics tool would benefit firms and consumers alike, thus warranting the continued investment of time and intellect by relevant specialists.

Once we have resolved this practical problem of substantial monetary and commercial impact by determining *whether* a proposed NLP-based technique improves decision making, we then need to theoretically understand *why* this is the case. Unlike typical IS research in which we are able to generate propositions based on theory, in the current research setting we have neither a body of extant strong theory from which to draw nor existing practice to observe in order to generate new theory. Thus, we are faced not only with immediate questions regarding the practical ramifications of NLP-based approaches to textual data analysis, but also with more abstract questions concerning how researchers in general should handle a lack of an overarching guiding theory in combination with an absence of observable practice.

In order to resolve the practical question posed, this experimental investigation aims to answer the following practice-oriented research question:

*RQ<sub>Practical</sub>: Can advanced natural-language-processing-based qualitative textual analysis techniques improve the decision-making capability of organizations?*

On a more conceptual level, we are additionally concerned with contributing to the IS research literature by establishing guidelines for cases in which the researcher has no strong guiding theory and where extant practice is clearly deficient (which we argue is the case with sentiment analysis of highly unstructured text). In this type of research setting, information systems design is driven not by an overarching theory or observation of direct practice, but by observation of solutions to similar, parallel problems that may or may not be explained through the application of a series of theoretically based insights. Through studying analogous scenarios, we expect to improve the focal practice by implementing a logically and practically designed information system that can then itself be observed and tested to generate new theory. This mode of scientific inquiry is often seen in applied fields such as medicine and engineering, yet is discussed very little in our practical field. Information systems theories are generated, ostensibly, for the purpose of improving practice. Sometimes, however, practice may still be improved despite a lack of adequate theoretical explanation.

We point to particularly salient examples of critical outcomes in applied fields to illustrate valid instantiations of this phenomenon (see Table 4-1). Notably, the entire paradigm of medical clinical trial research exemplifies the concept of putting into practice interventions not always theoretically predicated; experimentation is often the basis of gaining insight into the effectiveness of a therapy, resulting in diagnostic-based as opposed to theory-driven treatments (Freedman 1987). The practice of evidence-based medicine, in fact, relies on the concept of “best available external clinical evidence” from analogous cases, specifically in applying patient-centered clinical research to increase accuracy and precision of diagnostic tests (including clinical

examinations) and increase the efficacy and safety of therapeutic, rehabilitative, and preventive regimens (Sackett et al. 1996). Notably, external clinical evidence proceeds on the basis of invalidating previously accepted tests and treatments, effectively replacing them with more powerful, accurate, efficacious, and safer versions based on observation.

**Table 4-1**  
**Examples of practice instituted prior to theory development**

| Domain   | Focal Topic     | Theoretical Standing   |
|----------|-----------------|--|
| Physics  | Superconductors | In 1986, scientists discovered a new class of ceramic superconductor with the capability to conduct current with zero resistance at much higher temperatures than previously observed. As of now, we have no theory to explain this phenomenon, yet these materials are used in practice and are the subject of thousands of published research papers (Buchanan 2001; Varma 2010).  |
| Medicine | Antibiotics     | In 1847, a Viennese obstetrician noticed the dramatically high incidence of death from puerperal fever among women attended by doctors, but not midwives, during delivery. He drew a connection between puerperal fever and examinations by doctors coming directly from autopsy, subsequently instituting the practice of hand washing with chlorinated lime water before obstetrical examinations. Despite the absence of an explanatory germ theory, this step reduced childbirth mortality from 18% to 2.2% (Carter 1985). |
| Medicine | Antidepressants | Scientists do not know why antidepressants work on the human brain (Schafer 1999); despite a lack of understanding of action mechanisms and the proliferation of competing explanatory theories (Harmer, Goodwin, and Cowen 2009), U.S. doctors wrote over 260 million prescriptions for these drugs in 2011 (Anon. 2012).   |



In our quest to elucidate the IS research process for realms characterized by no guiding theory and no (or inadequate) observable practice, we ask the following theoretical question:

*RQ<sub>Theoretical</sub>: How should scholars advance knowledge in research areas characterized by a lack of strong guiding theory and inadequate or no observable practice?*

#### **4.1.1 Evidence-based IS**

To situate an “evidence-based” IS research approach similar to practices observed in other applied fields, we propose an understanding of IS research approaches according to dimensions of 1) strength of underlying theory and 2) quality of observable practice (Table 4-2). Though a novel area of research may begin based on observation of best practices in an organization from which theory is subsequently developed and tested (quadrant II), traditional IS research generally occurs in quadrant I, characterized by streams of research supported by existing (and extended) theory and grounded in organizational practice.

**Table 4-2**  
**Theory and practice in IS research**

|                           | High practice quality   | Poor, inadequate, or no practice  |
|---------------------------|---|---|
| Strong guiding theory     | <b>Quadrant I:</b><br>test existing theory,<br>propose new theory<br>(e.g., TAM, DOI, etc.) | <b>Quadrant III:</b><br>revise theory;<br>alternately, discover theory does<br>not work in practice (and no better<br>theory exists)<br>(e.g., machine translation “1.0”) |
| Weak or no guiding theory | <b>Quadrant II:</b><br>propose theory<br>(e.g., superconductor<br>research)                 | <b>Quadrant IV:</b><br>learning paradigm:<br>design and test; theory not here,<br>logical enterprise might work<br>(e.g., NLP-based analytics)                            |

Early machine translation (MT) is a relevant example illustrating the effects when the application of an overarching theory results in inadequate practice (quadrant III): theoretically-designed translation algorithms grounded in rules of human language (what we in this paper dub MT “1.0”) failed to produce a sustainable practice; the result was inadequate, inconsistent (unusable) translations from one natural language to another (Hutchins 1995). A clear methodological failure, the theoretical approach based on the notion of universal underlying principles of natural language was abandoned in favor of a statistical approach, the result of which are current-generation machine-learning-based statistical machine translation (SMT) algorithms capable of producing consistently reliable translations (although, it should be noted, there is no system yet capable of providing the holy grail of “fully automatic high quality translation of unrestricted text,” or “FAHQUT” (Anastasiou and Gupta 2011)).

The current research context is neither constrained by theory nor enabled by the examination of active practice (quadrant IV); instead we are situated within a learning paradigm and must look to parallel practice to help us ground in logic a viable new practice facilitated by a novel information system. Both relevant and recent, the analogous progression of statistical machine translation toward successful output provides applicable guidance for our design choices, based on the reckoning that just as translation from one language to another has proven to require a machine learning approach, so, we hypothesize, will the “translation” (interpretation) of highly unstructured text into meaningful intelligence capable of supporting organizational decision-making.

This paper makes the following contributions to theory and practice of unstructured data analysis. First, we compare the usefulness of sentiment analysis to other approaches used to derive intelligence from highly unstructured text, specifically the type of text that characterizes social media interaction. We empirically demonstrate that word-based sentiment analysis of social media text is no better than random sampling as the basis for grasping problems underlying customer-originating messages. Furthermore, we demonstrate that the use of a natural language processing-based approach can substantially enhance an analyst’s ability to detect problems with potential importance to organizational decision-making.

We also extrapolate from this particular study insight into the higher-level process of conducting meaningful research with practical implications in an area deficient of either strong guiding theory or feasible existing practice.

The paper continues as follows: the following section provides a context for this study, reviewing both important concepts of natural language processing as well as the

problems preventing effective sentiment analysis of certain types of text. Based on these concepts, we highlight certain features critical to a system capable of analyzing social media messages in a meaningful way. The third section proposes our model and propositions; section four describes our research methodology; section five presents the results from our experiment; section six discusses the findings and implications of our study and concludes the paper with some future research directions.

## **4.2 Literature Review and Development of Design Principles**

The mining of text encompasses a vast array of theoretical approaches and methods (Feinerer, Hornik, and Meyer 2008), including information retrieval, clustering, classification, entity-relationship and event extraction, and natural language processing (Hotho, Andreas, and Paaß 2005), each the focus of intense ongoing research. While these techniques are all interrelated in terms of their practical goal and are each likely to contribute in parallel to the ultimate success of machine-supported analysis of text, of particular interest to us in our quest to develop an approach robust against the idiosyncrasies of highly unstructured user-generated text is natural language processing.

The field of natural language processing (NLP) is too vast to cover in a single manuscript and in fact relates to each of the approaches and methods listed previously as well as comprising a long list of additional subfields; as such, we focus our review on those aspects particularly germane to the current research problem. Specifically, we orient our discussion of techniques to the analysis of highly unstructured text such as that generated within social media platforms. The general goal of NLP is to create algorithms capable of “understanding” natural language through techniques ranging

from the simple manipulation of strings to the automatic processing of natural language inquiries (Hotho, Andreas, and Paaß 2005). This methodology contrasts, for example, with information retrieval (IR), the goal of which is to return units of text matched according to pre-specified patterns. IR is essentially the confirmatory counterpart to NLP, although NLP can be incorporated into IR algorithms to increase their effectiveness through increased clarification of word ambiguity (Arazy and Woo 2007).

In contrast to NLP's potential to actually mine or discover meaningful intelligence from text, the immediate need of organizations to analyze their customers' messages propagating across myriad social media platforms has forced the reliance on an available, but we argue ineffective *in this setting*, methodology for deriving knowledge from customer chatter. Far less complex and sophisticated than NLP, sentiment analysis is the assessment of positive and negative customer sentiment in product or brand reviews or mentions that organizations use to "measure" the emotion underlying segments of consumer-generated text (Pang and Lee 2008). While sentiment analysis has the potential to impart insight into customers' reactions to a given organization and its products and services, we are somewhat dubious of the practical accuracy of this methodology on an automated and large scale. We argue that any analysis of text at the word level, which necessarily ignores the aggregate meaning of whole clauses, is tremendously susceptible to the misclassification of sentiment conveyed by idioms, negations, irony, and sarcasm. Misspellings further limit the validity of sentiment analysis. We also refer to recent demonstrations of the range of granularity lost through sorting comments and reviews according to simple negative/positive rating scales (e.g., Pavlou and Dimoka 2006).

Our assessment of the capabilities conveyed by sentiment analysis leads us to conclude that deeper insight into the impact of a given product or service requires a correspondingly deeper level of qualitative analysis. Toward the goal of deriving valuable business intelligence from consumer-generated unstructured data, organizations must become capable of qualitatively analyzing textual data on a large scale in near-real time, similar to the operational capabilities of just-in-time BI that reduce the latency between data acquisition and analysis (Chaudhuri, Dayal, and Narasayya 2011). However, a survey of current technologies indicates that while ratios of positive to negative mentions and counts of users and actions are readily available to firms, the ability to glean deep meaning from voluminous streams of social media-generated data is an expertise yet to be fully established and incorporated into organizational business intelligence-oriented monitoring.

#### **4.2.1 Design principles and their implications**

Ultimately, our observations of the capability gaps between desired social media analytics and the technologies currently available to firms lead us to develop a set of design principles that embody an ideal information system for mining highly unstructured text.

Dissecting sentences into buckets of unrelated words decontextualizes each instance of each word, disengaging a customer's intent from the *assembly* of words used to express that intent. Mental-models research indicates that humans understand patterns of words locally; multiple instances of a single word situated among different surrounding words are not perceived as semantically related by most speakers of English (Fox 1986). For example, we do not consider "my soup is *cold*" to have any

relation to “I have a head *cold*.” But if we extract *cold* from the rest of the sentence in which it exists, which is equivalent to what happens during sentiment analysis or other non-NLP based approaches, we then have no idea what the word actually means or whether it should be interpreted as a positive, negative, or neutral sentiment.

Collocation indexing, or the process of extracting overall syntax based on the identification of word combinations that carry specific meaning in natural text, has proven successful at word disambiguation in large scale systems that use naturally occurring text (Arazy and Woo 2007). This statistical NLP technique has proven to reduce the gap between the way humans think of information and the way in which it is represented by machines (Arazy and Woo 2007). At the very least, a reliable sentiment analysis approach to social media analytics would require the incorporation of this NLP-predicated capacity to identify meaningful word combinations with meaning separate from that of their individual components; human communication is replete with such complex expressions.

Such complexities as these force humans to rely on context to resolve a variety of textual puzzles, especially when deciphering highly unstructured text that, unlike formal written communication, does not benefit from boundaries or correction of syntax and semantics. Conceptually similar to collocation, humans consider neighboring words as we decipher an author’s original intent, not just when interpreting backward-referring and other ambiguous parts of speech and idiomatic turns of phrase, but also when decrypting misspellings, filling in omitted words, deciphering inaccurate autocorrections, and resolving sarcasm. Considering the critical loss of meaning inherent in any word-based approach unable to consider context, a critical desideratum of a useful social media analytics system is contextual sensitivity. We specify:

***Context-sensitivity principle:*** *A design must create output that is contextually situated by supporting sentence-level analysis, sentence-sequence analysis, and multi-party conversation analysis.*

The automated mining of text can be likened to the task of machine translation (MT) in that the goal of both is to interpret one set of words and translate them into an output of similarly-intended set of words, but in a form that is understandable to the recipient. Thus, while a language translator converts French sentences into English, a social media analytics system would interpret a Tweet or status update into an output that is meaningful to the organization deciding how it should react to the message. The output in this case may look like a phrase or sentence that conveys the latent (or even manifest) intent of the original text in terms relevant to the brand or organization.

Statistical machine translation (SMT) has emerged as the dominant, even “mainstream” machine translation approach over the last decade or so despite the competition of theoretically-driven, rules-based alternatives (Hutchins 2006). These theory-driven methods did not prove robust in practice and so subsided to a corpora-driven MT model based primarily on word frequency and word combinations derived from large volumes of real data (Hutchins 2006). Given the similarities of task, goal, and amount of data with which to begin training, we conclude that an effective approach to the “translation” of social media data into business intelligence should follow a parallel methodology that we label statistical machine interpretation (SMI). The basis of SMT, and subsequently SMI, is machine learning, a paradigm that calls for general learning algorithms typically grounded in statistical inference. Statistical machine translation and machine learning are interrelated in their analysis of large corpora of



real-world data during the training phase, from which an evaluation model is subsequently derived for new sentences (Lopez 2008).

Observation of the pattern of evolution of MT indicates a clear failure of predicating practice on grand linguistic theory, which is not to say that aspects of successful statistical machine translation cannot be explained by an amalgam of multiple theoretical insights or that a strong theory will not be discovered that explains the phenomenon. Inferring from this precedent, we presume that designing machine interpretation from a theory of universal human language will be similarly impractical. From a practice standpoint, our goal is to gain immediate insight into the application of SMI to the analysis of unstructured text. In order to achieve that objective in a useful way, we turn to statistical methods that we know work in a parallel methodology. Once a feasible interpretation system is designed and executed, then as academic researchers we will be better positioned to potentially explain phases or characteristics of the process from a theoretical perspective, backing into aspects of social theory, as it were, to extend or clarify our understanding of SMI. Any grand social theory that we may turn to at this point in the development of the technology, however, is simply not programmable in order to be practically functional.

Statistical machine learning is of critical importance to resolving tasks that cannot be solved strictly by classical programming techniques due to the lack of an available mathematical model, such as in instructing a machine how recognize the letter “A” in human hand; because it is not known how to write a program to perform this type of hand-written character recognition, the machine can instead be trained from examples (Cristianini and Shawe-Taylor 2000). SMT can improve a wide range of automated processes predicated on unstructured data; for example, it has enhanced the

efficacy of automated detection systems for combating fraudulent websites (Abbasi et al. 2010). Statistical learning theory (SLT), also known as the Vapnik-Chervonenkis theory, is the underlying computational learning theory that describes the learning process from a statistical perspective. Purely theoretical until the 1990's, SLT has since bolstered the development of highly effective algorithms, in particular support vector machines (SVM) (Vapnik 1999). SVMs are SLT-based learning algorithms belonging to the kernel methods class of pattern analysis, that, given a set of data, find patterns by embedding data into high dimensional feature space and looking for linear relationships in that space (Cristianini and Shawe-Taylor 2000).

Based on the dominance of SMT in performing accurate machine-based translations of text, and given the capabilities of SVMs to map new data points onto a space built from training data in order to predict categorization of the new information, an optimal social media analytics system should be grounded in similar mechanisms:

***Machine learning principle:*** *A design must allow a machine to learn from textual elements and associated codification (e.g., tags, classification, formalization) created by humans.*

Natural language processing scientists point out that progress in the field of statistical machine translation is largely driven by the availability of data (Koehn 2005). SMT thrives on the perpetuation of large quantities of parallel texts: original text paired with its translation into a target language. The process for translation, specifically between two natural languages, generally embody the following steps (Koehn 2005):

1. Gathering raw data (by crawling the web, or scraping social media sites)
2. Extracting and mapping parallel chunks of text (document alignment)
3. Breaking text into sentences (sentence splitting)
4. Preprocessing the corpus in preparation for SMT systems (normalization, tokenization)
5. Mapping original language sentences to target language sentences (sentence alignment)

This general procedure for translating between languages assumes the existence of many parallel texts available for alignment, for example the vast reserve of identical documents available in multiple languages on the United Nations website. However, for the task of social media text “translation” grounded in SMI, while we continue to accumulate vast amounts of “monolingual,” or original language data, on a daily basis, we lack correlated interpretations in forms usable by decision makers. Thus, our goal for an automated social media analytics system requires an additional preliminary step of manual interpretation on a large-scale basis to establish parallel corpora from which to derive a probability model for understanding future sentences.

The theoretical implications of creating such a body of mass translation, and by extension a body of less formal mass interpretation are numerous; examination of the translation genre within linguistics studies reveals a wide array of potential weaknesses or sources of interpretation variance. The generation, or *surface realization* step during which the “reasoner” produces a final readout, may diverge according to 1) available vocabulary and syntactic resources, 2) judgment as to what information should be explicitly restated and what might be left to inference, 3) the translator’s rhetorical

proclivity and 4) coherence and 5) cohesion as expressed through sentence structure and word distribution, and 6) the interpreter's ability to *find* a mapping of the information to a form that is linguistically expressible (McDonald 1993). Universal features additionally affecting translation as proposed in the literature include 7) simplification (the concept that translators subconsciously simplify the language, message, or both during translation), 8) explication (the tendency to spell things out, often by adding background information to the translation), and 9) normalization or conservatism (the tendency to exaggerate features of the target language and to conform to its typical patterns) (Baker 1996: 176). These multiple concerns are mapped to the domain of social media text interpretation in Table 4-3.

**Table 4-3**  
**Sources of inaccuracies across text interpretations**

| <b>Formal language translation<br/>(SMT)</b>                         | <b>Interpretation of unstructured text<br/>(SMI)</b>                                       |
|--|--|
| Limited vocabulary and syntactic experience                          | Limited organizational experience and domain knowledge                                     |
| Uncertainty regarding what information should be explicitly restated | Uncertainty regarding what information is critical to firm managers                        |
| Rhetoric   | Insertion of idiom or slang into interpretation  |
| Incoherence  | Insufficient level of logic or lucidity  |
| Lack of structural cohesion  | Ungrammatical interpretation   |
| Inability to map information to expressible words                    | Difficulty interpreting consumer language into meaningful knowledge about brand or product |
| Simplification   | Inadvertent omission of critical information   |
| Explication  | Insinuation of personal experience onto interpretation                                     |
| Normalization  | Incorrect projection of target vocabulary such as supposed organizational lingo            |

Recent work implies a utilitarian mechanism for contriving the critical mass necessary for viable machine interpretation, comparing the crowdsourcing of translation with statistical Machine Translation, noting that both approaches are able to cope with high input volume at high speeds (Anastasiou and Gupta 2011). Drawing on these

compatibilities and motivated by consideration of the potential for bias in any translation or interpretation, we suggest that an ideal strategy for interpretation would *combine* crowdsourcing and machine learning, employing crowdsourcing not as an alternative to SMI, but as its means. By leveraging aggregate information held by a large number of people, we can achieve the construction of parallel corpora from which an algorithm can then learn to interpret new data. Studies of this type of “prediction market,” which relies on the information contributions of a wide range of independently-acting people with some knowledge of the subject, has demonstrated that the averaged input of knowledgeable yet diverse masses is effective and surprisingly accurate (Sunstein 2006). Numerous case studies from the fields of economics and psychology illustrate the concept that the aggregation of information from groups of people often results in decisions that are better than what could have been determined by any single member of the group (see Surowiecki 2004); numerous inputs increase accuracy and clarity and tend to approach equilibrium (Leslie 2003).

Application of these advantages to the interpretation task heralds an averaging-out of inconsistencies across conversions that would help prevent building a probability based on extreme cases. By offering up thousands of segments of social-media-generated text to a large body of “reasoners” via a mechanism such as Amazon Mechanical Turk for interpretation, it is feasible to build a large pair of parallel texts. A variety of quality control mechanisms can be instituted such as a pre-test to screen out interpreters without a reasonable level of comprehension or language skills, as well as a range of manual post-translation checks by subject-matter experts. Interpreters would need a high level of proficiency in “texting” language, able to interpret ubiquitous yet cryptic texting short cuts such as “\*\$” (Starbucks), “<3” (love), “<333” (*really* love), or

“&3” (the HTML code for heart used to indicate love or positive affect toward an object). Additionally, interpreters would need to be well-versed in a variety of customs peculiar to specific social media platforms such as “RT” (retweet, or a message repeated by a different account on Twitter), “MRT” (modified retweet, indicating added or updated information), “dorbs” (Facebook abbreviation for adorable), etc. that might dictate or even reverse meaning of a consumer complaint, suggestion, etc. and that would need to be factored into the interpretation process.

Reiterating the difficulties inherent in the analysis of text potentially featuring idiomatic expression, sarcasm, slang, argots, and/or regional or communal vernacular, a system must be capable of contextual decryption. As a corollary to the contextual-sensitivity and machine-learning principles meant to address this need, a system must further include some mechanism for *evolving* its contextual capabilities and evaluation algorithms built from parallel corpora. A viable system for interpreting human language must be capable of evolving at the same rate as human colloquialism changes.

