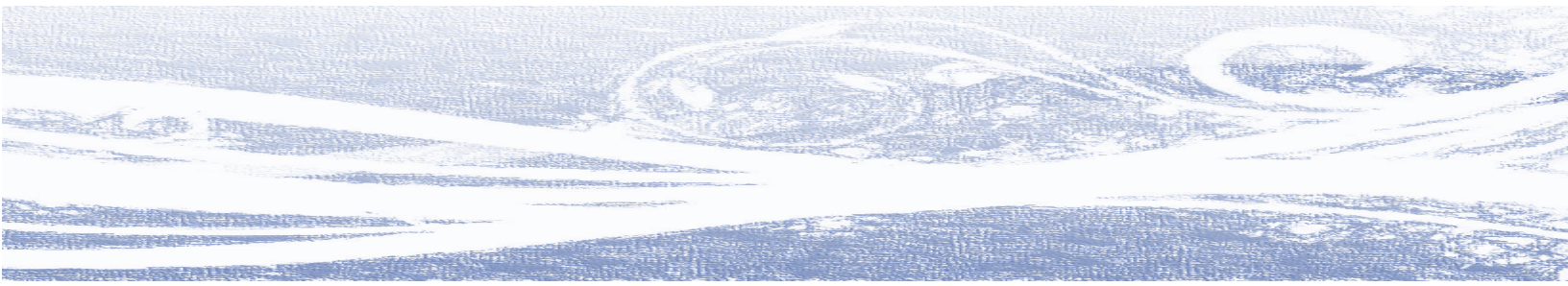
QRBD

QUARTERLY REVIEW OF BUSINESS DISCIPLINES

November 2020

Volume 7
Number 3



A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES
SPONSORED BY UNIVERSITY OF NORTH FLORIDA
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)

Blank inside front cover

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

FROM THE EDITORS

This issue of *Quarterly Review of Business Disciplines* begins with four papers on various aspects of Information Technology (IT), data usage, Artificial Intelligence (AI), and political blogging. The research of Liqiong Deng, University of West Georgia, delves into the paradoxical nature of user experience with IT. Cindi Smatt, Renée Pratt, Tamirat Abegaz, and Allie Dobbs, University of North Georgia, explore two popular and freely available data visualization tools D3.JS and Google Charts for retail customer data. Zixing Shen, New Mexico State University, focuses on the direct use of natural language processing (NLP) in the context of patient care. Hyun Jung Yun, Texas State University, revisits two traditional public opinion theories – the Spiral of Silence and the Bandwagon effect and analyzes textural-based interactive weblog dialogues in order to identify the importance of theoretical justifications of, or alterations to, these classical approaches in this new communication and information era.

The research of Jamie Birdwell, Florida Institute of