Building an SMT system for a language is largely contingent on the availability of parallel texts (Koehn 2005), whereby the core of the language model in the target language is the probabilistic phrase translation table learned from the parallel corpora. Considering the potential implied by crowdsourcing to create these corpora, a critical component of an ideal social media analytics system must be continuous input from human agents:

***Socio-technical principle:*** *A design must create a socio-technical system supporting human input (e.g., tagging, classification) and*

*evaluation during the original corpus-building phase and ongoing machine learning.*

In the current technological era, storage and processing capacities do not pose the roadblocks they once did. Parallel processing and inexpensive drives have obviated prior barriers to immediate evaluation of information. Organizations rely on real-time data analytics to support tactical decision-making and business processes, which in turn help firms become more customer-centric, increase revenues, and decrease costs (Watson et al. 2006). Well-informed business decisions rely on succinct and accurate models based on massive amounts of practical data (Wang 2007). However, these models have traditionally incorporated only structured data, meaning that massive amounts of qualitative data have not factored into major organizational decisions. It is estimated that perhaps as much as 85 percent of organizational data exist in the forms of emails, corporate documents, news articles, web pages, voicemail transcriptions (Lindvall, Rus, and Sinha 2003), and now social-media-facilitated comments, complaints, and suggestions, none of which typically factor into the processes of deriving business intelligence and, ultimately, creating business value (Feldman and Sanger 2007). Particularly in combination with mobile devices, consumer-generated data are produced and shared more quickly than ever and have a shorter lifespan, driving the development of real-time, database-intensive, scalable systems capable of handling the volume of data facilitated by social media platforms (Grinev et al. 2011). Given that processing data in batches is too slow to provide real-time support and that accumulated data can lose its significance in hours or even minutes, an effective social media analytics system

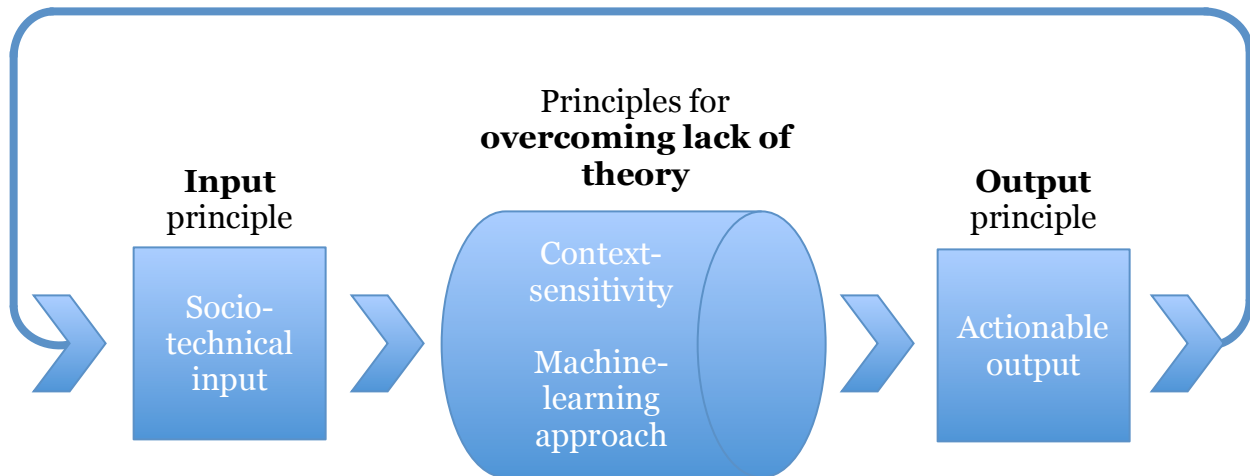


must process qualitative data at rates similar to structured data such as with traditional BI analytics:

***Actionability principle:*** *A design must produce results in a timely output suitable for decision making such that organizations can make well-timed assessments and take rapid action.*

#### **4.2.2 Exhaustiveness of design principles**

The design principles explicated above can be mapped to a traditional I/O diagram (Figure 4-1, below) to support the proposition that these are an exhaustive set of concepts for comprehensively informing a superior social media analytics system. We specify an input requirement, namely that the input must be derived from a socio-technical approach, and we specify a necessary condition for the output, namely that system output must support timely decision-making as opposed to historical analysis. Further, we redress the lack of strong theory available to guide our design by incorporating rich theoretical insight into the process requirements for unstructured data, namely that they be contextually evaluated by a machine learning algorithm capable of evolution. We further situate the process within a continuous loop of learning such that output is iteratively re-evaluated for legitimacy and input is dynamically replenished to ensure the maintenance of accurate and current interpretation models.



**Figure 4-1**  
**Input-process-output map of design principles**

To summarize, the literature supports our interest in natural language processing as the basis for a superior knowledge mining approach to sentiment analysis. Our review also reinforces the motivation behind our design choices, indicating that there is a need for dynamically rendered contextual preservation in any meaningful system for qualitatively analyzing highly unstructured text. By modeling the lessons encountered by machine translation over the decades, a statistical approach is the most likely route to a system appropriate to mine social media text data. An expedient and, according to the literature promising, mechanism for generating the necessary parallel interpretations is to crowdsource the task to a multitude of contributors. And as with any analytics system, results should be readily available to support decision-making. Further, any output upon which evaluation models are predicated should be repetitively re-evaluated for currency and accuracy.

### **4.3 Propositions and Model**

The major question we expect to answer through this investigation is whether an advanced, natural language processing (NLP)-based qualitative analysis technique predicated on machine-learning can improve the decision-making capacity of organizational managers, specifically in the context of highly unstructured text generated by consumers within social media environments. We derive from the literature discussed in the previous section a set of propositions whose outcomes will increase our knowledge of this domain.

Our discussions with practitioners indicate that some organizations employ teams of social media analysts to manually sort through and construe customer comments. This manual method, while likely effective for deriving knowledge due to the mechanisms of human perception, interpretation, judgment, and reasoning, is conversely not very efficient nor cost-effective once work hours are factored in. Additionally, interpretations that rely on the knowledge base of a single individual are, as indicated in earlier discussion, likely to be less accurate than an interpretation based on the aggregate wisdom of many. Further, when considering a team not of eighty or even twenty individual interpreters, but a team of one or two analysts, it is reasonable to expect information overload to occur once a tipping point is reached by the person (or two or three) attempting to read and interpret multiple streams of real-time social media data. Information overload occurs when the amount of input into a system exceeds processing capacity (Speier, Valacich, and Vessey 1999); as humans are limited in their cognitive processing capacity (Miller 1956), excess input causes information overload which in turn reduces the quality of decision-making and increases both time to decision and confusion during the process (see Speier 1999 and Eppler and Mengis

2004 for thorough IS-centric and cross-discipline reviews of information overload literature). In short, information overload impairs decision-making. We are interested in facilitating the opposite.

Taking a cue from the cognitive load concerns of information overload-related literature, we expect the manual processing of a large volume of real-time social media data to reduce decision-making capacity in an organizational setting by decreasing the number of problems and opportunities an analyst is able to identify, decreasing the accuracy of his or her assessment of these problems and opportunities, and subsequently reducing confidence of the analyst in his or her assessment. Particularly when real-time evaluation of rapidly-evolving (as opposed to static, historical) content is required, we expect deteriorated ability on the part of the analyst to decide accurately what messages convey critical problems and opportunities for the firm, accompanied by decreased confidence by the analysts that he or she thoroughly and accurately identified all the relevant problems and opportunities.

In contrast to organizations that employ manual analysis of social media, the prevalent automated approach currently underlying most social media monitoring systems is sentiment analysis. Despite its inadequacies as a measurement tool for highly unstructured text as argued earlier and supported by the literature, sentiment analysis is nonetheless a heuristic meant to sort negative messages from positive ones in a mechanical (quick) fashion. Because it conveys at least a rudimentary mechanism for sorting and narrowing, we reason that using the sentiment analysis output of a large volume of data will offer *some* degree of advantage over manual analysis of the same data. Diminishing the risks of cognitive overload and accelerating the decision-making process by starting with the assistance of some set of pre-identified messages instead of

the entirety of the original data set should ostensibly enable greater accuracy of problem and opportunity assessment, and in turn increased confidence in these assessments, than by starting with the full glut of information. An individual should be able to identify a greater number of problems and opportunities from a narrowed set than a full stream of social media messages. As such, we propose regarding sentiment analysis the following:

**Number of detected critical problems and opportunities**  
(sentiment analysis-based)

***P1a:** Individuals assisted by sentiment-based machine analysis of social media content will detect a greater number of key problems than individuals with no machine assistance.*

***P1b:** Individuals assisted by sentiment-based machine analysis of social media content will detect a greater number of key opportunities than individuals with no machine assistance.*

**Accuracy of detected critical problems and opportunities**  
(sentiment analysis-based)

***P2a:** Individuals assisted by sentiment-based machine analysis of social media content will more accurately detect key problems than individuals with no machine assistance.*

***P2b:** Individuals assisted by sentiment-based machine analysis of social media content will more accurately detect key opportunities than individuals with no machine assistance.*

### **Confidence in detection of critical problems and opportunities** (sentiment analysis-based)

***P3a:*** Individuals assisted by sentiment-based machine analysis of social media content will have greater confidence that they detected key problems than individuals with no machine assistance.

***P3b:*** Individuals assisted by sentiment-based machine analysis of social media content will have greater confidence that they detected key opportunities than individuals with no machine assistance.

In light of the shortcomings of sentiment analysis encountered in the literature review section, we expect it to provide a limited advantage to extracting knowledge from highly unstructured text, although we doubt its capacity on a large scale to match the decision-making support capabilities of an advanced NLP-based system. We reiterate the potential loss of meaning attributable to evaluation at the word level, in contrast to the retention possible with the contextualized evaluation of entire comments or clauses. The latter approach is far more robust to irony, sarcasm, misspellings, omitted words, and idiomatic expression, all pervasive characteristics of the type of highly unstructured text that comprises social media-enabled textual communication. Our design also recommends analysis driven by machine learning from a statistical model derived from an enormous set of pairs of crowdsourced interpretations of unstructured customer comments, Tweets, etc., One important advantage of this methodology is that it eliminates constraint to a simplistic bifurcated classification such as that required by a crude positive-or-negative assessment; with NLP, important consumer opinion not readily captured by an on/off scale can still be detected and retained in support of the goal of increased organizational decision-making capacity.

Based on the information-detection advantaged inherent in an NLP-based automated approach to the analysis of highly unstructured text, we propose the following relationships between NLP-based and sentiment-analysis or manual analysis of these data:

**Number of detected critical problems and opportunities**  
(NLP-based)

***P4a:** Individuals assisted by natural language processing-based machine analysis of social media content will detect a greater number of key problems than individuals assisted by sentiment-based machine analysis or with no machine assistance.*

***P4b:** Individuals assisted by natural language processing-based machine analysis of social media content will detect a greater number of key opportunities than individuals assisted by sentiment-based machine analysis or with no machine assistance.*

**Accuracy of detected critical problems and opportunities**  
(NLP-based)

***P5a:** Individuals assisted by natural language processing-based machine analysis of social media content will more accurately detect key problems than individuals assisted by sentiment-based machine analysis or with no machine assistance.*

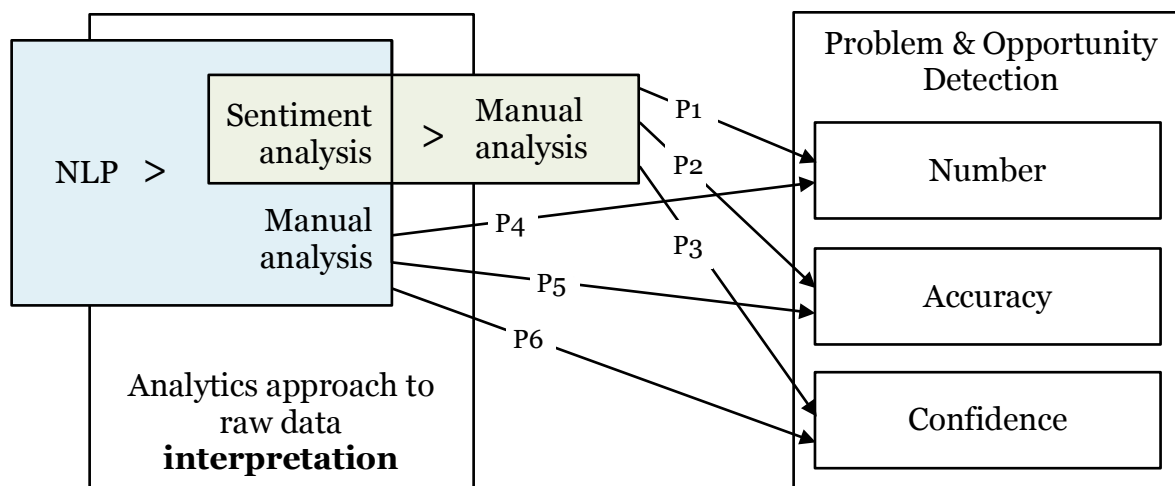
***P5b:** Individuals assisted by natural language processing-based machine analysis of social media content will more accurately detect key opportunities than individuals assisted by sentiment-based machine analysis or with no machine assistance.*

## Confidence in detection of critical problems and opportunities (NLP-based)

**P6a:** *Individuals assisted by natural language processing-based machine analysis of social media content will have greater confidence that they detected key problems than individuals assisted by sentiment-based machine analysis or with no machine assistance.*

**P6b:** *Individuals assisted by natural language processing-based machine analysis of social media content will have greater confidence that they detected key opportunities than individuals assisted by sentiment-based machine analysis or with no machine assistance.*

These six two-part propositions are graphically represented in the following figure. Further discussion of the predictive validity of our concepts and variables is detailed in Appendix A. Operationalizations of propositions are described in Table 4-4.



**Figure 4-2**  
**Research model highlighting propositions**



**Table 4-4**  
**Operationalization of propositions and concepts in experiment**

|   |   |
|---|---|
| Analytics approach to raw data interpretation | Treatment variable; expressed as contents of decision-assistance panel, 20 Tweets extracted from the raw data based on one of three methodologies. See Appendix D for contents of each Tweet set comprising the three levels of treatment. NLP group receives 20 Tweets evaluated as most important by an NLP approach, SA group receives top 10 most negative and top 10 most positive Tweets as evaluated by a sentiment analysis, and manual approach receives a random set of 20 continuous Tweets. |
| Number  | Self-reported counts of number of problems and number of opportunities identified.  |
| Accuracy                                      | Reflects external tally of problems and opportunities appearing both in subject's assessment and raw data. Appendix J describes this procedure in detail. Standardized counts range from 1 to 7.  |
| Confidence                                    | Operationalized as a 1 to 5 Likert scale. Subjects self-report this score.  |

#### **4.4 Research Methodology**

The propositions proposed in Section 4.3 and depicted in Figure 4-1 were tested through a single-factor, controlled laboratory experiment. The experiment required subjects to identify product- and firm-related opportunities and problems from a stream of customer-generated social media content. The independent variable was manipulated by giving participants different “decision-assistance panels” to consult during the task of identifying information important to organizational decision-making. The contents of this panel varied according to the three approaches being tested (manual, sentiment analysis, NLP analysis). All participants had the same raw data, the only difference

across treatments being the decision-assistance panel (whose contents were derived from the raw data).

#### **4.4.1 General task environment**

The object of this experiment was to determine the most effective analytics approach to deriving useful meaning from consumer-generated social media data, with the ultimate goal of supporting organizational decision-making. Given the relevance of the organizational setting for this particular study, the task assigned to subjects approximated a typical organizational responsibility: scanning the environment, in this case a social media environment, to discover critical information (i.e., problems and opportunities) to relay to senior decision makers for interpretation (Daft and Weick 1984).

#### **4.4.2 Experimental design**

The experimental design included three *between-subjects* conditions constituted by the *type of analytic approach* used to generate the decision-assistance panel: (1) *no analytical preprocessing*, (2) *sentiment-analysis based preprocessing*, and (3) *advanced NLP-based preprocessing*. These conditions correspond with approaches either currently employed in practice, or in the developmental stage: (1) manual text monitoring, (2) a standard analytic approach typically used in social media brand monitoring, and (3) a potentially useful innovation for firm-level social media monitoring, all of which are discussed in previous sections.

All subjects received three components via a browser-based interface designed for this experiment (see Appendix B for screen a shot of the experimental interface): (1) raw data (see Appendix C for an excerpt) on the left half of the browser window,

identical across groups; (2) a decision-assistance panel varying in content across conditions (see Appendix D for contents corresponding to each treatment) on the upper right quadrant, and (3) the debriefing questionnaire on the lower right quadrant (see Appendix K for content of the debriefing questionnaire). The experiment was implemented on a standard monitor in the same web browser to control for possible differential effects of look and feel. Font size, color, scrollability, etc., were hard-coded to ensure total uniformity across displays.

All participants (regardless of condition) received the same set of raw data, which was presented in identical order to ensure consistency of encountered problems and opportunities. We chose Twitter-mediated social media messages pertaining to the SunglassHut brand for the experiment. Tweets are appropriate because they are restricted to 140 characters, thus enabling us to control for maximum message length (and by extension, density of information conveyed in a single message). SunglassHut is an ideal brand for the experiment because it is a real business with a strong social media presence (resulting in abundant real raw data), and it sells products very accessible and familiar to our participant base of college students (fashionable and trendy sunglasses). We confirmed through an initial round of pilot testing that a stream of 300 Tweets provided sufficient information overload to prevent subjects from easily processing all messages manually and compelled subjects to rely on the decision-assistance panel provided to support task execution. This is a critical design feature since the ultimate goal of the experiment is to test the efficacy of *automated* processing. Search functionality was also provided to simulate keyword searchability.

Across conditions, we manipulated the decision-assistance panel, in each case a subset of 20 Tweets extracted from the raw data according to the three analytics

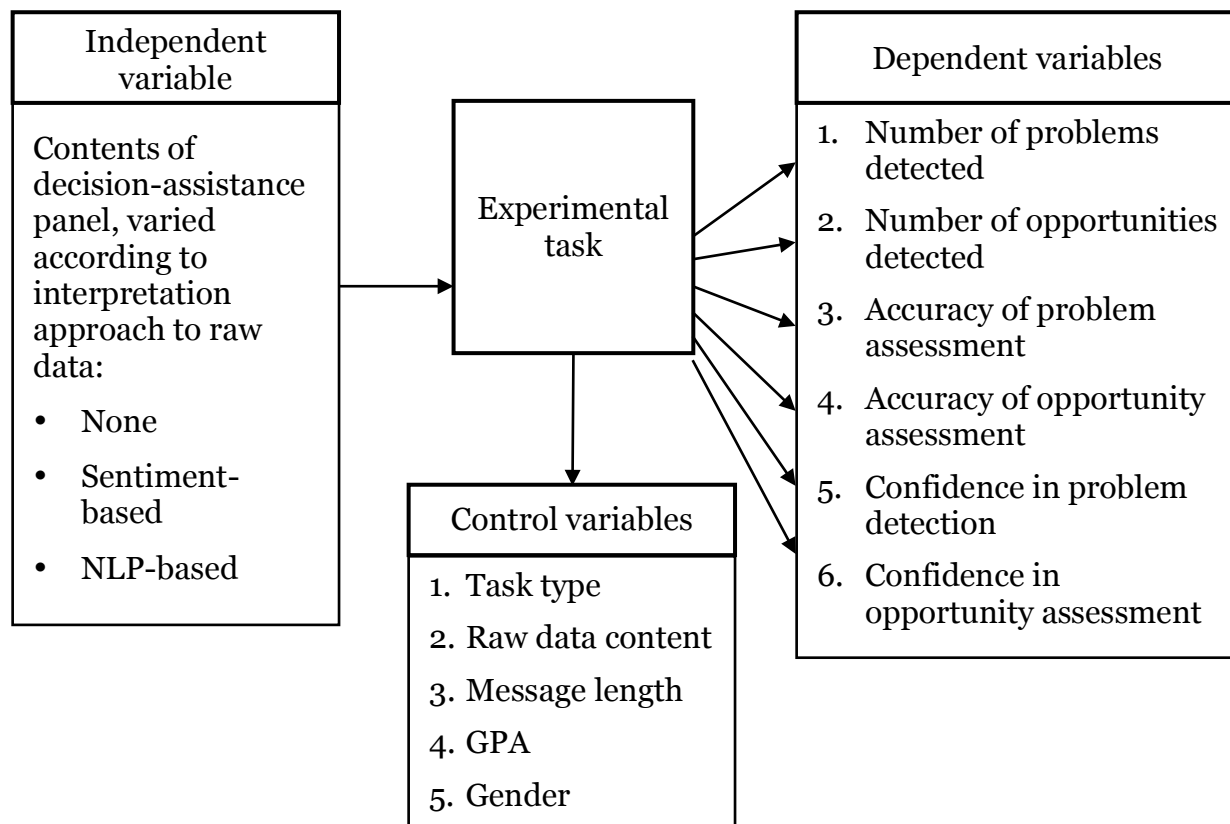
approaches being compared (none, sentiment-based, and NLP-based). The first condition received a panel of twenty continuous Tweets selected from the raw data. As opposed to providing twenty Tweets at random, we opted to provide random (thus different for each member of the group) *sets of continuous* Tweets in order to increase external validity. It is unlikely that Tweets would be read completely at random (i.e., out of order) in a business setting; rather, it is more likely a social media analyst or manager lacking the assistance of an analytics system would, at a minimum, read consumer-generated Tweets in real-time order via the Twitter console, or conduct keyword-specific searches via the basic search functionality provided by Twitter, which would retain chronology. Because understanding a given message may depend on reading it in series and embedded within a set of interactions, we considered that complete randomization of Tweets might unfairly bias the control group's understanding by obscuring context that would be clear in a natural setting. For these reasons, we preserved whatever intelligence might be communicated through message interconnectedness by randomly extracting unbroken series of Tweets.

The second condition received the first of two decision-assistance panels compiled according to automated approaches: the panel for Group Two consisted of twenty Tweets selected for inclusion based on sentiment analysis. The top ten most negatively and most positively scored Tweets from the raw data stream were included in this group's panel, which was meant to simulate the popular mechanism underlying a large percentage of currently-available free and fee-based social media monitoring systems. (The contents of the second panel are presented in Appendix D, while Appendix E presents our selection methodology in detail).

Finally, the third condition received a panel of twenty Tweets selected for inclusion based on a simulation of natural language processing. Simulation was a necessary step in our quest to determine the effectiveness of this nascent and as-yet incomplete technological advancement. Appendix D presents the contents chosen for this panel. We explicitly describe the simulation algorithm used to select this content in Section 4.4.2.2, “Independent variables.”

#### 4.4.2.1 Variables and variable relationships

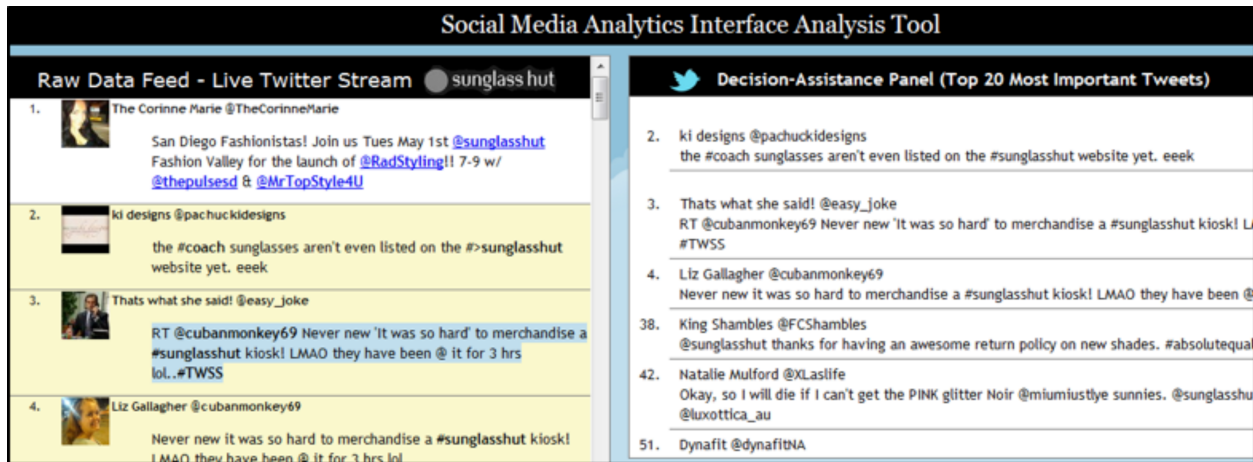
Figure 4-3 graphically depicts the relationships among independent, dependent, and control variables in the experiment:



**Figure 4-3**  
**Variable relationships**

#### 4.4.2.2 Independent variable

The single manipulated factor is the social media analytics approach used in an organizational setting, which we operationalize and manipulate by providing decision-assistance panel contents derived by different analytic approaches: none, sentiment-based, and NLP-based. Each panel comprises twenty Tweets for a balanced design.



**Figure 4-4**  
**Decision-assistance panel on right, raw data feed on left**

Part of the experimental interface is depicted in the preceding figure. A portion of the decision-assistance panel for the NLP treatment is shown on the right. The twenty component Tweets are individually highlighted within the raw data stream in yellow to enable the subject to easily detect where in the stream a given item appears, thus ensuring context is available in case Tweets build upon or respond to one another. Further, the subject can click on a Tweet in the decision-assistance panel and the raw data stream will scroll up or down to the active Tweet, additionally highlighting it in blue so that the subject can easily find it amidst a group of yellow highlighting.

## Manual Group

In order to prevent biasing the manual group by inadvertently featuring a random set of twenty consecutive Tweets that features particularly useful information or, alternately, particularly useless information, each member of the manual group receives a different set of twenty Tweets, the set chosen at random.

## NLP Group

The NLP group receives a human-executed output of a procedural algorithm based on a set of inclusion/exclusion steps derived from the design principles proposed in Section 4.2.1 intended to simulate machine output. As this human-mediated simulated output is neither fully representative of what a human using his or her own judgment might select as the most important twenty Tweets due to the constraints imposed by the methodology, nor a purely automated machine execution of a set of rules (as this technological capability is not yet reliably or robustly available) due to the unavoidable dimension of humanness that cannot be separated from the process, we suggest that the output represents the effects of a “best NLP.” This superlative output benchmarks what we hope to eventually achieve with a natural language processing-based analytics system.

Because we are testing the *theoretical* design of a class of advanced NLP-based social media analytics systems to determine, in part, the usefulness of pursuing such an enterprise, the decision-assistance panel contents for condition 3 are extracted from the raw data according to an algorithm specified to *simulate* an instantiation of the design. We outlined a procedure intended to produce output consistent with an automated analysis based on the design principles discussed in section 4.2.1, and tasked three independent raters with executing the algorithm manually to pare 300 Tweets to the

unequivocal top twenty most important customer-to-customer and customer-to-firm Tweets for Sunglass Hut management to learn from.

While the first three design principles (context-sensitivity, machine learning, and sociotechnical principles) are embodied in the design of the simulation algorithm, the final principle of actionability (or the requirement that a system must produce results in a timely output suitable for decision making such that organizations can make well-timed assessments and take rapid action) cannot be simulated, but must be assumed, due to the lack of time as a variable in this particular experiment. To approximate this assumption, we told experiment participants that the stream of raw data (the set of 300 Tweets) represents a live stream of real-time Twitter chatter.

Figure 4-5 graphically presents the steps that would be programmed into a machine algorithm intended to sift through unstructured social media texts for derived understanding; these steps are further translated for human action into the instructions given to the human sorters (see Appendix F). The procedure is explicated in the subsequent section.

### **NLP Output Selection Procedure**

Starting with a deck of 300 Tweets (our raw data scraped directly from Twitter), we first conducted steps to reduce the noise-to-signal ratio. Because noise filtering and signal specification in this experiment (in which we are strictly interested in customer-to-customer and customer-to-firm communications) dictates the straightforward elimination of advertisements, contest links, spam, coupon codes, non-English messages, employee- and firm-originated messages, and photo- or url-only messages, we completed this step for all messages unambiguously falling into one of these



categories prior to the raters' simulations. These "noise" Tweets did not require a judgment call to eliminate and would be excluded immediately by any accurate rater or simple automated filter. This noise-reduction step removed over half the Tweets from the raw data stream, leaving 140 relevant Tweets for the raters to siphon through the simulation process.

Following the sorting instructions presented in Appendix F, three independent sorters each arrived at a final set of 30 messages that, according to their interpretations, represented the theoretical output of an NLP-based machine algorithm, and *not* human expectation of what a machine *should* choose as most important. Thus, this set approximates a type of "best NLP" output, or output that approaches human capability, which is what NLP scholars strive toward. The algorithm specified, it should be noted, reflects only a portion of NLP capabilities that comprise myriad sub-fields of research, including Named Entity Recognition and Relation Extraction; thus, this algorithm can potentially be made even more "machine-like" by incorporating additional simulation steps. The sorting algorithm specified for this experiment is explicated following this section.

Of the 140 Tweets remaining in the deck after noise-elimination, subsequent to human processing of the NLP simulation, 46 Tweets occurred in at least one of the three final sets created by human sorting (33%), while 14 appeared in two lists and 15 Tweets occurred in all lists. To test the hypothesis that the lists overlap to such a degree merely due to chance, we consult the hypergeometric distribution, the distribution that overlapping probability is well known to follow (Fury et al. 2006). Based on the hypergeometric distribution of  $N = 140$  (see Appendix G for a detailed explanation of the statistic), it is statistically significant that two lists of 30 will overlap by 15 Tweets,

with a p-value < 0.0001; thus we conclude that a 50% concurrence is very significant and proceed on the assumption that the simulation outputs are due to the simulation algorithm and not chance, particularly for the 15 Tweets occurring in all three outputs.

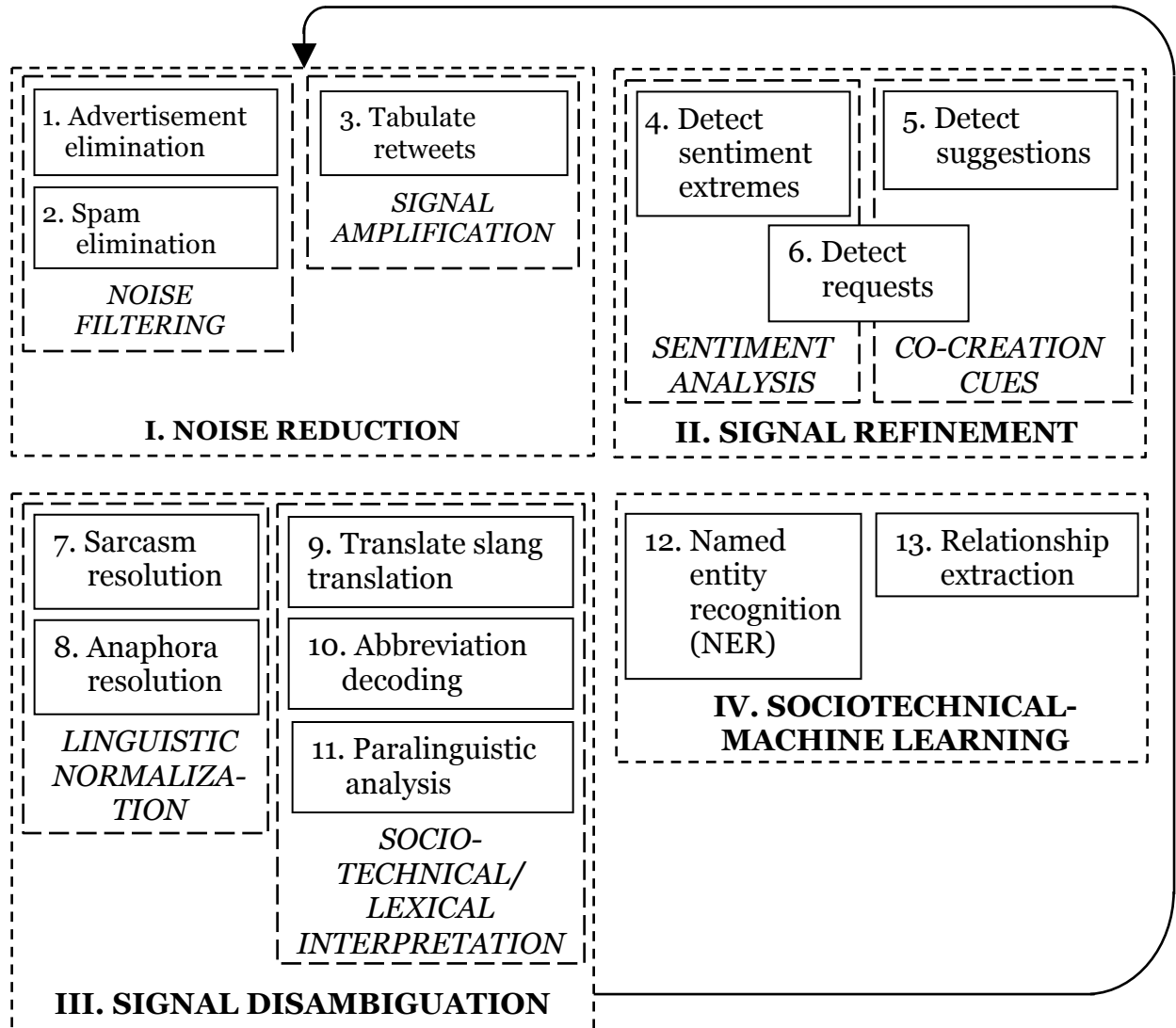
The degree of Tweet concurrence across outputs is summarized below in Table 4-5. We added the fifteen Tweets for which there was 100% consensus to the NLP-based decision-assistance panel without further analysis. To determine the appropriate final five Tweets, we presented to four completely new raters the fourteen Tweets appearing in two of the first three raters' outputs. We instructed the new raters to choose the top eight most important Tweets from this subset based on the criterion of likely importance to Sunglass Hut management. Of the 14 candidates, three were selected by all four independent raters, while two Tweets were chosen by three of the four; these five completed the decision-assistance panel for the NLP treatment. Thus, the NLP top twenty was rigorously designated through simulation of the algorithm with subsequent tie breaking by a total of seven independent raters.

**Table 4-5**  
**Overlap of NLP-simulation outputs**

|                  |    |
|------------------|----|
| Triple agreement | 15 |
| Double agreement | 14 |
| No agreement     | 17 |
| Tweets selected  | 46 |

### **NLP Simulation Algorithm**

The simulation algorithm for approximating an automated NLP-based outcome is depicted in Figure 4-5 and explained as follows.



**Figure 4-5**  
**Graphical representation of algorithm for NLP-based simulated output**

### **I. Reduction of Noise-To-Signal Ratio**

The algorithm for approximating the theoretical output of an NLP-based text analytics system via human rater begins with measures to reduce overall noise within the raw data stream. The initial step of (1) **eliminating advertisement** messages originating with Sunglass Hut or an affiliate removes text that does not represent

*unknown* or *undiscovered* intelligence, thus reducing superfluous data from the analysis. (2) **Removing spam** messages serves a similar noise-reduction purpose. Complementarily, a common Twitter custom of “retweeting” (denoted by the initials RT in the microblog entry) serves as an amplification mechanism through which a given status is rebroadcast through the system, essentially increasing signal strength. A particularly salient or useful message may be repeated numerous times, a signal indicating probable elevated importance of the original message. As such, (3) **tabulating retweets** facilitates the isolation of potentially significant focal messages. This step clearly defines the necessity of iteration in our human simulation: during the overall evaluation process, it is only once the “RT” signifier is encountered that the rater can know a previously-read message classifies as having been retweeted. Once a retweet is encountered, additional retweets of the same content augment the “importance assessment” of the focal message. Thus, manual evaluation of a set of Tweets requires two iterations for complete evaluation.

## II. Signal Refinement

Once irrelevant messages have been excluded and notable retweets earmarked, we begin to refine content selection by identifying important cues. Sentiment analyses evaluate content based on calculations of positive and negative word instances. These concepts are of high utility to an NLP-based analysis, so we incorporate (4) **detection of extreme sentiment** words while additionally detecting words indicative of (5) **suggestions** (“This product needs...”) and (6) **requests** (“I need help with my...”) that may serve as cues to further classes of knowledge and intelligence inherent in brand-oriented social media settings. Depending on their significance, suggestions and

requests for assistance or functionality embody two types of undiscovered knowledge that could add value and potentially alter organizational decisions

### III. Signal Disambiguation

Resolving ambiguity is the difficult and important goal of many NLP tasks such as parts-of-speech tagging (Hutchins 2006). Many of these “AI-complete” objectives depend on a range of types of human knowledge—grammatical rules, semantics, and facts about the real world—for successful resolution (Mairesse et al. 2007). This level of “contextual sensitivity” is extremely difficult to automate, in part because we lack accurately-labeled corpora for training machine learning systems (González-ibáñez and Wacholder 2011). An idiosyncratic feature of Twitter, the use of hashtagged keywords by comment authors to increase search accuracy, has facilitated sarcasm corpus building via the inclusion of the hashtag #sarcasm to eliminate ambivalence of intent. Sarcasm often reverses the polarity of a comment’s apparent sentiment, so if organizational decisions are to be influenced by information extracted from social media messages, that intelligence must control for sarcasm in order to be reliable. Thus, (7) **sarcasm resolution** is a critical step toward accurate simulation output. Further, once sarcasm is established in a given comment, this necessitates a reiteration of sentiment analysis to account for possible sentiment polarity reversal.

Additional ambiguity elimination takes place during (8) **anaphora resolution**, the determination of which previous noun a pronoun or other back-referring phrase corresponds with (Kao and Poteet 2007:1). This NLP subfield is rich with algorithms for pronominal anaphora resolution that enjoy high rates of correct analyses (e.g., Kennedy

and Boguraev 1996; Lappin and Leass 1994). Human simulation of this process will likely benefit from contextualized interpretation.

Additional interpretation occurs through the “translation” of (9) **slang**, (10) **abbreviations**, and (11) **paralinguistics**, or symbolic conventions used as shortcuts for standard concepts, phrases, or words. Ubiquitous yet cryptic, texting short cuts (“FWIW” means “for what it’s worth”) and emoticons (:p is the text equivalent of sticking one’s tongue out at someone else) can be captured appropriately through an interpretation corpus, and can significantly change the initial extraction of information. Once these symbols are expanded into equivalent words and phrases, the simulation process should be reiterated in the case that the expansion alters previous resolutions of sarcasm, etc. that need to be factored into the interpretation process.

#### **IV. Socio-technical Calibration**

Finally, message intent should be iteratively refined by additional clarification achieved through (12) **named entity extraction** (NER) and (13) **relationship extraction**. It is possible that additional knowledge or alteration to meaning previously assumed may be changed by the identification of brands, businesses, particular people, etc. NER is an NLP subtask whose goal is to identify within text all the names for specific types of things, typically persons, organizations, and locations (Sang and De Meulder 2003). Relationship extraction is a technique used to disambiguate relationships between objects and people. Upon resolution of these substeps, it is necessary to execute an additional round of refinement due to the alteration or clarification of meaning these procedures may convey.

#### **4.4.2.3 Dependent variables**

We are ultimately interested in how different social media analytics approaches support organizational decision-making. As proxy for this downstream construct, we measure the ability of participants to identify important problems and opportunities for the firm. We justify this operationalization by referring to the general assumption that organizational decision-making depends on information (Delbecq and Ven 1971; Galbraith 1974; Huber and McDaniel 1986) and that external problem and opportunity assessments are classic concerns of strategic planning (Houben, Lenie, and Vanhoof 1999). Opportunities convey chances to improve performance while problems are elements that could cause trouble for the business and therefore concern organizational managers.

Six dependent variables are measured. These variables gauge the number of problems and opportunities identified by the participant, the accuracy of the participant's problem and opportunity assessments, and the participant's confidence that he or she thoroughly detected the important problems and opportunities for the firm. While the same measures are repeated for problems and opportunities, questions regarding problems are posed to participants individually from questions regarding opportunities so as to avoid conflation of the two concepts and enable us to obtain more granular responses that measure distinct concepts properly.

#### **Number of problems and opportunities identified**

In the course of the experiment, subjects are tasked with tallying the number of problems and number of opportunities for the firm that they detect from customer-to-

firm or customer-to-customer Tweets (e.g., the raw data stream). We use these self-reported values as direct measures for these two dependent variables.

### **Accuracy of problem and opportunity assessment**

In addition to indicating the number of problems and opportunities they identify, participants are asked to describe in their own words (i.e., not cut and paste) the problems and opportunities they detect in the data and from which SunglassHut management might derive valuable intelligence. These written responses provide the basis for quantitative tallies of accuracy (see Appendix H for details on how this measure was derived and tallied). Accuracy gauges the degree to which a participant's assessment aligns with the set of predetermined most important problems and opportunities within the raw data stream.

### **Confidence in problem and opportunity assessments**

Finally, participants are asked to rate on a five-point Likert scale their level of confidence that they have detected important problems and opportunities from the raw data stream. This single-item scale provides an additional reflection of subjective decision quality (Häubl and Trifts 2000). Referring to a variety of recent studies demonstrating little to no difference in the predictive validity of multiple-item versus single-item versions of the same measure for such constructs as happiness (Abdelkhalek 2006), attitude toward advertisement, attitude toward brand (Bergkvist and Rossiter 2007), and global self-esteem (Robins, Hendin, and Trzesniewski 2001), we operationalize confidence as a single-item scale. Relevant multi-item scales found in the literature invariably incorporate components of self-efficacy, which is different from what we are trying to measure in this case. We want to capture only how assured a



subject is that she identified all critical issues in a particular case, not how assured the participant is in her general ability to detect problems. The range (see Table 4-6 below) follows standard 5-point Likert wording and arguably captures differences that are distinct enough that participants can identify with one level of confidence, exclusively.

**Table 4-6**  
**Confidence in problem detection and**  
**opportunity detection scale wording**

- 1: *“Not at all confident”*
- 2: *“Slightly confident”*
- 3: *“Somewhat confident”*
- 4: *“Very confident”*
- 5: *“Extremely confident”*

#### **4.4.2.4 Control variables**

Task type and raw data content are held constant by giving all participants the same task, objectives, instructions, and raw data from which to work. The interface is constant across groups, only differing in content (but not length) of the decision-assistance panel. In order to control for the possibility of variably dense messages, we opted to create the experiment with Tweets since they are limited to 140 characters. Finally, we measure GPA and gender to test for possible differences in responses due to these influences.

## 4.5 Data Analysis

### 4.5.1 Background information on subjects

A total of 85 undergraduate management information systems majors were recruited as subjects from a southeastern U.S. university campus and randomly assigned to one of three decision aid conditions, with 29 students in group 1, 29 in group 2, and 27 in group 3. Of the 85 participants, all between ages 18 – 22, 30 reported as female (approximately 34%) and 3 did not report gender. MIS majors are considered appropriate participants in this social media-oriented study because they are likely candidates to intern in organizations that have implemented or are interested in implementing some type of social media analytics system. It is reasonable to expect a student intern to be assigned to manage or monitor this type of technology and report key information to managers for further analysis or decision-making. The students appeared to be engaged in the assignment and generally interested in the topic of research. Participants volunteered afterwards that the task was “fun.”

### 4.5.2 Experimental results

A one-way MANOVA reveals a significant multivariate main effect for analytics approach, Wilkes'  $\lambda = 0.5236$ ,  $p = 0.0001$ . Given the significance of the overall test, the univariate main effects are examined. Significant univariate main effects for analytics approach are indicated for **number** of problems identified,  $p = 0.00095$ , **accuracy** of problem assessment,  $p < 0.0001$ , **confidence** in problem identification,  $p = 0.0491$ , and **accuracy** of opportunity assessment,  $p = 0.0081$ . Due to unequal variances across treatment group responses for accuracy of opportunity assessment, we use the Kruskal-Wallis nonparametric test on this variable (Neter, Wasserman, and Kutner 1990: 642).

Significant treatment pairwise differences are obtained in a linear contrast of **number** of problems identified between NLP and SA, and NLP and random. The mean number of problems identified is 3.71 using sentiment analysis, 5.86 using NLP, and 4.69 with a random set of Tweets. A similar pattern of pairwise differences is obtained for **accuracy** of problems assessment between NLP and SA and NLP and random. The mean accuracy rating of problem assessments is 2.41 using sentiment analysis, 4.38 using NLP, and 2.88 with a random set of Tweets. Finally for problem variables, significant differences were obtained in **confidence** in problem assessment between NLP and SA. The mean number of confidence levels indicated by participants is 3.1 using sentiment analysis, 3.66 using NLP, and 3.19 with a random set of Tweets.

Significant analytics approach pairwise differences were obtained for **accuracy** of opportunity assessment between NLP and SA and NLP and random. The mean accuracy rating of opportunity assessments is 2.1 using sentiment analysis, 3.17 using NLP, and 2.42 with a random set of Tweets.

ANOVA statistics for all dependent variables are presented in Table 4-7, including means of all measures.

**Table 4-7**  
**Means and ANOVA results for dependent variables**

|   | <b>Sentiment Analysis</b> | <b>NLP</b>              | <b>Manual</b>     | <b>Grand</b> | <b>Pr&gt;F</b> |
|---|---------------------------|-------------------------|-------------------|--------------|----------------|
| <b>Number – problems detected</b>       | 3.71 <sup>a</sup>         | <b>5.86<sup>b</sup></b> | 4.69 <sup>a</sup> | 4.77         | 0.000495*      |
| <b>Accuracy – problem detection</b>     | 2.41 <sup>a</sup>         | <b>4.38<sup>b</sup></b> | 2.88 <sup>a</sup> | 3.24         | <0.0001*       |
| <b>Confidence – problem assessment</b>  | 3.1 <sup>a</sup>          | <b>3.66<sup>b</sup></b> | 3.19 <sup>a</sup> | 3.23         | 0.0491*        |
| Number – opportunities detected         | 6.14                      | 6.24                    | 7.72              | 6.66         | 0.123          |
| <b>Accuracy – opportunity detection</b> | 2.1 <sup>a</sup>          | <b>3.17<sup>b</sup></b> | 2.42 <sup>a</sup> | 2.57         | 0.008076*      |
| Confidence – opportunity assessment     | 3.59                      | 3.55                    | 3.65              | 3.60         | 0.888          |

\* The highest mean for each variable is bolded and highlighted. Different superscripts indicate significantly different means for the four variables (bolded in the first column) with significant ANOVA F-tests.

### **Effect on Problem Number**

The ANOVA model indicates that the analytics approach to social media text introduces significant variation to the number of problems subjects are able to identify. To determine which groups differed, we performed planned linear contrasts (see Table 4-23, Appendix L). These results indicate that the NLP analytics approach enables participants to identify a greater number of problems underlying the raw data than either sentiment analysis or manual approaches, but no difference in numbers identified

is detected when comparing sentiment analysis to a manual approach. Therefore, **P4a** is supported while **P1a** is not.

### **Effect on Problem Accuracy**

The ANOVA model indicates that the analytics approach to social media text introduces significant variation to the ability of subjects to accurately identify problems underlying the data. To determine which groups differed, we performed planned linear contrasts (see Table 4-26, Appendix L). Results indicate that the NLP analytics approach enables participants to make more accurate assessments of problems underlying the raw data than sentiment analysis or manual approaches, but no difference is detected when comparing sentiment analysis to a manual approach. Therefore, **P5a** is supported while **P2a** is not.

### **Effect on Problem Confidence**

The ANOVA model indicates that the analytics approach introduces significant variation among participants' confidence in their ability to identify critical problems. Linear contrasts (Table 4-29, Appendix L) support a difference between NLP and sentiment-based approaches, although a difference is not supported between NLP and manual approaches. Thus **P6a** is supported by the data while **P3a** is not supported.

### **Effect on Opportunity Assessment Accuracy**

Finally, the ANOVA model indicates that the analytics approach to decision assistance affects the ability of participants to accurately identify and assess opportunities. The linear contrast (Table 4-33, Appendix L) indicates that the NLP approach enables a statistically greater degree of accuracy in identifying opportunities compared to sentiment analysis and manual approaches. A significant difference is not

detected between sentiment analysis and manual approaches. **P5b** is supported while **P2a** is not.

### **Effect on Other Opportunity Variables**

In contrast with the significant effect of the treatment on all problem detection measures, opportunity detection is not demonstrably affected along two of the three dimensions, number identified and confidence in assessment; these p-values are 0.123 and 0.888 respectively while power for these tests are 0.48 and 0.51. There are a few potential explanations for this disparity, the first being that it is possible that “opportunity” is a fuzzier concept for students to grasp at a firm level, while “problems” are likely more straightforward to recognize. It may lead to confusion that problems in general can be reconstructed as opportunities; for example, the problem of “lack of promotional pricing leading customers to defect to other brands” could be reformulated as an opportunity to provide more promotions in order to retain customer loyalty. The converse is not true, however—an existing opportunity cannot be reformulated as an existing problem.

Examination of the opportunity assessments supports this notion, given that multiple reports include reworded problems, for example, “better customer service could increase customer base,” “should list the products on website ASAP,” “educate the retailers on how to merchandise the product,” and “products are relatively more expensive. They could offer coupons to customers” all really reflect more of a focus on problems appearing in the raw data than opportunities, as voiced by the customers. These data still convey important information, but strictly speaking do not fall within the categorization of opportunity. It is important to note that while participants did not

report a greater number of opportunities or feel more confident in their ability to identify opportunities across treatments, the NLP group still outperformed the other groups in terms of accuracy of opportunity assessment. While it would be optimal to accurately identify a greater number of opportunities, it can still be helpful to organizational decision-making to accurately identify a small number of opportunities. The converse is not true, however. It would be detrimental to identify a greater number of opportunities *inaccurately*; thus, we argue accuracy is a more important variable, comparatively, to perform well on, all else being equal.

**Table 4-8**  
**Summary of hypothesis testing**

| <b>Hypothesis</b>   | <b>Supported?</b> |
|---|-------------------|
| <b>1a:</b> SA > manual (number of problems detected)                  | No                |
| <b>1b:</b> SA > manual (number of opportunities detected)             | No                |
| <b>2a:</b> SA > manual (accuracy of problem detection)                | No                |
| <b>2b:</b> SA > manual (accuracy of opportunity detection)            | No                |
| <b>3a:</b> SA > manual (confidence in problem assessment)             | No                |
| <b>3b:</b> SA > manual (confidence in opportunity assessment)         | No                |
| <b>4a:</b> NLP > SA and manual (number of problems detected)          | Yes               |
| <b>4b:</b> NLP > SA and manual (number of opportunities detected)     | No                |
| <b>5a:</b> NLP > SA and manual (accuracy of problem detection)        | Yes               |
| <b>5b:</b> NLP > SA and manual (accuracy of opportunity detection)    | Yes               |
| <b>6a:</b> NLP > SA and manual (confidence in problem assessment)     | Yes               |
| <b>6b:</b> NLP > SA and manual (confidence in opportunity assessment) | No                |

## 4.6 Implications of Results

The most interesting result of the analysis is lack of support for our general conjecture that while only a very crude heuristic, sentiment analysis should still provide some measureable advantage beyond manually sifting through raw data. The results indicate that sentiment analysis provides no advantage over simply reading random sets of consumer chatter; that is, both perform poorly compared to the NLP analysis. At the very least, these results should warn organizations to be cautious when attempting to link sentiment to actual business outcomes.

There appears to be a substantial difference in problem detection versus opportunity detection that warrants further investigation. While we see that the analytics approach significantly impacts the participants' ability to detect problems, the same pattern does not hold across the board for opportunity detection. Post hoc examination of the raw social media stream indicates that there are approximately three times as many important opportunities embedded in the data as important problems, using the theoretical output of the NLP simulation as our basis for evaluation. It is possible that the relative abundance of opportunities leveled the playing field for participants, making it possible to detect enough opportunities manually from the raw data that the participants did not need to rely on the decision-assistance panel for help. This of course obviates our manipulation, as we are not interested in what subjects can detect from the entirety of raw data manually, but in what they can detect from the raw data while relying on the decision-assistance panel of Tweets drawn according to various methodologies.

It may be possible to overcome this issue by doubling the number of Tweets appearing in the raw data feed to 600, to guarantee that the subjects are undoubtedly



overwhelmed and ensure reliance on the decision-assistance panel. If a usable NLP system were available to us, we would no longer need to rely on human simulation of the design principles to produce contextual, machine-learning-based, socio-technical, real time output. Given the results of this study, effort to produce such a system seems reasonable; if this technology were available to use and the simulation step were no longer simulated but machine-compiled, a far greater number of Tweets could be incorporated into the interface because we would no longer be constrained by needing to make the task reasonably compact to prevent overwhelming the human raters processing the raw data during simulation.

It is still worthwhile to note that although participants do not identify greater numbers of opportunities or feel more confident about their opportunity identification using one analytics approach over another, accuracy of opportunity identification is still superior with the use of an NLP-based decision-assistance panel. It could be the case that number and confidence do not matter as much to ultimate organizational decision-making as accuracy; while the weighted importance of these factors is outside the scope of this research, this question nonetheless provides an interesting issue to address in future research.

It is also worthwhile to point out that in every individual case, although not statistically significant in difference, manual approach enables a slightly better performance than sentiment analysis. While this may not be of much import taken singly, if we consider the set of results as a whole we can consider them in light of the binomial distribution probability. Thus while the results for any given test may have no better than a 50% chance of coming out in favor of manual analysis, the probability that all six tests would favor manual analysis is 0.0156, indicating significance in aggregation

that sentiment analysis is not only no better than reading Tweets manually, but is likely detrimental to the analyst's potential ability to glean information important to organizational decision-making.

#### **4.7 Discussion**

This study examines the effects of various analytics approaches to mining intelligence from highly unstructured text, specifically the type of unconstrained consumer-generated text that generally materializes within social media settings. We have selected and compared three types of approaches—two that are typically employed in organizational settings today (i.e., sentiment analysis and manual analysis), as well as an innovative, natural language approach designed to resolve the shortcomings inherent in the two currently standard approaches. While there are few studies in Information Systems research that take advantage of knowledge accumulated in the field of qualitative text data mining, the advent of the social media age commands attention to such technologies, particularly as they are likely to convey advantages to organizations willing to embrace such novel approaches to better understanding their environment. Tapping into the wealth of knowledge in customer-to-customer social-media-enabled exchanges via advanced contextual qualitative mining approaches is forward looking and sophisticated, particularly compared to existing methods. We see exciting opportunities at this intersection of computer science, linguistics, organizational science, and information systems.

Natural language processing and machine learning are, similarly, overlapping fields experiencing steady advancement. Again, few studies in IS turn to computational linguistics to contemplate how we might plan to leverage these burgeoning technologies

in organizational settings, but as these practices become more sophisticated it becomes imperative that information systems scholars understand the consequences (and benefits) of replacing current protocols with such advanced capabilities. For example, the short list of papers examining how NLP can increase the efficacy of certain organizational tasks includes a manuscript providing evidence of the superiority of NLP techniques for enhanced information retrieval (Arazy and Woo 2007), while a statistical machine learning based design of a new class of fake website detection systems proves more accurate than existing systems (Abbasi et al. 2010). These studies clearly herald positive implications for the areas of e-commerce, security, and information asset management, but there is a wide range of organizational practices that may be studied to contribute additional empirical knowledge to our understanding of natural language processing and the practices it enables. Our research contributes to this knowledge gap by applying the automated process of extracting meaningful information from natural language input to the context of social media measurement, an increasingly critical concern for organizations (Hoffman and Fodor 2010).

Additionally, we hope to contribute at a conceptual level to the IS research literature by offering this paper as an example of research situated in circumstances whereby the researcher has neither guiding theory to shape the study, nor adequately effective extant practice upon which to improve. We present a logically justified experiment that empirically tests a design thought to be superior to current practices and that can eventually be instantiated, observed, and tested to generate new theory capable of improving subsequent practice. Despite the lack of a theoretical foundation for the results at hand, we demonstrate that modeling information systems design after a parallel but not distinctly related practice can still improve the focal practice.

In this study, the design of a better approach to detecting critical information from social media chatter is motivated by the presumption of a more useful objective than merely monitoring positive versus negative sentiment. While understanding customer sentiment is appropriate and relevant to a variety of research questions as well as consumer-oriented practices (such as providing peer reviews of products or movies), we suggest that the same capability can be subsumed much more accurately by a natural language processing-based mechanism, particularly as applied to highly unstructured text. An expanded capacity will enable social media analysts not just to detect sentimental extremes, but to discover a wide range of intelligence underlying customer comments, suggestions, requests for assistance, product-related issues, and other components that may prove important to decision-making (despite being neither extremely positive nor extremely negative).

The most important contribution this study makes to social media research is the demonstration that using sentiment analysis to learn from customers is likely no more effective than reading streams of consumer chatter by hand. This result is valuable for improving social media monitoring practices. Empirical proof that an NLP approach is superior to sentiment analysis suggests that efforts to build an information system based on the design principles motivating this paper constitute a worthwhile and beneficial goal, with the potential to substantially increase the amount of knowledge firms may glean from tapping into customer-to-customer exchanges and enhance the effectiveness with which they monitor and respond to customer-to-firm communications.

#### **4.7.1 Limitations of the study**

Social science researchers are subject to the universal dilemma of attempting to reconcile the following conflicting objectives (McGrath 1981):

1. Maximizing external validity by increasing the generalizability of the findings to population.
2. Maximizing internal validity by strictly controlling independent variables.
3. Maximizing realism by studying the phenomenon in a realistic setting.

While there is no way to simultaneously achieve these desiderata to their fullest, we strived in our research design to give strong consideration to each. Although we maximize internal validity at the expense of realism and generalizability (which is the appropriate stance for an experiment), certain design features serve to preserve these objectives to some degree.

We suggest the following justifications for selecting an experiment as our methodology:

##### **1. NLP technology is new.**

As NLP technology is still evolving, systems implementing this technology are new (even theoretical). Little is known about the possible effects of natural language processing in a social media setting, and we approached this study open to the possibility that an NLP approach might not convey a substantial advantage over sentiment analysis, thus proving futile any further efforts to make such a system. Therefore, our research question required precise manipulation of the treatment in order to isolate the effects of the approach on our response variables.

## **2. There are few, if any, examples of NLP-based information systems in use in organizations.**

Because reliable execution of NLP is still developing, this is not a technology readily in use for monitoring text flows in organizations yet. Thus, we do not have the ability to study this technology in the field, particularly in the context of social media measurement.

## **3. Statistical power is increased in a laboratory setting.**

As statistical power is a function of variation in the sample, a controlled setting with homogenous subjects increases the power of the study. Students are relatively homogenous and are available in numbers sufficient for adequate power.

Alternately, the shortcomings of a laboratory experiment include reduced external validity and realism. We discuss these as they pertain to this study:

### **1. Low external validity.**

A laboratory experiment with student subjects is not typical of organizational life because it lacks interpersonal involvement, employee accountability, and the ongoing culture of an enterprise. Students in general will unlikely display the motivation to perform at the same level as when preparing a report for their manager. However, to address the artificiality of the setting, we instructed the students to approach the task as if they were interns in an organization and were tasked with reporting back to their boss on the details gleaned from using the social media analytics system. Nevertheless, the results of this study must be cautiously applied to different settings and populations.

## **2. Low realism.**

If is difficult to convincingly simulate the real world in a laboratory setting. Students are not as highly motivated, committed, or involved as an actual intern in a real organization. While the students generally appeared to treat the task seriously, act responsibly (e.g., being punctual to their time slot), and even consider the task “fun” (as one participant mentioned after his submission), it is still likely that given the same task in business setting, they might demonstrate greater diligence.

## **3. Limited sample size.**

The sample of 85 participants is limited in size, resulting in an average cell size of 28. This potentially entails the limitations that accompany small sample sizes.

## **4. Reliance on simulation**

Because the proficiencies suggested by the design principles are not generally available in a usable format, building an actual NLP algorithm meeting all the criteria delineated was outside of the scope of this project. As such, we required simulation of the algorithm, executed by human sorters who are subject to differences in opinions, interpretations, external experiences, and other factors that would not result in variations by a machine. While we maintained rigorousness in our methodologies and gave full attention to ensuring consistency across raters to preclude bias, we cannot escape the fact that the simulation was ultimately subject to human predisposition. As NLP capabilities become realized to a greater degree, further research replacing human simulation of machine algorithms with actual machine algorithms is necessary to discover whether the findings of this study hold true to a truly automated NLP output. It is also important to have human analysis as a benchmark against which to compare

alternative NLP algorithms. Furthermore, any output based on a machine-learning method will necessarily only be as good as the model from which the algorithm operates; wide variance of system performance is possible depending on the amount of training data available, how many “translations” comprise the parallel corpora, and how appropriate the interpretations are. As we are able to begin to implement these types of systems in practice, we will be better able to develop theory regarding how and why they work better or worse than systems based on other approaches, which will in turn allow us to improve upon system design and implementation.

## **5. Task novelty.**

Students were asked to execute a task that they may have had little practice with in the past, especially considering that most have not yet held full-time employment in an organization. While identifying problems and opportunities is a classic task in a business setting, the terminology and undertaking of the problem may be somewhat foreign to undergraduates, even business majors. Subjects more familiar with SWOT analysis or practiced in formulating written business reports may perform differently and produce different outcomes.

As we noted earlier, the conception of “opportunity” might constitute a more nebulous idea to students than “problem,” and thus be more difficult to isolate and assess. This complication could be addressed by implementing a short training session at the start of each round of experiments in which a short feed of Tweets is projected for all to see and the proctor of the experiment walks the participants through an explanation of which Tweets contain information appropriate to include in a report to managers. Critically, the proctor would specify precisely what a problem looks like and



what an opportunity looks like. While we gave the students verbal direction of what might constitute a problem or opportunity (see Appendix J for complete instructions given), we merely suggested a few examples (e.g., that a problem could be something like a customer complaint while an opportunity could be something like a customer request that an intern could address or praise for some type of customer service policy is working well and should be continued). We intentionally limited the number of examples mentioned to subjects before the experiment out of concern that we would bias them into looking specifically for those particular types of information, thus reducing their sensitivity to other types of issues. Because we stressed that the examples mentioned represent just a couple of rhetorical examples that should not limit or bias the problems and opportunities discussed in participants' assessments, it may have posed a problem to give more instruction than we did. Alternately, future experiments following this outline could devise example problems and opportunities not representative of the instances found within the raw data, thus possibly avoiding bias, or at least preventing it from affecting results.

## **6. Operationalization of the variables.**

The measures used in this study produced scores for number identified, accuracy of assessment, and confidence in assessment for both problems and opportunities. These measures included simple counts, qualitative scoring by judges, and one single-item scale (confidence) borrowed from the marketing literature. While we attempted to maximize the probability that our techniques measured the concept intended, one can never be completely certain that a particular concept has been accurately operationalized and measured.

To summarize, this study suffers from the general limitations associated with experimental research in a laboratory setting. This implies that any generalizations of the finding must be applied with appropriate caution.

#### **4.7.2 Conclusions**

This experiment not only serves to compare the effects of NLP-based text mining with sentiment analysis and manual approaches, it also serves as an empirical test of the efficacy of our design principles for improving practice. The main implications of the study are that for mining organizational intelligence from highly unstructured text, NLP outperforms sentiment analysis, and sentiment analysis does not appear to perform any better than manually analyzing social media data as they stream by in real time. Rather, the evidence implies that sentiment analysis may actually serve to inhibit organizational success in detecting critical problems and important opportunities from customer-to-customer communication. While this interpretation follows from formal statistical analysis of the data, the results of this single experiment are insufficient evidence to discourage the pursuit of further investigation of the impact of natural language processing of social media data on organizational decision-making. While we draw some conclusions about what implications the results might herald for practice, this initial experiment should be considered a source of ideas for future research rather than a proposal for the practical application of the results.

#### **4.7.3 Future research**

The findings of this study support the pursuit of natural language processing proficiency and the application of the technology to information systems design. It also indicates that organizations relying on sentiment analysis to monitor customer input

and requests for output should be wary of its ability to effectively and accurately detect important information. While we have suggested potential explanations as to what factors might have confounded some of the opportunity measures, we do not know if there truly is no effect on opportunity number or confidence in detection despite the fact that the same factors of problem detection are affected. There may be something wrong with this experiment such that results were duly influenced, or we may merely lack power to detect a more subtle effect than that demonstrated on problem detection.

As a first experiment of a research program focused on the applicability of natural language processing to social media data, we suggest future research that can be formulated with reference to this study.

In particular, we note the effort expended prior to conducting the experiment to perfect various qualitative components of the research design, sometimes relying on trial and error and typically involving extensive time and effort of multiple participants. It is convenient, however, that we now have a framework for conducting future similar experiments with variations on certain elements of this study.

A particularly useful reusable feature of the experiment is the experimental interface. Because it was implemented using style sheets and server-side includes, it would be a trivial matter to reuse the interface with any type of raw data or measures for subsequent comparisons. For example, it would be useful to include differing sentiment analysis outputs in future variations on this experiment. While we used a typical word-based approach with a standard sentiment lexicon supplemented by industry-relevant terms, as NLP capabilities extend to support reliable phrase- or sentence-level analyses, empirical testing of the effects of those data on organizational decision-making would be

meaningful and useful. In certain contexts, more sophisticated sentiment analyses may prove useful and interesting.

While extensive testing enabled us to devise a methodology for simulating NLP output based on the principles of context, machine learning, and social-technical input, a logical future step would be to implement one or all of these features in a machine-executed algorithm, and conduct a similar investigation to this study but with data closer to that which would be the output of our desired system.

Another useful variation would be an investigation into differences in the robustness of analytics approaches across industries or even products. Do parallel corpora built from general data have the capacity to produce a statistical model capable of interpreting social media data pertaining to a narrow subject, or would this approach dilute the effectiveness of a more tailored model? Conversely, if training data are derived from a brand- or product-specific community, can the subsequent model accurately translate generic social media data, or be portable to a second product-specific social media discourse? Are statistical models relatively modular, or are they extremely sensitive to context? How sensitive? These research questions and more can be implemented rather painlessly within the structures and processes developed for our current experimental study.

Extending beyond the organizational setting, the medical field is a parallel practical domain in which the program of research suggested here could further evolve and provide additional insights into both text data mining approaches as well as social-media-facilitated behaviors of organizations and individuals. Pertaining to the former, machine learning algorithms have performed successfully in building classifiers for medical text reports, with critical implications for potential improvement in treatment

and diagnosis (Wilcox and Hripcsak 2003). Increased effectiveness of such efforts is critical, considering that the sheer volume of biomedical literature necessitates the application of text mining protocols to enable humans to locate, retrieve, and manage relevant information within such a vast sea of text (Spasic et al. 2005).

Given that the goal of biomedical research is to discover knowledge and practically implement it in the forms of diagnosis, prevention and treatment, developing technological capabilities to connect individual elements of biomedical knowledge with practitioners capable of using this knowledge is potentially a life-or-death endeavor. While isolated knowledge exists in the thousands of medical journals actively published to date (5619 currently indexed by Medline<sup>10</sup>), the fact remains that no human or even group of humans is in a position to process and absorb the entirety of this information in a meaningful way to make necessary, timely connections. Additionally, institutional characteristics of medicine further exacerbate this disconnect, including highly specialized fields and subfields and traditionally poor communication between disciplines (Cohen and Hersh 2005).

Manual efforts by humans to link previously-unconnected medical discoveries in unrelated and isolated journal articles indicate that advances in treatment and prevention are hiding in plain sight; much necessary data is extant and codified, ready to bear knowledge but for some mechanism to connect the related points. Large databases of scientific literature can yield discovery simply through the connection of concepts via logical inference. Thus, if A influences B in one article, B influences C in another articles, A may very likely influence C which thus bears subsequent

---

<sup>10</sup> U.S. National Library of Medicine at the National Institutes of Health:  
[http://www.nlm.nih.gov/bsd/num\\_titles.html](http://www.nlm.nih.gov/bsd/num_titles.html)

experimental and clinical evaluation(Weeber et al. 2003). Discoveries made according to this logic include the connection of magnesium as a treatment for migraine headaches through the linkage of 11 medical publications (Smalheiser and Swanson 1994), evidence of the therapeutic efficacy of fish oil on Reynaud's syndrome (Swanson 1986), the suggestion that thalidomide might be useful for treating acute pancreatitis, chronic hepatitis C, *Helicobacter pylori*-induced gastritis, and myasthenia gravis (Weeber et al. 2003), and the combined analysis of previously isolated statistics demonstrating the pervasiveness of iatrogenic illness (injuries and deaths caused by medical treatment) as the true leading cause of death and injury in the US (783,936 such deaths in 2001, followed by 699,697 deaths attributable to heart disease and 553,251.5 due to cancer) (Null et al. 2005). Given the impact of such manual efforts to uncover unrealized knowledge from the biomedical literature, the potential of an automated system to achieve similar future results, more quickly and on a larger scale, is extremely promising for the advancement of practicable medical knowledge.

Combining the domains of social media and medicine also appears to be an auspicious avenue of practical research; Web-based and often social, "Medicine 2.0" applications have emerged to target health care consumers, caregivers, patients, health professionals, and biomedical researchers. Medically-oriented social platforms enable such goals as health-related social networking, participation in health care decisions, and collaboration among and between providers and consumers of health care (Eysenbach 2008). Though not structurally dissimilar from customer-firm interchanges that unfold across social media platforms, as physicians incorporate social media tools into the treatment of their patients, a wide new realm of interesting and potentially measurable data are becoming available to support investigation into the effectiveness

of patient-physician social media exchange. Some possible research questions with very practical ramifications on human health and wellbeing include whether social media-based approaches can increase the effectiveness of such efforts as fighting against endemic obesity, increasing patient success with substance abuse cessation, or improving outcomes through increased adherence to pharmaceutical therapies. It is possible that social media tools may mediate a variety of health-related behaviors such that physicians can more successfully treat patients and patients can be more empowered to participate effectively in their treatment plans. Effects of social-mediated interactions may vary greatly from the effects observed in a customer-firm relationship, particularly given the very different nature of the goals and possible outcomes.

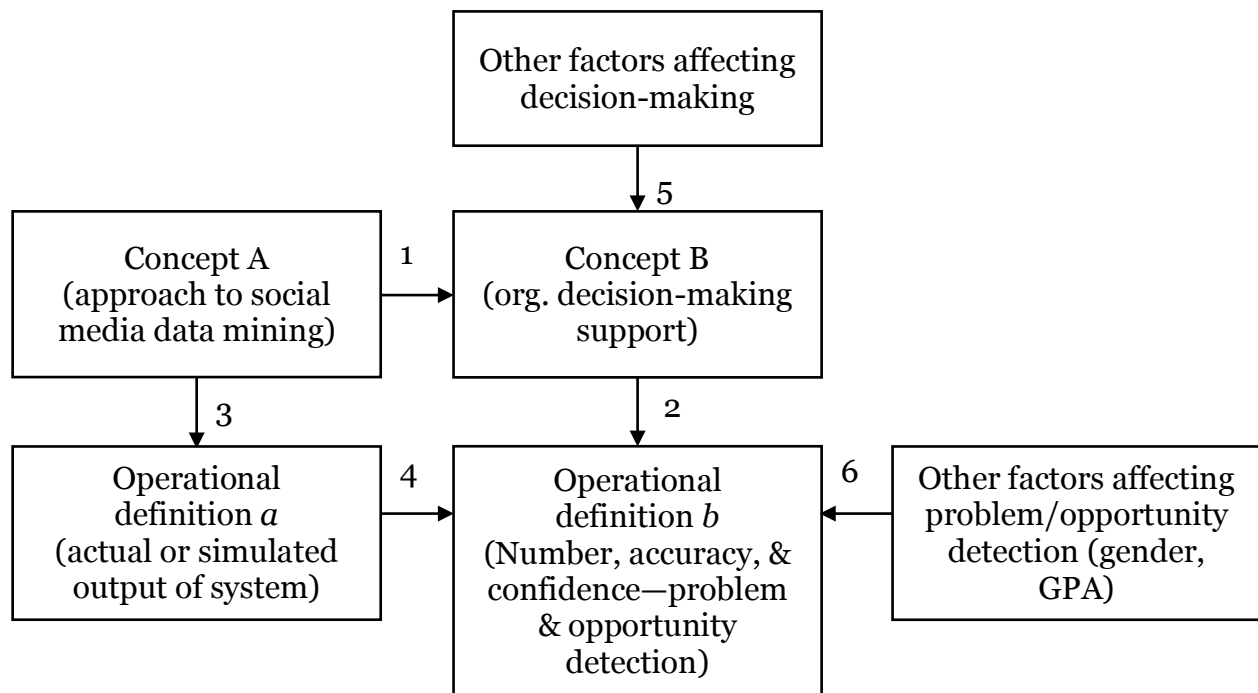
Whether applied to the context of a traditional organizational setting or ported into biomedical research or healthcare settings, it should be reiterated that this is the first study in which analytics approaches to mining qualitative social media data have been compared. Implications must be drawn cautiously, as it is only with the accumulation of knowledge through multiple studies of an area that inferences can be decisively drawn.

## **Appendix A – Regarding the Construct Validity Dilemma**

In our quest to test the relationship between constructs X and Y, we have developed operational definitions (x) for X and (y) for Y, then devised a setting in which to test the empirical relationship between x and y; from this investigation we draw an inference about the relationship between X and Y (McGrath 1981). There are four relationships implicated (see Figure 4-3): X-Y, which is conceptual and cannot be tested; X-x and Y-y which are definitional and can only be tested indirectly; and finally an empirical relationship, x-y. This is the “empirical lever” by which we assess the other three relationships.

We are interested in the conceptual X-Y relationship (1 in Figure 4-3) or the effect of the approach underlying our social media analytics system design on organizational decision-making. We presume that X-x (3) and Y-y (2) hold, so that we can use x-y (4) to test X-Y (1). As McGrath points out, the knowledge we acquire in the course of an experiment is contingent on assumptions. In the case that x-y is strong, we interpret it as evidence for X-Y. In our case, we interpret the strength of certain components of x-y as evidence in support of the conceptual relationship proposed between the approach underlying our social media analytics system design and organizational decision-making. Those components of x-y that do not appear to hold may be indicative that x is an inaccurate measure of X, y is an inaccurate measure of Y, that X-Y does not hold, or that the x-y data at hand is insufficient evidence. Figure 4-3 graphically presents the mappings of relationships among our concepts and measures tested in our experiment.





**Figure 4-6**  
**Rationale for predictive validity of decision-making quality**  
**(Extrapolated from McGrath 1981)**

## Appendix B – Experimental Interface

Social Media Analytics Interface Analysis Tool

**Raw Data Stream: Sunglass Hut Tweets** sunglass hut

**Decision-Assistance Panel (Random selection of 20 Tweets)**

|      |   |
|------|---|
| 182. | J9aholic: RT @Wedaad: @LifeisSavage and @Janez_Vermeiren @SunglassHutSA's sexiest men, yes?... YES!! :)   |
| 183. | Wedaad: @LifeisSavage and @Janez_Vermeiren @SunglassHutSA's sexiest men, yes?... YES!! :)   |
| 184. | agirlinbloom: RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!   |
| 185. | liezegouws: \ "@SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!\ "@RuanBeneke                             |
| 186. | BeautyDivaSA: @SunglassHutSA is it any designer shades Or only specific ones that qualify for the special?  |
| 187. | Puma_Bester: RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!  |
| 188. | denziljacobs: RT \034@SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!\035                                 |
| 189. | RT @danfast: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc in molestie odio. This is a retweet  |
| 190. | Pixie_Luna: RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!   |
| 191. | LifeisSavage: Ha! Did you hear me butcher my question! RT @SunglassHutSA: RT @_TOOLZ_: @LifeisSavage is better in person! Loves! Hahaha! #CosmoSexyMan      |
| 192. | Kirsty_Kirst: RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!   |
| 193. | Need an average length tweet? Aliquam venenatis risus sed lectus cursus luctus. Al  |
| 194. | louisludik: RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!   |
| 195. | marc_rainsford: RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!                                       |
| 196. | snanza: Nice e RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!  |
| 197. | thembi_l: Yay RT @janine_j: Yay!!!! RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!                   |
| 198. | janine_j: Yay!!!! RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!                                     |
| 199. | janine_j: Yay!!!! RT @SunglassHutSA: Valentine's Offer: Buy 2 pairs of designer sunglasses and get 40% off ur 2nd pair!                                     |
| 200. | SunglassHutSA: \ "@mirwin_z: Your staff at @SunglassHutSA Gateway are also friendly and helpful;) thanks for the great service!Thank you for the feedback:) |
| 201. | SunglassHutSA: RT @teeteenteta: Check out my @GLAMOUR_sa http://t.co/q9AL1Ltk cc @SunglassHutSA - rocking my @StellaMcCartney                               |
| 202. | Odile83: Just saw such a fab pair of shades at @SunglassHutSA! #birthdaywishlist  |

**Reminder:** You may opt out of answering any question that makes you feel uncomfortable by leaving the selection field unchanged.

**PROBLEMS**

1. How many problems regarding the "Sunglass Hut" brand did you detect from the Twitter stream?

2. Please describe the problems you identified:

3. On a scale of 1 (lowest) to 5 (highest), how confident are you that you were able to spot all the important problems for "Sunglass Hut"?  
☒ 1 - Not at all confident  
 ☐ 2 - Slightly confident  
 ☐ 3 - Somewhat confident  
 ☐ 4 - Very confident  
 ☐ 5 - Extremely confident

**OPPORTUNITIES**

4. How many opportunities regarding the "Sunglass Hut" brand did you detect from the Twitter stream?

5. Please describe the opportunities you identified:

6. On a scale of 1 (lowest) to 5 (highest), how confident are you that you were able to spot all the important opportunities for "Sunglass Hut"?  
☒ 1 - Not at all confident  
 ☐ 2 - Slightly confident  
 ☐ 3 - Somewhat confident  
 ☐ 4 - Very confident  
 ☐ 5 - Extremely confident

SUBMIT THIS FORM

DISCARD MY ANSWERS

**Figure 4-7**  
Screen shot of browser-based experiment interface

## Appendix C – Raw Data Set Used in Experiment

The 300 Tweets comprising the raw data stream given to all experiment participants were obtained using Twitter’s search interface, scraped as HTML, and then imported into the R statistical environment for subsequent sentiment analysis (for condition 2) or simulation of NLP-based analysis (condition 3). The focal brand of our search was Sunglass Hut, so we first executed a search for the phrase “Sunglass Hut.” Subsequently, we searched for the term “sunglasshut” in order to detect results related to two types of Twitter-specific terms:

- **@SunglassHut:** the @recipient is a Twitter mechanism for indicating the name or handle (online pseudonym) of a Twitter account’s owner, often an individual but possibly a brand or business. A messaging account can use “@recipient” to tag the recipient’s Twitter account in the Tweet, embedding a link to the recipient’s account from within the message. When “@recipient” is located within the body of the Tweet, it typically indicates that the message is directed toward that Twitter account holder. For example, a consumer may direct a complaint or suggestion to Sunglass by including “@SunglassHut” in the Tweet.
- **#SunglassHut:** the hashtag (#) essentially enables Twitter authors to add metadata to their Tweets, usually either to ensure the message is easily findable in conjunction with a particular concept search, or to add contextual meaning to a limited microblog entry (140 characters maximum). For example, the Tweet **I love my new job :) #hellodiscount :)))) #sunglasshut #makingbank #blessed** lends insight into the author’s reasons for valuing his or her job: first, the author enjoys a discount, as indicated by the hashtag and underscored by the smile emoticon with four

parentheses instead of one (superlative to a simple one-parenthesis smile emoticon). Additionally, the Tweet author indicates feeling “blessed” due to the fact that he or she is earning high wages (making bank).

Relevant to the research context at hand, hashtags may also reduce ambiguity; whereas “I love my job” could be meant sarcastically, the inclusion of contextualizing hashtags signals to a reader that in this case, the phrase is meant literally. Thus, the conveyance of metadata through hashtags is extremely useful for ensuring accuracy of interpretation of a short (140-character) message.

Aside from reviewing each Tweet to ensure it did not contain anything potentially offensive to our student research pool (one message including a derogatory remark were stricken from the sample), we did not further manipulate the raw data stream in any artificial way.

## Appendix D – Decision-Assistance Panel Contents

**Condition 1.** The decision-assistance panel for the first condition, compiled without any degree of automated analytical preprocessing, contained a randomly selected set of twenty continuous Tweets. The following set in one such sample, containing Tweets 77 through 96 of the raw data stream:

**Table 4-9**  
**20 Tweets comprising manual treatment – sample set**

|     |  |
|-----|--|
| 77. | Michelle Gray @mogy37:<br>Check out Inner Circle - the ultimate style hotspot.<br>bit.ly/I1jopF  |
| 78. | Jason Bring @JasonBring:<br>Ordering these sunglasshut.com/webapp/wcs/sto  |
| 79. | Latarsha ;) @FashionableLola:<br>Bout to check them out @sheislex @FashionableLola check out<br>sunglasshut online   |
| 80. | @sheislex:<br>@FashionableLola check out sunglasshut online<br>In reply to Latarsha ;)   |
| 81. | Wendy Christidis @Wendy9295:<br>Check out Inner Circle - the ultimate style hotspot.<br>bit.ly/IsV7WW  |
| 82. | Sandra Mills @Sandym68:<br>Check out Inner Circle - the ultimate style hotspot.<br>bit.ly/HRTnnV   |
| 83. | Tracy Boulter @JesseBelles:<br>Check out Inner Circle - the ultimate style hotspot.<br>bit.ly/J0eYPq   |
| 84. | Shea @Best_21:<br>#BirthdayGift Item # 8 : google.com/imgres?q=sungl   |
| 85. | Rochelle Fox @Rochelle_Fox:<br>@sunglasshut Think I am going to have to pop into a store<br>before @MBFWA after seeing this!!!<br>In reply to Sunglass Hut |
| 86. | Sunglass Hut @sunglasshut  |

|     |  |
|-----|--|
|     | Sunglass Hut Inner Circle - Burberry's Latest Collection<br>fb.me/1TOv0ijk7  |
| 87. | J A C Q U E @babyjxxx<br>@sroman12 @treyhoe #sunglasshut instagr.am/p/JoH-VKMAV3/  |
| 88. | Bob Tierney @BT_612<br>Hmm. I think I'll visit @sunglasshut and pick up a nice pair of freshies this wknd. #selfreward #stayclean #bigsmile                                  |
| 89. | Ellie Rigsby @EllieNicoleRigs<br>#sunglasshut #tulsa #downtown #friends #fun #stairs<br>instagr.am/p/JoBPJmPrln/   |
| 90. | Ellie Rigsby @EllieNicoleRigs<br>Downtown Tulsa #fun #friends #downtown #tulsa #sunglasshut<br>instagr.am/p/JoA-Yevr1X/  |
| 91. | ThatGirl Laura w @WoodsLaura<br>Went to the #sunglasshut today....Tried on soooo many pairs of sunglasses ...I feel the need for a new pair = hopefully soon that can happen |
| 92. | Joe Geniti @fondAfondal<br>@iamnotrichard there's a pair of versace sunglasses that you would rock the shit out of. #sunglasshut #comevisit<br>In reply to richard           |
| 93. | Nuffnang Australia @nuffnangAU Don't forget you can win your way to Fashionopolis @sunglasshut. Includes flights, accom and tix to the event bit.ly/IMvr4S #NNF2012          |
| 94. | Ellie Rigsby @EllieNicoleRigs<br>Tulsa Downtown #fun #friends #downtown #sunglasshut<br>instagr.am/p/Jn8fCaPrzS/   |
| 95. | Lady Melbourne @ladymelbourne<br>Morning all! Got your tickets to Fashioniopolis by @sunglasshut yet? What are you waiting for?! is.gd/Olktdw #NNF2012                       |
| 96. | Toyah Harris @Bestmomalive25<br>I need a new pair of RayBans 2 add 2 my collection<br>sunglasshut here I cum   |

**Condition 2.** This panel, produced according to word-level sentiment analysis-based analytics processing, contained the following top ten most negative and top ten most positive Tweets, selected according to methodology outlined in Appendix E:

**Table 4-10**  
**20 Tweets comprising SA treatment**

|      |   |
|------|---|
| 3.   | Thats what she said! @easy_joke:<br>RT @cubanmonkey69 Never new 'It was so hard' to merchandise a #sunglasshut kiosk! LMAO they have been @ it for 3 hrs lol. . #TWSS                     |
| 7.   | Ellie Rigsby @EllieNicoleRigs:<br>Flea market earrings :) #yellow #work #sunglasshut #fun #fashion #sunglasses #vintage instagr.am/p/JxBSfGvrzh/  |
| 24.  | UA Hitmaker Shawty @lilsteve_beatz:<br>SunglassHut is a great place with great associates. I'm having fun on this business trip, just coolin...   |
| 35.  | Sunglass Hut @sunglasshut:<br>Sunglass Hut is proud to announce this week's lucky winners of the Inner Circle competition and a \$200 Sunglass. . . fb.me/1G1RyGJ                         |
| 48.  | Sofia Sbordonni @SofiaSbordonni:<br>#sunglasshut #bored instagr.am/p/Js6yETK30i/  |
| 53.  | Alex Hardy @alexhardyuk:<br>@gordonhf I saw Tarik today. He now works at Westfield #sunglasshut selling genuine fake Gucci. Free shot of #raki with every pair.<br>In reply to Gordon H-F |
| 96.  | Toyah Harris @Bestmomalive25:<br>I need a new pair of RayBans 2 add 2 my collection sunglasshut here I cum  |
| 100. | Amanda Palisi @PrincessssAmanda:<br>White #wayfarers ? Yeah I think I need to get these. . #sunglasshut instagr.am/p/JnLqUluisF/  |
| 124. | Luci @LuciLuu_x3:<br>I want a mean dark pair of Chanels!!! Oh the only thing i miss of #sunglasshut!!!  |
| 133. | South Beach Playboy @dapolashes:<br>@ThisisAshleyB1 @CranDan1 uv gt a frikin nerve Mr go sunglasshut 7 times and buy nutin #savediana   |
| 135. | Thom Whilton @couturing:<br>@LisaCouturing @persoleyewear @sunglasshut LOVE IT! The orange Steve McQueens are a classic!<br>In reply to Lisa Teh (message 140)                            |

|      |  |
|------|--|
| 141. | Lisa Teh @LisaCouturing:<br>Love the sexy, new advertising campaign for @StellaMcCartney's new eco range at @sunglasshut. Looks like gre <a href="https://www.instagram.com/p/JjU8ahtuRq/">instagr.am/p/JjU8ahtuRq/</a>  |
| 193. | DealForGirl.com @dealforgirl:<br><a href="https://dealforgirl.com/view/14897">dealforgirl.com/view/14897</a> Sunglass Hut Coupons: \$20 or \$50 Bonus Code with Full Priced Purchase + Free Next Day Shipping  |
| 198. | Gracie Dzienny @GracieDzienny:<br>How do you like our new shades?! Fun shopping stop @sunglasshut @ihavetude and @nickdeeez :) <a href="https://pic.twitter.com/mFBfJ4gC">pic.twitter.com/mFBfJ4gC</a>   |
| 210. | . . ::: Menzi :::. . @MrMenziN:<br>Tjo RT @Lady_Crunk: Waiting 2hours outside SunglassHut's JandB Met tent for @Janez_Vermeiren RT MrMenziN: What's the craziest groupie?  |
| 211. | Olwethu-Thando Klaas @Lady_Crunk:<br>Waiting 2hours outside SunglassHut's JandB Met tent for @Janez_Vermeiren RT @MrMenziN: What's the craziest groupie stunt you've pulled before?  |
| 240. | BradsDealsApparel @BradsApparel:<br>Sunglass Hut: Up to 50% Off + Ship Free: Save up to 50% on sale sunglasses at <a href="https://SunglassHut.com">SunglassHut.com</a> . Even better w. . . <a href="https://bit.ly/I85HQD">bit.ly/I85HQD</a><br>Hank @HenryHerrera |
| 270. | @BL11Hannah @raybanglassesus:<br>@sunglasshut ugh! That is the worst! I work on elevated train tracks and I dropped mine on accident #Done! :-(<br>In reply to Hannah Curlee (message 282)   |
| 282. | Hannah Curlee @BL11Hannah:<br>So sad my new @raybanglassesus from @sunglasshut were stolen today at b'fast! I just bought them!<br>#anotherpairgone  |
| 293. | krislloyd enriquez @krislloyd84:<br>PSA: ladies if ur looking for hot shades for summer do checkout #Melodies by #MJB @maryjblige I highly approve! found at @sunglasshut  |



**Condition 3.** The final panel, generated according to advanced NLP-based simulated analytics processing, contained the following top twenty most important Tweets, selected according to the algorithm depicted in Figure 4-5 (p. 153)Error! Reference source not found.:

**Table 4-11**  
**20 Tweets comprising NLP treatment**

|     |  |
|-----|--|
| 2.  | ki designs @pachuckidesigns<br>the #coach sunglasses aren't even listed on the<br>#sunglasshut website yet. eeek   |
| 3.  | Thats what she said! @easy_joke<br>RT @cubanmonkey69 Never new 'It was so hard' to merchandise<br>a #sunglasshut kiosk! LMAO they have been @ it for 3 hrs<br>lol. . #TWSS     |
| 4.  | Liz Gallagher @cubanmonkey69<br>Never new it was so hard to merchandise a #sunglasshut<br>kiosk! LMAO they have been @ it for 3 hrs lol  |
| 38. | King Shambles @FCShambles<br>@sunglasshut thanks for having an awesome return policy on<br>new shades. #absolutequality  |
| 42. | Natalie Mulford @XLaslife<br>Okay, so I will die if I can't get the PINK glitter Noir<br>@miumiustlye sunnies. @sunglasshut are you getting any in<br>Australia? @luxottica_au |
| 51. | Dynafit @dynafitNA<br>@sunglasshut seriously awesome crusty service at 29th<br>street in boulder. Thanks Dave + Evan!!!  |
| 54. | Drew S. @uknowitsdrew<br>Props to @sunglasshut for giving me customized sunglasses!!<br>#ThanksDiana #HappyBirthdayToMe  |
| 55. | Allison Pior @allisonp6<br>Loving @sunglasshut right now. Ran in to get a new pair of<br>aviators and they exchanged my wayfarers that tobex<br>destroyed also. #happysaturday |
| 58. | 3DayCapt @3DayCapt<br>@sunglasshut great customer service at Stonestown Galleria,<br>SF, replaced defective Maui Jims, \$300+ sunglasses, no<br>questions asked.               |
| 73. | Jeremy Smith @nothelsemttrs<br>#Sunglasshut needs to get some coupons out asap before I<br>have to get some cheaper glasses for my groomsmen at<br>lenscrafters :(             |

|      |  |
|------|--|
| 74.  | Kusai Zarrugh @kazarrugh<br>#SunglassHut said they can fix my #rayban for free. Awesome :)   |
| 108. | Lauren @StylizedEx<br>Pretty disappointed with the customer service I have received from @sunglasshut!   |
| 118. | Thom Whilton @couturing<br>@LisaCouturing @persoleyewear @sunglasshut those are the ones! Foldables are a huge trend in sunnies!<br>In reply to Lisa The (message 122)   |
| 128. | MolliesFund @MolliesFund<br>Proper eyewear is another component to practicing #safesun. We like this pair from @sunglasshut for the ladies.<br>bit.ly/J34RGj   |
| 141. | Lisa Teh @LisaCouturing<br>Love the sexy, new advertising campaign for @StellaMcCartney's new eco range at @sunglasshut. Looks like gre <a href="https://www.instagram.com/p/JjU8ahtuRq/">instagr.am/p/JjU8ahtuRq/</a> |
| 215. | IamConquer @msconquer<br>#Sunglasshut Brooklyn is on point. Thank u guys for hooking me out with the stylish #Rayban, gonna look sueve this Saturday.  |
| 244. | Danielle Hervey @daniellecuz<br>RT @thecoolhunter: Clever way to promote sunglasses <a href="https://pic.twitter.com/ayLfBkGf@sunglasshut_sa">pic.twitter.com/ayLfBkGf@sunglasshut_sa</a>                              |
| 269. | Joseph Aleo @josephaleo<br>Closed? In the afternoon? (at @sunglasshut)<br><a href="https://path.com/p/1wvWnQ">path.com/p/1wvWnQ</a>  |
| 272. | Alisson Cancado @alissoncancado<br>@sunglasshut i have a problem my on my ship please contact me i sent a email but didnt fix it yet.  |
| 281. | Lorena Azizeh @LorenaAzizeh<br>I wish @sunglasshut would train their staff. Went in to get the artist series glasses but no one knew what I was talking about or cared!  |

## Appendix E – Description of Tweet Selection, Sentiment-Based Condition

We generated the decision-assistance panel of Condition 2 by analyzing the raw data according to a standard (word-level) sentiment-based approach. We adopted the methodology<sup>11</sup> (Miner et al. 2012) that leverages Hu and Liu's opinion lexicon (Hu and Liu 2004) and the below function, `score.sentiment`,<sup>12</sup> within the R statistical software environment, a GNU project developed at Bell Laboratories and available as Free Software. While many free (as well as for-fee) sentiment-based analytical tools are commercially available, all depend on the thoroughness and accuracy of the lexicon upon which their analyses are based. To achieve the most unbiased and realistic results possible, we augmented Hu and Liu's lexicon used with superlatives expected to appear in the colloquial speech, particularly in the given setting (fashion-related microblogs).

```
score.sentiment = function(sentences, pos.words, neg.words,
  .progress='none')
{
  require(plyr)
  require(stringr)

  # we've got a vector of sentences. plyr will handle a list
  # or a vector as an "l" for us
  # we want a simple array ("a") of scores back, so we use
  # "l" + "a" + "ply" = "laply":
  scores = laply(sentences, function(sentence, pos.words,
neg.words) {
```

---

<sup>11</sup> Originally presented by Jeffery Breen at the June 2011 Boston Predictive Analytics Meetup and subsequently published in *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications* (Miner et al. 2012).

<sup>12</sup> Function publically available at <http://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/>.

```

    # clean up sentences with R's regex-driven global
    substitute, gsub():
        sentence = gsub('[:punct:]]', '', sentence)
        sentence = gsub('[:cntrl:]]', '', sentence)
        sentence = gsub('\\d+', '', sentence)
        # and convert to lower case:
        sentence = tolower(sentence)

    # split into words. str_split is in the stringr package
    word.list = str_split(sentence, '\\s+')
    # sometimes a list() is one level of hierarchy too much
    words = unlist(word.list)

    # compare our words to the dictionaries of positive and
    negative terms
    pos.matches = match(words, pos.words)
    neg.matches = match(words, neg.words)

    # match() returns the position of the matched term or NA
    # we just want a TRUE/FALSE:
    pos.matches = !is.na(pos.matches)
    neg.matches = !is.na(neg.matches)

    # and conveniently enough, TRUE/FALSE will be treated as
    1/0 by sum():
    score = sum(pos.matches) - sum(neg.matches)
    return(score)
}, pos.words, neg.words, .progress=.progress )

scores.df = data.frame(score=scores, text=sentences)
return(scores.df)
}

```

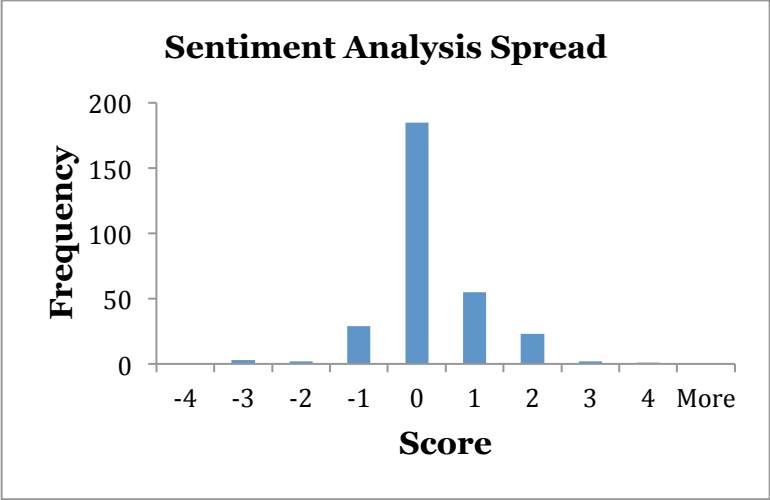
Prior to scoring the raw data, we appended Hu and Liu's lexicon with the following positive and negative words expected to appear in the colloquial speech of

discussants of Sunglass Hut. This particular set of words was compiled by identifying clearly positive and negative words within the contents of the raw data to ensure accuracy of the sentiment analysis of that data set; we would proceed with additional augmentation based on additional raw data for future experiments.

**Table 4-12**  
**Words appended to SA lexicon**

| Positive      |              |               |              | Negative     |             |
|---------------|--------------|---------------|--------------|--------------|-------------|
| 1. cared      | 11. friends  | 21. practical | 31. stunna   | 1. accident  | 11. forget  |
| 2. coolin     | 12. funny    | 22. prized    | 32. style    | 2. ass       | 12. forgot  |
| 3. crusty     | 13. genuine  | 23. promote   | 33. thanks   | 3. closed    | 13. frikin  |
| 4. customized | 14. giveaway | 24. props     | 34. train    | 4. craziest  | 14. mean    |
| 5. deal       | 15. heart    | 25. rad       | 35. trend    | 5. destroyed | 15. need    |
| 6. deals      | 16. hotspot  | 26. replaced  | 36. ultimate | 6. dropped   | 16. peeling |
| 7. discount   | 17. kiss     | 27. rock      | 37. woohoo   | 7. eek       | 17. probs   |
| 8. dreamjob   | 18. luv      | 28. rocking   | 38. yay      | 8. eww       | 18. shits   |
| 9. exclusive  | 19. new      | 29. sale      |              | 9. fake      | 19. waiting |
| 10. fix       | 20. party    | 30. service   |              | 10. fee      |             |

The sentiment analysis resulted in eight clear positive and negative extremes: Tweets 124 (-3), 211 (-3), 270 (-3), 210 (-2), 282 (-2), 35 (+3), 141 (+3), and 24 (+4). The spread of sentiment scores across the stream of 300 Tweets (ranging from -3 to +4) is small due to the limited number of words possible within a 140-character limit. To fill out the top ten most negative and top ten most positive decision-assistance panel, we randomly selected five Tweets from the set of twenty-nine that scored -2 and seven from the set of twenty-three that scored +2. The clear majority of Tweets scored as overall neutral (0).



**Figure 4-8**  
**Sentiment analysis score histogram**

**Table 4-13**  
**Distribution of SA scores**

| Score | Frequency |
|-------|-----------|
| -4    | 0         |
| -3    | 3         |
| -2    | 2         |
| -1    | 29        |
| 0     | 185       |
| 1     | 55        |
| 2     | 23        |
| 3     | 2         |
| 4     | 1         |

## **Appendix F – Instructions Given to NLP Simulation Card Sorters**

### **Algorithm simulation – Human process**

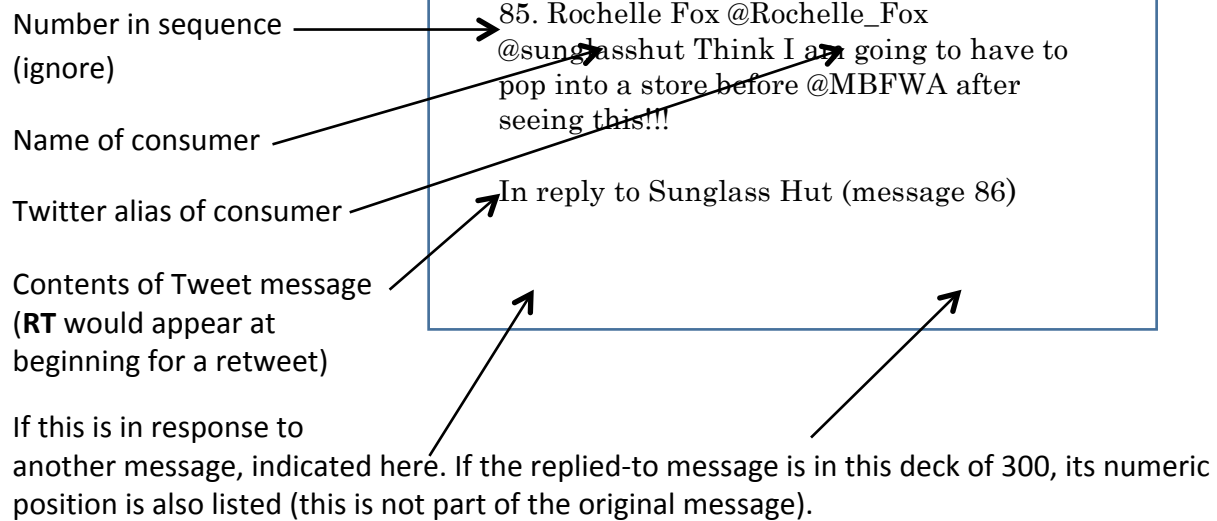
During this procedure, you will sort out a pile of Tweets (microblog messages with a maximum of 140 characters) and come up with a set of the most important messages you would hand off to senior management to read if you could only give them 20 (also please note your additional ten, for a total of 30). The messages you choose should contain information that can tell the firm something about its customers. These messages might contain, for example, problems the firm needs to address, requests that an intern might be able to handle, or indications of what the firm or its subsidiaries are doing well (which thus give the firm insight into what its customers value). You don't have to order your results, just let me know which comprise your top 20 and which are only in the top 30.

One technical note—keep in mind that “RT” means retweet, which means it's a rebroadcasting of the original message. This may indicate increased importance.

Because it contains the sorting instructions, here is the actual sorting process you will read at the end of the last page. Keep it in mind:

**Final Sort:** Considering the types of things discussed above that are considered important in a firm context, the final step is to sort the remaining cards into stacks of “fluff,” or messages that don't really contain anything interesting or useful to the company, and “important.” The stack of important can break down however you like, but should ultimately contain the 20 cards you consider to be the most important messages for a social media intern to hand over to management.

## Anatomy of a card



Many Tweets also contain hashtags (e.g., #sunglasshut, #dreamjob, #sarcasm). These are often used as metatags to give context and meaning to a message and facilitate searching.

## Card Reduction Procedure:

(Because **Step 1** does not require judgment, all parts of it have been completed for you. It is listed as an **appendix** following these instructions to let you know what has already been removed from the stack of cards.)

Steps 2 and 3 list out what has likely occurred naturally as you've read and decoded the language on each card. These are steps that a machine must be programmed to do, and are important for clarification of meaning to an automated system. These are listed for your consideration, but do not require action on your part.

### 1. Signal Disambiguation

#### a. Linguistics normalization

- i. **Resolve sarcasm:** Keep in mind that Twitter users often use the hashtag "#sarcasm" to explicitly clarify that their message should be interpreted as sarcastic.
- ii. **Resolve misspellings:** e.g.: **IamConquer** @msconquer:  
#Sunglasshut Brooklyn is on point. Thank u guys for hooking me out with the stylish #Rayban, gonna look **sueve** this Saturday.



## **b. Interpretation**

- i. **Translate slang:** e.g., “sunnies” = “sunglasses”
- ii. **Decode abbreviations and text speak:** e.g., “<3 this” means “love this!”

## **2. Incorporation of Sociotechnical-Machine Learning procedures**

- a. **Named Entity Recognition (NER):** Whereas it is usually clear to humans that “Sue married Joe” contains two named entities, this must be programmed for a machine.
- b. **Relationship extraction:** Whereas it is usually clear to humans that “Sue married Joe” contains a relationship, this must be programmed for a machine.

\*\*\* Now, getting down to selection of your final 20/30: Note that we aren’t asking you to come up with extremely positive and extremely negative messages, per se—these messages might contain emotional words, but not be relevant to the company. However, emotional words *in combination with other info* may indicate something really important that the company might need to know about the consumer.

In your final sorting, please keep the following criteria in mind (step 4). These are aspects of messages we can program a machine to detect in order to identify potentially important messages that management might need to look at.

## **3. Signal Refinement**

### **a. Sentiment Analysis**

- i. **Detect extreme sentiment:** Sentiment extremes are important (though not the whole story). Extreme language will likely signal to us an important message, but more specifically we are looking for problems the firm should address, opportunities in the form of customer service, etc. As well as messages that convey approval regarding aspects of the company’s services or brand that should be noted by mgmt., as well.

### **b. Co-creation cues**

- i. **Detect suggestions:** Suggestions directed toward the firm are likely important messages. These may herald useful opportunities as well as suggest addressable problems.
- ii. **Detect requests:** Requests directed at the firm, especially those entail customer-follow up, are likely important messages. These may also herald useful opportunities for excellent customer service, or suggest addressable problems.

### Final Sort:

Considering the types of things discussed above that are considered important in a firm context, the final step is to sort the remaining cards into stacks of “fluff,” or messages that don’t really contain anything interesting or useful to the company, and “important.” The stack of important can break down however you like, but should ultimately contain the 20 cards you consider to be the most important messages for a social media intern to hand over to management.

Please list the numbers of your top 20 here:

|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

Please list the additional ten that round out your top 30:

|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|

### Appendix to card sorting instructions:

The following steps were done to cull cards prior to human sorting procedure, but are included here for your reference to understand what was discarded and why.

#### 2. Noise-to-signal ratio reduction

##### a. Noise Filtering

###### i. Remove advertisements, contests, etc., e.g.:

**London Luton Airport @LDNLutonAirport:** @sunglasshut are looking for a full time sales associate at the airport. Apply here: [ow.ly/anxTl](https://ow.ly/anxTl)) **or i ♥ pretty things @iheartdesign2:** Inner Circle - Fashion Tips Blog From Sunglass Hut [fb.me/1wrYb18pW](https://fb.me/1wrYb18pW)

###### ii. Remove spam, coupon codes, etc., e.g.:

**US Promo Coupons @uscouponcodes:** Sunglass Hut - Shop special offers at Sunglass Hut: Shop special offers at Sunglass Hut. To claim this Discount . . . [bit.ly/mAwsVc](https://bit.ly/mAwsVc))

**b. Signal Amplification**

- i. **Consider Retweets:** When a message contains “RT” (meaning retweet), this strengthens the original message because it means it is being re-broadcast through the system. Simply keep track of RTs because this may bump up their importance.

**c. Signal Specification**

- i. **Remove non-English messages:** Foreign language translations is outside the scope of this information system.
- ii. **Remove employee-originated messages:** Remove Tweets by Sunglass Hut employees regardless of content, as these are outside of our dyad of interest (we only want to look at customer-to-customer or customer-to-firm messages in this study). E.g., eliminate:

**Megan Morfe @megnelizabeth:** Meeting in Virginia for work! #sunglasshut

or

**High Heffner @ImSoMarilynx3:** Off to #work #dancing for them #dollas Lmao at #SunglassHut  
instagr.am/p/JIbrD5TXx1/

- iii. **Remove photo- or link-only messages:** If a Tweet contains only a link to a photo or a website WITH NO contextualizing hashtags, eliminate it because we do not have capabilities for automated analysis of image or video content. However, if hashtags are included, retain the message because it’s possible (though not likely) the hashtags convey something important. E.g.,

*eliminate -* **Barrett P. @Barrett\_Primmer:**  
mobile.usablenet.com/mt/www.sunglas

*retain -* **JuanCuba @juancubanation:** #Fashion #lanvin  
#chanel #sunglasshut #sgh instagr.am/p/JsX74PD9-t/

## Appendix G – Hypergeometric Distribution

Borrowing a technique from genetics used to estimate the probability that overlapping sets of genes occur by chance alone in the analysis of differential expression under two conditions, we are able to determine the likelihood that two NLP simulation outputs overlap by  $m$  Tweets by chance (Fury et al. 2006). If two lists of genes are selected out of  $N$  genes randomly, the probability of  $m$  genes in common between the two lists of lengths  $n_1$  and  $n_2$  is known to follow the hypergeometric distribution.<sup>13</sup> Given these integers  $N$ ,  $n_1$ ,  $n_2$ , and  $m$  where  $\max(n_1, n_2) \leq N$  and  $m \leq \min(n_1, n_2)$ , the hypergeometric distribution is defined as

$$P(m) = \frac{C(n_1, m)C(N - n_1, n_2 - m)}{C(N, n_2)} = \frac{\binom{n_1}{m} \binom{N-n_1}{n_2-m}}{\binom{N}{n_2}}$$

where  $C(n, m)$  is the number of possibilities of choosing  $m$  objects out of  $N$  objects:  $C(N, m) = N!/[m!(N-m)!]$ . When  $n_1$  genes are randomly chosen from a total of  $N$  genes, and another random sampling leads to  $n_2$  genes, the probability that the two lists of genes have  $m$  in common is exactly the hypergeometric probability  $P(m)$ . Thus, if  $n_1 = 30$  Tweets are selected from a total of  $N = 140$  possible Tweets and a second sample produces  $n_2 = 30$  Tweets from the same population,  $p(m)$  that the two lists will share  $m = 15$  Tweets in common due to chance equals, approximately,  $\frac{1.78}{3.1719 \times 10^{-6}}$ .

The explanation is as follows:

---

<sup>13</sup> The hypergeometric distribution relates to the binomial distribution, which describes the probability of  $m$  successes in  $N$  **with** replacement.

1. The total number of possible choices for the two lists of Tweets is  $C(140,30) \cdot C(140,30)$ .
2. There are  $C(140,30)$  possibilities for choosing the first list.
3. Among the 30 Tweets in the first list, there are  $C(140,15)$  possibilities for choosing 15 Tweets to be in common with the second list.
4. In the second list, besides the 15 Tweets that are in common with the first list, the remaining  $30 - 15$  Tweets are chosen among the  $140 - 30$  “leftover” Tweets not in the first list, thus  $C(140-30,30-15) = C(110,15)$  possibilities.

The  $P(m)$  is simply  $(\#2 \times \#3 \times \#4) / \#1$ . Note that  $n_1$  and  $n_2$  are interchangeable without changing the  $P(m)$  value.

$P(m)$  for overlap of two lists ( $n_1 = n_2 = 30$ ) drawn from  $N=140$  possible Tweets sharing  $m = 15$  Tweets in common:

$$P(m) = \frac{\binom{n_1}{m} \binom{N-n_1}{n_2-m}}{\binom{N}{n_2}} = \frac{C(n_1, m) C(N - n_1, n_2 - m)}{C(N, n_2)} = \frac{C(30, 15) C(110, 15)}{C(140, 30)}$$

$$\text{where } C(a, b) = \frac{a!}{[b! (a - b)!]}$$

$$= \frac{\left[ \frac{30!}{15! (30 - 15)!} \right] * \left[ \frac{110!}{15! (110 - 15)!} \right]}{\frac{140!}{[30! (140 - 30)!]}} = \frac{\frac{30!}{15! (15)!} * \frac{110!}{15! (95)!}}{\frac{140!}{30! (110)!}}$$

$$\approx \frac{[1.55 \times 10^8] * [1.18 \times 10^{18}]}{3.20 \times 10^{30}} \approx \frac{1.83 \times 10^{26}}{3.20 \times 10^{30}}$$

$$\approx \frac{1.83}{3.20 \times 10^4}$$

More interestingly, the sum of  $P(m)$  for  $m$ 's equal or larger than the observed value (i.e., the p-value) is:

$$\sum_{k=m}^{\min(n_1, n_2)} p(k) = \sum_{k=0}^{\min(n_1, n_2)} p(k) - \sum_{k=0}^{m-1} p(k).$$

In the R statistical package, p-value is calculated using the accumulative distribution of hypergeometric distribution,  $phyper(m, n_1, n - n_1, n_2)$ :

$$\begin{aligned} \text{p-value} &= phyper(\min(n_1, n_2), n_1, n - n_1, n_2) - phyper(m-1, n_1, n - n_1, n_2) \text{ if} \\ &m > 0, \text{ and p-value} = 1 \text{ if } m = 0. \end{aligned}$$

Significance of overlapping  $m = 15$  from  $n = 140$  between list  $n_1 = 30$  and list  $n_2 = 30$  is indicated by p-value =  $6.65 \times 10^{-5}$ .

## **Appendix H – Problem/Opportunity Categorization of NLP-simulated Output for Accuracy Scoring**

To score “accuracy of problem identification” and “accuracy of opportunity identification,” we implement a numerical scoring methodology derived from qualitative category identification. Assessment accuracy scoring comprises a calculus of combining a positive component for the percentage of important problems (or opportunities) identified from the raw data, plus a negative component based on the number of problems (or opportunities) identified but not actually appearing in the raw data and those Tweets considered irrelevant according to the instructions given (for example, messages issued not by a consumer but by an employee or another company).

We draw on the outputs of the NLP-based simulation algorithm to identify the “most important Tweets” that form the basis of the positive score component. To compensate for variation introduced by human judgment, we consider a wider range than just the top twenty Tweets narrowed down in multiple rounds of NLP-simulation sorting; we expand this list to all Tweets contained in the sorting outputs of at least two of the three original sorters (see Table 4-5) for a total of 29 Tweets with double agreement. We do not consider Tweets output by a single rater, as the chance is too large that these are outliers.

During the experiment, we ask subjects to assess the problems and opportunities contained in the raw data *in their own words*, as opposed to simply cutting and pasting individual Tweets that they think contain problems or opportunities. We phrase the task in this manner to encourage evaluation by the participants as opposed to a simple tally, to simulate the type of real-life synthesis required when a manager requests a summary

report from an employee. In this scenario, a mere list of largely un-evaluated data points unconnected by narrative would be inappropriate.

To avoid penalizing accuracy scores due to granularity lost to the generalization inherent in the summarization process, we similarly organize the 29 Tweets incorporated into our target list into descriptive categories. Instead of tallying according to individual Tweets identified, we tally proportionately according to categories represented in the assessment. These categories and proportions are derived from the sorts of two independent raters with discrepancies reconciled by one of the authors as follows:

Separate lists of unsorted problems (7 of 29) and opportunities (22 of 29) were given to two raters; both were asked (independently) to sort the problems and opportunities into any number of categories, according to dimensions **of their choosing**. As such, two exploratory analyses were essentially conducted from scratch. Rater #1 organized problems into four categories, as did Rater #2 (see Table 4-16). Rater #1 structured opportunities into five categories, again as did Rater #2 (see Table 4-17). Categories devised by the raters are compared and contrasted in Table 4-16 and Table 4-17, with resultant categories (as synthesized by the author from the two raters' conceptualizations) indicated.

Accounting for Tweets cross-listed under multiple categories, proportions are assigned to each resultant category (based on number of Tweets represented by the category) in order to form the seven-point scale used to score problem and opportunity assessment accuracies. For each category of problem indicated by an experimental subject's assessment, a corresponding positive score component is thus added to the



accuracy rating, weighted by proportion of the number of individual Tweets belonging to that category.

Once categories of problems and opportunities were aggregated and score proportions assigned, actual participant data was scored by an author of the paper with extreme familiarity with the raw social media data. Participant assessments were assigned random IDs, shuffled into random order, and thus scored blindly. The scorer quantitatively tallied each problem and opportunity assessment's accuracy by adding score components for problems and opportunities accurately identified from the raw data according to the distribution devised by prior categorization. Scores ranged from 1 (for no accurately identified problem/opportunity) to 7 (for identification of problems/opportunities from every single category pre-specified. IDs were then used to rematch these scores with the original record for analysis.

The following list presents the 29 problems and opportunities sorted by the two qualitative raters. Each rater received the lists of problems and opportunities in random order with ID numbers assigned; list item numbers are standardized here for tabular presentation purposes.

The ultimate list of problem and opportunity categories and scores are presented in the following table, followed by the list of candidate "most important" Tweets given to the raters to sort.

**Table 4-14**  
**Resultant problem and opportunity categories**

**Problems:**

1. Problems with web professionalism: 1/6
2. Problems with store-front professionalism: 1/6
3. Problems with staff professionalism: 1/6
4. Problems with staff knowledge/training: 2/6
5. Pricing/promotional problems: 1/6

**Opportunities:**

1. Opportunity to retain customers by continuing certain practices: 1/6
2. Opportunity to capitalize on trends and promote particular styles: 2/6
3. Opportunity to retain customers by responding to requests: 1/6
4. Opportunity to continue effective marketing/increase brand awareness: 1/6
5. Opportunity to reinforce professionalism in particular stores: 1/6

**Table 4-15**  
**Problem and opportunity Tweets presented to category sorters**

**PROBLEMS:**

1. the #coach sunglasses aren't even listed on the #sunglasshut website yet. Eeek
2. RT @cubanmonkey69 Never new 'It was so hard' to merchandise a #sunglasshut kiosk! LMAO they have been @ it for 3 hrs lol. . #TWSS
3. Closed? In the afternoon? (at @sunglasshut) path.com/p/1wvwnQ
4. Never new it was so hard to merchandise a #sunglasshut kiosk! LMAO they have been @ it for 3 hrs lol

5. Pretty disappointed with the customer service I have received from @sunglasshut!
6. #Sunglasshut needs to get some coupons out asap before I have to get some cheaper glasses for my groomsmen at lenscrafters :(
7. I wish @sunglasshut would train their staff. Went in to get the artist series glasses but no one knew what I was talking about or cared!

### **OPPORTUNITIES:**

1. Loving @sunglasshut right now. Ran in to get a new pair of aviators and they exchanged my wayfarers that tobex destroyed also. #happysaturday
2. Love walking into sunglasshut and they clean my raybans lmfao
3. Okay, so I will die if I can't get the PINK glitter Noir @miumiustlye sunnies. @sunglasshut are you getting any in Australia? @luxottica\_au
4. @sunglasshut can you help me out with a marketing contact email for a travel contest I'm producing in Asia. #dreamjob
5. Props to @sunglasshut for giving me customized sunglasses!! #ThanksDiana #HappyBirthdayToMe
6. #SunglassHut said they can fix my #rayban for free. Awesome :)
7. :D Worth it!!!! RT @MrMenziN: Tjo RT Lady\_Crunk: Waiting 2hours outside SunglassHut's JandB Met tent for ... m.tmi.me/olydD
8. @MadonnaMDNAday my most prized possession. If only u knew what I went through 2 get this from sunglasshut img.ly/ggmH #askmadonna
9. @sunglasshut thanks for having an awesome return policy on new shades. #absolutequality
10. @sunglasshut seriously awesome crusty service at 29th street in boulder. Thanks Dave + Evan!!

- 11.@sunglasshut great customer service at Stonestown Galleria, SF, replaced defective Maui Jims, \$300+ sunglasses, no questions asked.
- 12.@LisaCouturing @persoleyewear @sunglasshut those are the ones! Foldables are a huge trend in sunnies!
- 13.Proper eyewear is another component to practicing #safesun. We like this pair from @sunglasshut for the ladies. [bit.ly/J34RGj](http://bit.ly/J34RGj)
- 14.RT @thecoolhunter: Clever way to promote sunglasses [pic.twitter.com/ayLfBkGf@sunglasshut\\_sa](http://pic.twitter.com/ayLfBkGf@sunglasshut_sa)
- 15.Love the sexy, new advertising campaign for @StellaMcCartney's new eco range at @sunglasshut. Looks like [grinstagram.com/p/JjU8ahtuRq/](http://grinstagram.com/p/JjU8ahtuRq/)
- 16.#Sunglasshut Brooklyn is on point. Thank u guys for hooking me out with the stylish #Rayban, gonna look sueve this Saturday.
- 17.I need a connect for shades ( dont mind a little discount) any followers work at #SunglassHut ? #FreeGstacksTho
- 18.@sunglasshut i have a problem my on my ship please contact me i sent a email but didnt fix it yet.
- 19.So adorable @SunglassHutSA: Rocking Ray-Ban Kids sunglasses-Mason Disick son of @KourtneyKardash #Rayban #Sunglasshut [yfrog.com/obz4bqbbj](http://yfrog.com/obz4bqbbj)
- 20.@couturing I think you need the foldable @persoleyewear glasses from @sunglasshut too! So cool and practical!
- 21.eraserhead77 I'd check out [sunglasshut.com](http://sunglasshut.com) they have pretty good deals sometimes :)
- 22.Go get you a cheesesteak RT @jaebee2fly: These are too many guilty pleasures in this airport..Charley's ...Ben and Jerry's ..SunglassHut ????

**Table 4-16**  
**Categories of problems devised by independent raters**

| Problems  |                    |                    |                    |   |                    |                  |
|---|--------------------|--------------------|--------------------|---|--------------------|------------------|
| Rater #1  |                    | Rater #2           |                    | Aggregate   |                    |                  |
| Category Specified                                      | Constituent Tweets | Category Specified | Constituent Tweets | Resultant Problem Category                            | Constituent Tweets | Proportion score |
| Website not current with merchandise                    | 1                  | Marketing problem  | 1, 2, 3            | Problems with professionalism—online or store-front * | 1, 2, 3            | 2/6              |
| Lack of store-front professionalism                     | 2, 3               |                    |                    |   |                    |                  |
| Generic or specific store complaints – to follow up on  | 4, 5, 6            | Staffing problem   | 4                  | Problems with staff professionalism and knowledge **  | 4, 5, 6            | 2/6              |
|   |                    | Training problem   | 5, 6               |   |                    |                  |
| Lack of promos – potential customer shift to competitor | 7                  | Pricing problem    | 7                  | Pricing/promotional problems                          | 7                  | 2/6              |

\* For sake of granularity in tallying accuracy, problems with online professionalism and store-front professionalism split into two categories of 1/6 proportion each.

\*\* For a similar reason, problems with staff professionalism and knowledge were split into problems with staff professionalism and problems with staff knowledge/training, also with 1/6 proportion score each.

**Table 4-17**  
**Categories of opportunities devised by independent raters**

| Opportunities   |                              |  |                                |   |                                |                    |
|---|------------------------------|--|--------------------------------|---|--------------------------------|--------------------|
| Rater #1  |                              | Rater #2   |                                | Aggregate   |                                |                    |
| Category Specified  | Constituent Tweets           | Category Specified   | Constituent Tweets             | Resultant Opportunity Category  | Constituent Tweets             | Proportion Score   |
| Return policy/service offered/campaign pleases customers                            | 8, 9, 10, 11, 12, 13, 17, 18 | Opportunity to retain a customer                           | 8, 9, 10, 12, 13, 15, 17       | Opportunity to retain customers by continuing certain practices   | 8, 9, 10, 11, 12, 13, 17       | $7/26 \approx 1/6$ |
| Trends in sunglasses to be aware of (via blogging customers, etc – with wide reach) | 20, 21, 22, 23, 24, 25       | Opportunity to sell more sunglasses of a certain type      | 11, 14, 18, 20, 21, 23, 24, 25 | Opportunity to capitalize on trends and promote particular styles ***   | 14, 18, 20, 21, 23, 24, 25, 27 | $8/26 \approx 2/6$ |
| Questions/requests from customers that can be followed up on                        | 14, 15, 16, 19               | Opportunity to partner with another person or organization | 16                             | Opportunity to retain customers by responding to requests   | 14, 15, 16, 19                 | $4/26 \approx 1/6$ |
|   |                              | Opportunity to increase brand awareness                    | 19, 22, 26, 29                 | Opportunity to continue effective marketing/increase brand awareness  | 22, 26, 28, 29                 | $4/26 \approx 1/6$ |
| Specific store CS compliments – follow up on (all cross-categorized)                | 10, 17, 18                   |  |                                | Opportunity to reinforce professionalism in particular stores   | 10, 17, 18                     | $3/26 \approx 1/6$ |
| Generic catch-all: C2C mentions, etc.   | 26, 27, 28, 29               | Unclear (uncategorized) Tweet                              | 27, 28                         | *** To enable more granular scoring, category split into 1) capitalize trends and 2) promote particular trends, each with 1/6 score proportion. |                                |                    |

## Appendix I – Informational Letter Emailed to Prospective Participants

April 26, 2012

Dear Students:

I am a graduate student under the direction of Dr. Rick Watson in the Department of MIS at The University of Georgia. I invite you to participate in a research study entitled “Automated Analysis of Highly Unstructured Data: An Empirical Investigation into Decision-Making Usefulness.” The purpose of this study is test different social media analytics interfaces.

For legal reasons, participants must be 18 years of age or older

**Your participation will involve viewing a stream of Tweets and identifying important problems and opportunities in the data, and should only take about 30 minutes of your time.** In order to make this study valid, some information about the study will be withheld until its completion. Your involvement in the study is voluntary, and you may choose not to participate or to stop at any time without penalty or loss of benefits to which you are otherwise entitled. The data you submit will be anonymous and not marked with any information to identify the participants. The results of the research study may be published, but your name will not be used. In fact, the published results will be presented in summary form only. Your identity will not be associated with your responses in any published format. Because there will be no way to identify you based on your submissions, you will be unable to withdraw your submission after you are debriefed (or given previously undisclosed information about the research project following completion of your participation).

The findings from this project may provide information on the benefits of timely, automated, machine-learning approaches to text mining of social media-generated data. There are no known risks or discomforts associated with this research. **To compensate you for your time, you will be eligible to receive extra credit in an MIS course.** An alternate assignment will be available during the same experiment signup times to those unable or unwilling to participate in the experiment.

Multiple time slots for this experiment will be open throughout the day of May 1, 2012 (Reading Day), in Caldwell 305—I will be in contact via email with a url for a signup sheet this week. If you have any questions about this research project, please feel free to

call me, Keri Larson, at (706) 542-6999 or send an e-mail to [kmlarson@uga.edu](mailto:kmlarson@uga.edu). Questions or concerns about your rights as a research participant should be directed to The Chairperson, University of Georgia Institutional Review Board, 629 Boyd GSRC, Athens, Georgia 30602; telephone (706) 542-3199; email address [irb@uga.edu](mailto:irb@uga.edu).

By signing up for the study and submitting the questionnaire via web browser, you are agreeing to participate in the above described research project.

Thank you for your consideration! Please keep this letter for your records.

Sincerely,

A handwritten signature in blue ink, appearing to be 'Keri Larson', with a stylized, flowing script.

Keri Larson  
Department of MIS



## Appendix J – Instructions Read to Participants before Experiment

Your assignment –

You are an intern working for Sunglass Hut, and the company has recently purchased a Twitter analytics software system. Your boss has just asked you to come prepared to a meeting in 45 minutes with a report on **problems** and **opportunities** that Sunglass Hut management should know about in order to possibly act on, based on what your customers are saying about the company via Twitter.

**Problems** could be something a customer complaint, for example, while an **opportunity** could be something like a customer request that an intern (like you!) could address, or praise for some type of customer service policy is working well and should be continued. These are just a couple of possible examples and should not limit or bias the problems and opportunities you discuss in your assessment.

To prepare, you access the Twitter analytics software which shows a live stream of Tweets mentioning your company, Sunglass Hut. To help make sense of the large number of Tweets, the software presents a selection of these in the “Decision-Assistance Panel.” This panel can help you narrow down the information to present to your boss, which he will pass on to the appropriate channel. After reviewing the data at your disposal, you will indicate to me in the online questionnaire the number of problems you identified, and then describe these problems in your own words. Then you will tell me how confident you are that you’ve identified all the important problems for Sunglass Hut. You will repeat these three steps for opportunities you identify in the data. Finally, and remember your experimental results are completely anonymous, please indicate your GPA and gender so we can determine whether or not these two things make any difference in the results.

The purpose of the experiment –

The purpose of this experiment is to determine the *best way* to select useful Tweets to appear in the Decision-Assistance Panel in order to inform company management what important things they can learn from their customers on social media.

Course credit –

Please write your name below to indicate that you've put forth your best effort in this experiment, and also write the course number and professor name that you'd like to receive extra credit for and return this sheet to me.

Thank you so much for your help!

Keri

## Appendix K – Debriefing Questionnaire

The debriefing questionnaire contained items used to measure dependent variables indicative of a social media analytics system's ability to support organizational decisions making through the identification of problems and opportunities within customer communications. For each decision-assistance panel condition (no analysis, sentiment-based analysis, or NLP-based analysis), we measured the effect of using the decision-assistance panel on the following outcomes for each subject:

1. Number of problems accurately identified
2. Number of opportunities accurately identified
3. Accuracy of problem assessment
4. Accuracy of problem assessment
5. Confidence in problem detection
6. Confidence in opportunity detection

-----  
(Start of questionnaire contents):

### **PROBLEMS**

1. How many problems regarding the "Sunglass Hut" brand did you detect from the Twitter stream?
2. Please describe the problems you identified (in your own words):
3. On a scale of 1 (lowest) to 5 (highest), how confident are you that you were able to spot all the important problems for "Sunglass Hut"?

### **OPPORTUNITIES**

1. How many opportunities regarding the "Sunglass Hut" brand did you detect from the Twitter stream?
2. Please describe the opportunities you identified:
3. On a scale of 1 (lowest) to 5 (highest), how confident are you that you were able to spot all the important opportunities for "Sunglass Hut"?

## Appendix L – Detailed Analysis of Experimental Results

This appendix:

1. Presents the detailed results of the statistical analysis of the data collected during the experiment.
2. Outlines the requirements of the ANOVA model used in the analysis of the experimental data.
3. Presents and discusses the results of the statistical tests used to ensure fulfilment of requirements.

### I. Statistical Methods

The major statistical techniques used to analyze the experimental data are analysis of variance with subsequent linear contrasts. First, a MANOVA ensures there are significant effects present in the data; subsequent ANOVA models detect those dependent variables for which significant differences exist, and then linear contrasts are used to test hypotheses related to observed significant effects.

#### a. Analysis of Variance

It is customary to begin the analysis of a single-factor study by determining whether or not the factor level means  $\mu_i$  are equal (Neter, Wasserman, and Kutner 1990: 546). The *factor effects* ANOVA model for a fixed effect single-factor study is:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where  $\sum \tau_j = 0$  and where:

$Y_{ij}$  is the value of the response variable in the  $j$ th trial for the  $i$ th factor level or treatment

$\mu$  is the overall mean (a constant component common to all observations)

$\tau_i$  is the effect of the  $i$ th factor level (a constant for each factor level)

$\varepsilon_{ij}$  are independent  $N(0, \sigma^2)$

$i = 1, \dots, r; j = 1, \dots, n_i$

The  $Y_{ij}$  observations are assumed to be independent and approximately normal with constant variance.

The test for equality of factor means is expressed in terms of the factor effects  $\tau_i$ .

The alternatives we wish to consider are:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_r = 0$$

$$H_a: \text{not all } \tau_i \text{ are equal}$$

The test statistic used for choosing between these alternatives is

$$F^* = \frac{MSTR}{MSE}$$

where MSTR is the treatment mean square, or the treatment sum of squares divided by its associated degrees of freedom, and MSE is the error mean square, or the error sum of squares divided by its associated degrees of freedom. Because  $F^*$  is distributed as  $F(r - 1, n_T - 1)$  when  $H_0$  holds and large values of  $F^*$  lead to the conclusion  $H_a$ , the decision rule to control the level of significance at  $\alpha$  is:

$$\text{If } F^* \leq F(1 - \alpha; r - 1, n_T - r), \text{ conclude } H_0$$

$$\text{If } F^* > F(1 - \alpha; r - 1, n_T - r), \text{ conclude } H_a$$

where  $F(1 - \alpha; r - 1, n_T - r)$  is  $(1 - \alpha)100$  percentile of the appropriate F distribution.

The ANOVA model is used for an omnibus test to detect significant differences in dependent variable means across treatments at the 5% level of significance. The aov procedure of the statistical package R is used to compute the ANOVA model. Based on the omnibus test, we conclude that significant differences exist for five variables: number of problems identified, accuracy of problem identification, confidence in problem assessment, and accuracy of opportunity assessment.

**Table 4-18**  
**Omnibus ANOVA F-test**

| Between-subjects source of variation                | Df | SS    | F                 | Pr>F      |
|---|----|-------|-------------------|-----------|
| Number of problems identified                       | 2  | 65.95 | 8.38              | 0.000495* |
| Accuracy of problem identification                  | 2  | 60.72 | 13.93             | 6.31e-06* |
| Confidence in problem identification/assessment     | 2  | 5.04  | 3.13              | 0.0491*   |
| Number of opportunities identified                  | 2  | 40.7  | 2.15              | 0.123     |
| Accuracy of opportunity identification              | 2  | 17.4  | ( $\chi^2$ ) 9.64 | 0.008076* |
| Confidence in opportunity identification/assessment | 2  | 0.15  | 0.12              | 0.888     |

Prior to testing whether the treatment means are the same for each dependent variable across groups, we conduct tests to establish the appropriateness of using the analysis of variance model by ensuring all model requirements are met.

## II. Requirements of the ANOVA model

The ANOVA model requires the following conditions:

1. Homogeneity of variances
2. Independent samples
3. Normality of error terms

### a. Homogeneity of Variance

The error terms  $\varepsilon_{ij}$  should have constant variance for all factor levels. When samples are not large, the appropriateness of this assumption can best be studied from residual plots against fitted values. In conjunction with this visual test, the Bartlett test (Neter, Wasserman, and Kutner 1990: 614) is also used to assess homogeneity of variance. The test statistic for deciding between:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2$$

$$H_a: \text{not all } \sigma_i^2 \text{ are equal}$$

is:

$$B = \frac{1}{C} \left[ (df_T) \log_e MSE - \sum_{i=1}^r (df_i) \log_e s_i^2 \right]$$

Where:

$$MSE = \frac{1}{df_T} \sum_{i=1}^r df_i s_i^2$$

$$\text{and } C = 1 + \frac{1}{3(r-1)} \left[ \left( \sum_{i=1}^r \frac{1}{df_i} \right) - \frac{1}{df_T} \right]$$

B is approximately distributed as  $\chi^2$  with  $r - 1$  degrees of freedom when  $H_0$  holds, so the decision rule for controlling Type I error at  $\alpha$  is:

If  $B \leq \chi^2 (1 - \alpha; r - 1)$ , conclude  $H_0$

If  $B > \chi^2 (1 - \alpha; r - 1)$ , conclude  $H_a$

where  $\chi^2 (1 - \alpha; r - 1)$  is the  $(1 - \alpha)100$  percentile of the  $\chi^2$  distribution with  $r - 1$  degrees of freedom. When the Bartlett test is used for single-factor ANOVA, we have:

$$df_i = n_i - 1 \text{ and } df_r = \sum_{i=1}^r (n_i - 1) = n_T - r.$$

At  $\alpha = 0.05$  and  $r = 3$  treatment groups, we require  $\chi^2 (0.95, 2)$ . From the  $\chi^2$  distribution table, we find  $\chi^2 (0.95, 2) = 5.99$ . Therefore, our decision rule is

If  $B \leq 5.99$ , conclude  $H_0$

If  $B > 5.99$ , conclude  $H_a$

**Table 4-19**  
**Bartlett's test for homogeneity of variance**

| Dependent variable                    | B statistic |
|---------------------------------------|-------------|
| Number of problems identified         | 0.4558      |
| Problem identification accuracy       | 0.798       |
| Problem identification confidence     | 6.8601*     |
| Number of opportunities identified    | 0.5125      |
| Opportunity identification accuracy   | 9.8037*     |
| Opportunity identification confidence | 3.7771      |
| * Significant at the 5% level         |             |



Bartlett's test, at the 5% level, indicates that the sample variances across problem identification confidence are not equal. Running the Bonferroni outlier test in R on the data indicates there are no studentized residuals with Bonferroni  $p < 0.05$ . However, treating the largest absolute studentized residual as an outlier and removing it from the dataset resolves the homogeneity problem. The same process does not resolve the homogeneity problem for opportunity identification accuracy. Instead we use a non-parametric test for mean differences.

### **b. Independent Samples**

Samples should be drawn from independent populations, which is typically achieved by random assignment of a treatment to an experimental subject. In this experiment, sessions were run hourly over a nine-hour period of time. Each time period accommodated approximately equal numbers of treatments based on the number of students signed up for each time slot. On each hour, experimental workstations were set up in order, alternating from treatment 1 to treatment 2 to treatment 3; as students entered the lab in no particular order, each sat at the next available workstation. Monitors were off on all machines prior to the start of each session, the effect of which was random assignment of each student to a treatment to ensure independence of samples.

### **c. Normality of Error Terms**

The normality of the error terms can be studied graphically as normal probability plots of the residuals (Neter, Wasserman, and Kutner 1990: 125). Further, when sample sizes are not large, all residuals  $e_{ij}$  for all treatments can be combined, as long as there

are no major departures from constant error variances across groups. Our plots of residuals do not indicate any serious departures from normality, as the patterns of points are all reasonably linear. Furthermore, Monte Carlo simulation-based studies demonstrate robustness of the one-way ANOVA test against normality violations (Schmider et al. 2010).

### **III. Power Analyses**

Statistical power is the probability of correctly rejecting the null hypothesis when it is false in our sample. Based on suggested effect sizes of 0.1, 0.25, and 0.4 for small, medium, and large effect sizes (Cohen 1992), we estimate the required number of participants for a one-way ANOVA with three factor levels to be approximately 63 for a large effect size (21 per cell). Because this research domain is new and we do not have a strong sense of the magnitude of the underlying phenomenon, any estimate of effect size clearly gives us only a very rough estimate for the number of subjects needed. If the effect is medium, we need 156 participants (52 per cell), and for a small effect size we need almost 1000. However, we are limited by the number of students who sign up for the experiment, so we rely on post hoc power analyses to give us slightly more insight into the possible power of our ANOVA test.

Post hoc analysis suggests that for a medium effect size, we have about a 50 percent chance of detecting significant differences, although for a large effect size our power approaches 89 percent.

Table 4-20 shows the expected power based on cell size for this particular experimental design.

**Table 4-20**  
**Expected power of ANOVA tests according to cell size**

|  |                  | Cell size |        |        |        |        |
|--|------------------|-----------|--------|--------|--------|--------|
|  |                  | 25        | 26     | 27     | 28     | 29     |
| Effect size<br>(suggested by<br>Cohen) | 0.1<br>(small)   | 0.1082    | 0.1107 | 0.1133 | 0.1159 | 0.1186 |
|  | 0.25<br>(medium) | 0.4600    | 0.4765 | 0.4927 | 0.5086 | 0.5243 |
|  | 0.4<br>(large)   | 0.8689    | 0.8828 | 0.8954 | 0.9068 | 0.9171 |

#### IV. Statistical Results and Tests of Hypotheses

The statistical findings for each of the six dependent variables are presented in this section.

##### a. Problem Number

The number of problems identified from the raw data with the aid of the decision-assistance panel is measured on a scale of 1 to 10+. Table 4-21 summarizes the number of problems identified by each group:

**Table 4-21**  
**Number of problems identified:**  
**mean score**  
**(standard deviation and cell size)**

| Decision-Assistance Panel Approach |                            |                            |                            |
|------------------------------------|----------------------------|----------------------------|----------------------------|
| Sentiment<br>Analysis              | NLP                        | Manual                     | Total                      |
| 3.71<br>(1.94          28)         | 5.86<br>(2.18          29) | 4.69<br>(1.78          26) | 4.77<br>(2.15          83) |

The ANOVA model demonstrates a statistically significant effect of decision-assistance panel approach on this dependent variable. The results of the ANOVA model for problem number are presented in Table 4-22:

**Table 4-22**  
**Number of problems**  
**analysis of variance**

| Variation Source          | Df | Sum Sq | F    | Pr(>F)       |
|---------------------------|----|--------|------|--------------|
| Decision-assistance panel | 2  | 65.95  | 8.38 | 0.000495 *** |
| Residuals                 | 80 | 314.70 |      |              |

### Tests of hypotheses related to problem number

We conduct linear contrasts to detect the location of significant differences in means. The results for problem number are presented in Table 4-23:

**Table 4-23**  
**P-values for paired comparisons of problem number means**

|            | diff      | p adj         |
|------------|-----------|---------------|
| SA-NLP     | 2.147783  | 8.591382e-05* |
| SA-manual  | 0.978022  | 0.142948      |
| NLP-manual | -1.169761 | 0.01111268*   |

**P1a: Number of detected critical problems (sentiment analysis)**

Individuals assisted by sentiment-based machine analysis of social media content will detect a greater number of key problems than individuals with no machine assistance.

***Not supported***

**P4a: Number of detected critical problems (NLP)**

Individuals assisted by natural language processing-based machine analysis of social media content will detect a greater number of key problems than individuals assisted by sentiment-based machine analysis or with no machine assistance.

***Supported***

There is no significant difference detected in the number of problems identified by the group using the sentiment-analysis-based aid (SA) versus the control group (manual). However, the NLP group outperformed the SA and manual groups to a significant degree, supporting hypothesis H4a. It is interesting to note that the group relying on a sentiment-analysis-based decision-assistance panel was able to identify fewer problems, on average, than the group processing the raw data manually.

**b. Problem Accuracy**

The accuracy of subjects' problem identification was numerically scored according to a weighted calculus resulting in a scale ranging from extremely inaccurate

(1) to extremely accurate (7). Table 4-24 summarizes the accuracy of problem assessments according to group:

**Table 4-24**  
**Accuracy of problem identification:**  
**mean score**  
**(standard deviation and cell size)**

| <b>Decision-Assistance Panel Approach</b> |                              |                              |                              |
|---|------------------------------|------------------------------|------------------------------|
| <b>Sentiment Analysis</b>                 | <b>NLP</b>                   | <b>Manual</b>                | <b>Total</b>                 |
| 2.41<br>(1.45            29)              | 4.38<br>(1.61            29) | 2.88<br>(1.34            26) | 3.24<br>(1.69            84) |

The ANOVA model demonstrates a statistically significant effect of decision-assistance panel approach on this dependent variable. The results of the ANOVA model for problem accuracy are presented in Table 4-25:

**Table 4-25**  
**Accuracy of problem identification**  
**analysis of variance**

| Variation Source          | Df | Sum Sq | F     | Pr(>F)       |
|---------------------------|----|--------|-------|--------------|
| Decision-assistance panel | 2  | 60.72  | 13.93 | 6.31e-06 *** |
| Residuals                 | 81 | 176.52 |       |              |

### **Tests of hypotheses related to problem accuracy**

The results of the linear contrast for problem accuracy are presented in Table 4-26.

**Table 4-26**  
**P-values for paired comparisons of accuracy of problem identification**

|            | diff       | p adj         |
|------------|------------|---------------|
| SA-NLP     | 1.9655172  | 3.64657e-06*  |
| SA-manual  | 0.4708223  | 0.4266818     |
| NLP-manual | -1.4946950 | 0.0001136953* |

**P2a: Accuracy of detected critical problems (sentiment analysis)**

Individuals assisted by sentiment-based machine analysis of social media content will more accurately detect key problems than individuals with no machine assistance.

***Not supported***

**P5a: Accuracy of detected critical problems (NLP)**

Individuals assisted by natural language processing-based machine analysis of social media content will more accurately detect key problems than individuals assisted by sentiment-based machine analysis or with no machine assistance.

***Supported***

**c. Problem Confidence**

The degree of confidence that all important problems were detected is measured on a five-point Likert scale, with the following anchors:

- 1: “*Not at all* confident”
- 2: “*Slightly* confident”
- 3: “*Somewhat* confident”
- 4: “*Very* confident”
- 5: “*Extremely* confident”

Table 4-27 summarizes the confidence in problem detection of each group:

**Table 4-27**  
**Confidence in problem identification:**  
**mean score**  
**(standard deviation and cell size)**

| <b>Decision-Assistance Panel Approach</b> |                                    |                                    |                                    |
|---|------------------------------------|------------------------------------|------------------------------------|
| <b>Sentiment Analysis</b>                 | <b>NLP</b>                         | <b>Manual</b>                      | <b>Total</b>                       |
| 3.1<br>(0.94                  29)         | 3.66<br>(0.67                  29) | 3.19<br>(1.06                  26) | 3.23<br>(0.92                  84) |

The ANOVA model demonstrates a statistically significant effect of decision-assistance panel approach on this dependent variable. The results of the ANOVA model for problem identification confidence are presented Table 4-28:

**Table 4-28**  
**Confidence in problem identification**  
**analysis of variance**

| Variation Source          | Df | Sum Sq | F     | Pr(>F)   |
|---------------------------|----|--------|-------|----------|
| Decision-assistance panel | 2  | 5.04   | 3.128 | 0.0491 * |
| Residuals                 | 81 | 65.28  |       |          |



### **Tests of hypotheses related to problem confidence**

The results of the linear contrasts for problem confidence are presented in **Table**

**4-29:**

**Table 4-29**  
**P-values for paired comparisons of problem confidence means**

|            | diff        | p adj       |
|------------|-------------|-------------|
| SA-NLP     | 0.55172414  | 0.02650408* |
| SA-manual  | 0.08885942  | 0.9162916   |
| NLP-manual | -0.46286472 | 0.02383028* |

#### **P3a: Confidence in critical problem detection (sentiment analysis)**

Individuals assisted by sentiment-based machine analysis of social media content will detect a greater number of key problems than individuals with no machine assistance.

***Not supported***

#### **P6a: Confidence in critical problem detection (NLP)**

Individuals assisted by natural language processing-based machine analysis of social media content will have greater confidence that they detected key problems than individuals assisted by sentiment-based machine analysis or with no machine assistance.

***Supported***

#### **d. Opportunity Number**

The number of opportunities identified from the raw data with the aid of the decision-assistance panel is measured on a scale of 1 to 10+. Table 4-30 summarizes the number of opportunities identified by each group:

**Table 4-30**  
**Number of opportunities identified:**  
**mean score**  
**(standard deviation and cell size)**

| <b>Decision-Assistance Panel Approach</b> |                              |                              |                              |
|---|------------------------------|------------------------------|------------------------------|
| <b>Sentiment Analysis</b>                 | <b>NLP</b>                   | <b>Manual</b>                | <b>Total</b>                 |
| 6.14<br>(3.27            28)              | 6.24<br>(2.86            29) | 7.72<br>(3.08            26) | 6.66<br>(3.12            83) |

The ANOVA model F-test does not indicate a statistically significant effect of decision-assistance panel approach on this dependent variable.

#### **Tests of hypotheses related to opportunity number**

##### **P1b: Number of detected critical opportunities (sentiment analysis)**

Individuals assisted by sentiment-based machine analysis of social media content will detect a greater number of key opportunities than individuals with no machine assistance.

***Not supported***

#### **P4b: Number of detected critical opportunities (NLP)**

Individuals assisted by natural language processing-based machine analysis of social media content will detect a greater number of key opportunities than individuals assisted by sentiment-based machine analysis or with no machine assistance.

***Not supported***

#### **e. Opportunity Accuracy**

The accuracy of subjects' opportunity identification was numerically scored according to a weighted calculus resulting in a scale ranging from extremely inaccurate (1) to extremely accurate (7). Table 4-31 summarizes the accuracy of opportunity identification according to group:

**Table 4-31**  
**Accuracy of opportunity identification:**  
**mean score**  
**(standard deviation and cell size)**

| <b>Decision-Assistance Panel Approach</b> |                   |                   |                    |
|---|-------------------|-------------------|--------------------|
| <b>Sentiment Analysis</b>                 | <b>NLP</b>        | <b>Manual</b>     | <b>Total</b>       |
| 2.1<br>(0.82 29)                          | 3.17<br>(1.42 29) | 2.42<br>(1.47 26) | 2.57<br>(21.33 84) |

The ANOVA model demonstrates a statistically significant effect of decision-assistance panel approach on this dependent variable. The results of the ANOVA model for opportunity accuracy are presented in Table 4-32.

**Table 4-32**  
**Accuracy of opportunity identification**  
**analysis of variance**

| Variation Source          | Df | Sum Sq | F     | Pr(>F)     |
|---------------------------|----|--------|-------|------------|
| Decision-assistance panel | 2  | 17.4   | 8.699 | 0.00599 ** |
| Residuals                 | 81 | 129.2  |       |            |

**Tests of hypotheses related to opportunity accuracy**

The results of the simultaneous pairwise comparisons for opportunity accuracy are presented in Table 4-33:

**Table 4-33**  
**P-values for paired comparisons of accuracy of opportunity identification**

|            | diff      | p adj         |
|------------|-----------|---------------|
| SA-NLP     | 1.0689655 | 0.0006156996* |
| SA-manual  | 0.3196286 | 0.6679747     |
| NLP-manual | 0.7493369 | 0.003121333*  |

**P2b: Accuracy of detected critical opportunities (sentiment analysis)**

Individuals assisted by sentiment-based machine analysis of social media content will more accurately detect key opportunities than individuals with no machine assistance.

***Not supported***

### **P5b: Accuracy of detected critical opportunities (NLP)**

Individuals assisted by natural language processing-based machine analysis of social media content will more accurately detect key opportunities than individuals assisted by sentiment-based machine analysis or with no machine assistance.

***Supported***

### **f. Opportunity Confidence**

The degree of confidence that all important opportunities were detected was measured on a five-point Likert scale, with the following anchors:

- 1: “*Not at all* confident”
- 2: “*Slightly* confident”
- 3: “*Somewhat* confident”
- 4: “*Very* confident”
- 5: “*Extremely* confident”

Table 4-34 summarizes the confidence in opportunity detection of each group:

**Table 4-34**  
**Confidence in opportunity detection:**  
**mean score**  
**(standard deviation and cell size)**

| <b>Decision-Assistance Panel Approach</b> |                              |                              |                              |
|---|------------------------------|------------------------------|------------------------------|
| <b>Sentiment Analysis</b>                 | <b>NLP</b>                   | <b>Manual</b>                | <b>Total</b>                 |
| 3.59<br>(0.73            29)              | 3.55<br>(0.69            29) | 3.65<br>(0.94            26) | 3.60<br>(0.78            84) |

The ANOVA model F-test does not indicate a statistically significant effect of decision-assistance panel approach on this dependent variable.

### **Tests of hypotheses related to opportunity confidence**

#### **P3b: Confidence in critical opportunity detection (sentiment analysis)**

Individuals assisted by sentiment-based machine analysis of social media content will detect a greater number of key opportunities than individuals with no machine assistance.

***Not supported***

#### **P6b: Confidence in critical opportunity detection (NLP)**

Individuals assisted by natural language processing-based machine analysis of social media content will have greater confidence that they detected key opportunities than individuals assisted by sentiment-based machine analysis or with no machine assistance.

***Not supported***

## 4.8 References

- Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker, J. F. 2010. "Detecting Fake Websites: The Contribution of Statistical Learning Theory," *MIS Quarterly* (34:3), pp. 1-28.
- Abdel-khalek, A. M. 2006. "Measuring Happiness with a Single-Item Scale," *Social Behavior and Personality* (34:71962), pp. 139-150.
- Agichtein, E., Castillo, C., Donato, D., Gionis, A., and Mishne, G. 2008. "Finding High-Quality Content in Social Media," in *Proceedings of the international conference on Web search and web data mining* ACM, pp. 183-193.
- Anastasiou, D., and Gupta, R. 2011. "Comparison of crowdsourcing translation with Machine Translation," *Journal of Information Science* (37:6), pp. 637-659.
- Arazy, O., and Woo, C. 2007. "Enhancing Information Retrieval Through Statistical Natural Language Processing: A Study of Collocation Indexing," *MIS Quarterly* (31:3), pp. 525-546.
- Baker, M. 1996. "Corpus-based translation studies: The challenges that lie ahead," in *Terminology, Lsp and Translation: Studies in Language Engineering in Honour of Juan C. Sager (Benjamins Translation Library)*, H. Somers (ed.), (18th ed, )John Benjamins Pub Co., pp. 175-186.
- Bergkvist, L., and Rossiter, J. R. 2007. "The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs," *Journal of Marketing Research* (44:2), pp. 175-184.
- Buchanan, M. 2001. "Mind the Pseudogap," *Nature* (409:4 January 2001), pp. 8-11.
- Bunescu, R., and Mooney, R. 2007. "Extracting Relations from Text: From Word Sequences to Dependency Paths," in *Natural Language Processing and Text Mining*, A. Kao and S. R. Poteet (eds.), Springer, pp. 29-44.
- Carter, K. C. 1985. "Ignaz Semmelweis, Carl Mayrhofer, and the Rise of Germ Theory," *Medical History* (29:1), pp. 33-53.
- Chaudhuri, S., Dayal, U., and Narasayya, V. 2011. "An Overview of Business Intelligence Technology," *Communications of the ACM* (54:8), pp. 88.
- Cohen, A. M., & Hersh, W. R. (2005). "A Survey of Current Work in Biomedical Text Mining," *Briefings in Bioinformatics*, 6(1), 57-71.
- Cohen, J. 1992. "A Power Primer," *Psychological Bulletin* (112:1), pp. 155-159.
- Cristianini, N., and Shawe-Taylor, J. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, (1st ed.) Cambridge University Press, pp. 204.
- Culnan, M. J., Mchugh, P. J., and Zubillaga, J. I. 2010. "How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value," *MIS Quarterly Executive* (9:4), pp. 243-260.

- Daft, R. L., and Lengel, R. H. 1986. "Organizational Information Requirements, Media Richness and Structural Design," *Management Science* (32:5), pp. 554-571.
- Daft, R. L., and Weick, K. E. 1984. "Toward a Model of Organizations as Interpretation Systems," *The Academy of Management Review* (9:2), pp. 284-295.
- Delbecq, A. L., and Ven, A. H. V. de. 1971. "A Group Process Model for Problem Identification and Program Planning," *The Journal of Applied Behavioral Science* (7:4), pp. 466-492.
- Demetriou, G., and Kawalek, P. 2010. "Benefit-Driven Participation in Open Organizational Social Media Platforms: The Case of the SAP Community Network," *Issues in Information Systems* (XI:1), pp. 601-611.
- Eppler, M. J., and Mengis, J. 2004. "The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines," *The Information Society* (20:5), pp. 325-344.
- Eysenbach, G. (2008). "Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness," *Journal of Medical Internet Research*, 10(3), e22.
- Feinerer, I., Hornik, K., and Meyer, D. 2008. "Text mining infrastructure in R," *Journal of Statistical Software* (25:5), pp. 1-54.
- Feldman, R., and Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, pp. 422.
- Fox, B. A. 1986. "Local patterns and general principles in cognitive processes: Anaphora in written and conversational English," *Text* (6:1), pp. 25-51.
- Freedman, B. 1987. "Equipoise and the Ethics of Clinical Research," *New England Journal of Medicine* (317:3), pp. 141-145.
- Fury, W., Batliwalla, F., Gregersen, P. K., and Li, Wentian. 2006. "Overlapping Probabilities of Top Ranking Gene Lists, Hypergeometric Distribution, and Stringency of Gene Selection Criterion," *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference* (1:m), pp. 5531-4.
- Galbraith, J. R. 1974. "Organization Design: An Information Processing View," *Interfaces* (4:3), pp. 28-36.
- González-ibáñez, R., and Wacholder, N. 2011. "Identifying Sarcasm in Twitter : A Closer Look," *Computational Linguistics* (2010), pp. 581-586.
- Grinev, M., Grineva, M., Hentschel, M., and Kossman, D. 2011. "Analytics for the Real-Time Web," in *Proceedings of the VLDB Endowment*, pp. 1391-1394.
- Harmer, C. J., Goodwin, G. M., and Cowen, P. J. 2009. "Why Do Antidepressants Take So Long to Work? A Cognitive Neuropsychological Model of Antidepressant Drug Action," *The British Journal of Psychiatry* (195:2), pp. 102-8.



- Hearst, M. A. 1999. "Untangling Text Data Mining," in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* College Park, MD: Association for Computational Linguistics, pp. 3-10.
- Hoffman, D. L., and Fodor, M. 2010. "Can You Measure the ROI of Your Social Media Marketing?," *MIT Sloan Management Review* (52:1).
- Hotho, A., Andreas, N., and Paaß, G. 2005. "A Brief Survey of Text Mining," *Machine Learning* (20:1), pp. 19-62.
- Houben, G., Lenie, K., and Vanhoof, K. 1999. "A Knowledge-based SWOT-analysis System as an Instrument for Strategic Planning in Small and Medium Sized Enterprises," *Decision Support Systems* (26:2), pp. 125-135.
- Hu, M., and Liu, B. 2004. "Mining and Summarizing Customer Reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168-177.
- Huber, G. P., and McDaniel, R. R. 1986. "The Decision-Making Paradigm of Organizational Design," *Management Science* (32:5), pp. 572-589.
- Hutchins, W. J. 1995. "Machine Translation: A Brief History," in *Concise History of the Language Sciences: from the Sumerians to the Cognitivists*, E. F. K. Koerner and R. E. Asher (eds.), Pergamon Press, pp. 431-445.
- Hutchins, W. J. 2006. "Example-based machine translation: a review and commentary," *Machine Translation* (19:3-4), pp. 197-211.
- Häubl, G., and Trifts, V. 2000. "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science* (19:1).
- Inmon, W. H., and Nesavich, A. 2008. *Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence*, Prentice Hall, pp. 241.
- Kao, A., and Poteet, S. R. 2007. *Natural Language Processing and Text Mining*, A. Kao and S. R. Poteet (eds.), Springer, pp. 272.
- Kennedy, C., and Boguraev, B. 1996. "Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser," in *Proceedings of the 16th conference on Computational linguistics*, pp. 113-118.
- Koehn, P. 2005. "Europarl: A Parallel Corpus for Statistical Machine Translation," in *MT Summit* (Vol. 11), pp. 79-86.
- Lappin, S., and Leass, H. J. 1994. "An Algorithm for Pronominal Anaphora Resolution," *Computational Linguistics* (20:4), pp. 535-561.
- Larson, K., and Watson, R. T. 2011. "The Value of Social Media: Toward Measuring Social Media Strategies," in *ICIS 2011 Proceedings* Shanghai, China, pp. 1-18.
- Leslie, M. 2003. "NetWatch," *Science* (301:5638), pp. 1299.

- Lindvall, M., Rus, I., and Sinha, S. S. 2003. "Software systems support for knowledge management," *Journal of Knowledge Management* (7:5), pp. 137-150.
- Lopez, A. 2008. "Statistical machine translation," *ACM Computing Surveys* (40:3), pp. 1-49.
- Maedche, A., and Staab, S. 2000. "Mining Ontologies from Text," in *Knowledge Engineering and Knowledge Management Methods, Models, and Tools: 12th international conference, EKAW Juan-les-Pins*: Springer, pp. 169-189.
- Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. 2007. "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *Journal of Artificial Intelligence Research* (30), pp. 457-500.
- McCallum, A., and Li, Wei. 2003. "Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons," in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL Association for Computational Linguistics*, pp. 188-191.
- McGrath, J. E. 1981. "Dilemmatics: The Study of Research Choices and Dilemmas," *American Behavioral Scientist* (25:2), pp. 179-210.
- Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information," *The Psychological Review* (63:2), pp. 81-97.
- Miner, G., Elder IV, J., Hill, T., Nisbet, R., Delen, D., and Fast, A. 2012. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Academic Press, pp. 1000.
- Neter, J., Wasserman, W., and Kutner, M. H. 1990. *Applied Linear Statistical Models*, (3rd ed.), pp. 1181.
- Null, G., Dean, C., Feldman, M., & Rasio, D. (2005). "Death by Medicine," *Journal of Orthomolecular Medicine*, 20(1), 21-34.
- Pang, B., and Lee, L. 2004. "A Sentimental Education : Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 271.
- Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval* (2:1-2), pp. 1-135.
- Pavlou, P. A., and Dimoka, A. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," *Information Systems Research* (17:4), pp. 392-414.
- Road, B., and Linguistics, C. 1993. "Issues in the Choice of a Source for Natural Language Generation," *Computational Linguistics*.
- Robins, R. W., Hendin, H. M., and Trzesniewski, K. H. 2001. "Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale," *Personality and Social Psychology Bulletin* (27:2), pp. 151-161.

- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., and Richardson, W. S. 1996. "Evidence Based Medicine: What it is and what it isn't," *British Medical Journal* (312:7023), pp. 71-72.
- Sang, E. F. T. K., and Meulder, F. De. 2003. "Introduction to the CoNLL-2003 Shared Task : Language-Independent Named Entity Recognition," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL* Association for Computational Linguistics, pp. 142-147.
- Schafer, W. R. 1999. "How Do Antidepressants Work? Prospects for Genetic Analysis of Drug Mechanisms," *Cell* (98:5), pp. 551-554.
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., and Bühner, M. 2010. "Is It Really Robust?," *European Journal of Research Methods for the Behavioral and Social Sciences* (6:4), pp. 147-151.
- Smalheiser, N. R., & Swanson, D. R. (1994). "Assessing a Gap in the Biomedical Literature: Magnesium Deficiency and Neurologic Disease," *Neuroscience Research Communications*, 15(1), 1-9.
- Spasic, I., Ananiadou, S., McNaught, J., & Kumar, A. (2005). "Text Mining and Ontologies in Biomedicine: Making Sense of Raw Text," *Briefings in Bioinformatics*, 6(3), 239-51.
- Speier, C., Valacich, J. S., and Vessey, I. 1999. "The influence of task interruption on individual decision making: An information overload perspective," *Decision Sciences* (30:2) Wiley Online Library, pp. 337-360.
- Sunstein, C. R. 2006. *Infotopia: How Many Minds Produce Knowledge*, Oxford University Press, pp. 288.
- Surowiecki, J. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, Doubleday, pp. 336.
- Swanson, D. R. (1986). "Fish oil, Raynaud's Syndrome, and Undiscovered Public Knowledge," *Perspectives in Biology and Medicine*, 30(1), 7-18.
- Tanabe, L., Scherf, U., Smith, L. H., Lee, J. K., Hunter, L., and Weinstein, J. N. 1999. "MedMiner: an Internet Text-mining Tool for Biomedical Information, with Application to Gene Expression Profiling," *BioTechniques* (27:6), pp. 1210-4, 1216-7.
- The Use of Medicines in the United States: Review of 2011*. 2012. Parsippany, NJ.
- Vapnik, V. N. 1999. "An Overview of Statistical Learning Theory," *IEEE Transactions on Neural Networks* (10:5), pp. 988-99.
- Varma, C. 2010. "High-temperature Superconductivity: Mind the Pseudogap," *Nature* (468:10 November 2010), pp. 184-185.

Wang, J. 2007. "Support vector machines based on K-means clustering for real-time business intelligence systems Xindong Wu Chengqi Zhang," *Intelligence* (1:1), pp. 54-64.

Watson, H. J., Wixom, B. H., Hoffer, J. A., Anderson-Lehman, R., and Reynolds, A. M. 2006. "Real-Time Business Intelligence: Best Practices at Continental Airlines," *Information Systems Management* (23:1), pp. 7-18.

Weeber, M., Vos, R., Klein, H., de Jong-van den Berg, L. T. W., Aronson, A., & Molema, G. (2003). "Generating Hypotheses by Discovering Implicit Associations in the Literature: A Case Report of a Search for New Potential Therapeutic Uses for Thalidomide," *Journal of the American Medical Informatics Association*, 10(3), 252-259. doi:10.1197/jamia.M1158.

Wilcox, A. B., and Hripcsak, G. 2003. "The Role of Domain Knowledge in Automating Medical Text Report Classification," *Journal of the American Medical Informatics Association* (10:4), pp. 330-339.

## CHAPTER 5

### CONCLUSION

#### **5.1 Discussion across Papers**

This three-paper dissertation offers a series of manuscripts that evolves from the conceptual delineation of social media events and processes of concern to firms, to the specification of a social media analytics design capable of exploiting valuable product- and brand-oriented consumer-generated text as it flows across social media settings. By starting from a granular understanding of prevalent social-mediated cross-interactions, firms can focus their efforts on managing and measuring the most critical components of the glut of unstructured data proliferating across social media platforms daily. We recognize three levels of measurement inherent in the assessment of social media exchange, ranging from the existing capabilities of counts and sentiment analysis to a deeper level of qualitative analysis capable of analyzing textual data on a large scale to derive useful insight into the impact of a given product, service, brand, or campaign. This latter ability to glean deep meaning from voluminous streams of social media-generated data is a proficiency yet to be fully established. Toward the goal of arriving at an instantiation of this proficiency, we address the complications associated with qualitative analysis of highly unstructured data lacking semantic or syntactical constraints through set of design principles. These principles indicate a class of social media measurement tools expected to positively affect organizational decision-making and confer the ability to competitively manage social media initiatives in an

environment characterized by extensive two-way communication and collaboration between and among firms and consumers. The overall practical objective of this dissertation is the extraction of actionable, accurate, useful knowledge from consumer social media interactions.

Synthesizing across all three studies, the most important discovery resulting from our investigation is the evidence that the analytics approach underlying the majority of social media monitoring systems currently in practice, sentiment analysis, does not convey an advantage to organizations interested in mining knowledge from their customers. This leaves organizations with the option of employing teams of analysts to slowly monitor fragments of customer-to-customer and customer-to-firm messages, which is likely to result in a biased and disjointed understanding of customer complaints, requests, needs, wants, and suggestions. Alternately, toward the goal of a more comprehensive and reliable understanding of these components, we experimentally demonstrate that natural language processing heralds substantial promise to firms concerned with more than just the extreme sentiments of their customers. Furthermore, NLP is capable of improving the effectiveness of sentiment analysis for better understanding consumer opinions, although we stress this is just one important component of the overall scope of knowledge extraction from social media chatter.

As a whole, this dissertation represents an important foray into social media research. Although an important concern for organizations given the explosive growth in the number of firms now interacting with customers through social media channels (Boyd and Ellison 2008), it is still a relatively new area of information systems investigation. As such, this dissertation is presented as a contribution to the theory and

practice of both social media and the realm of qualitative textual analytics. As the start of a program of research focused on the incorporation of NLP advances into the mining of highly unstructured text, we expect the scope of context to potentially expand beyond social media monitoring capabilities. A wide range of textual data exists as potential sources of unknown or obscured knowledge, one example being the enormous magnitude of medical texts so voluminous as to be beyond the range of any human to encompass the knowledge contained therein (Spasic et al. 2005). As such, the risk is great that connections among symptoms, diseases, causes, and treatments across thousands of studies will never be identified (Cohen and Hersh 2005). However, as machine algorithms become more and more sophisticated in their abilities to understand the human written word, the more expediently and effectively such knowledge can be detected and thus preserved, potentially supporting critical innovations in medicine. It is with an eye toward this type of ultimate accumulation of knowledge that this dissertation is submitted.

## **5.2 References**

- Boyd, D. M., and Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Cohen, A. M., and Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in bioinformatics*, 6(1), 57-71.
- Spasic, I., Ananiadou, S., McNaught, J., and Kumar, A. (2005). Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in bioinformatics*, 6(3), 239-51.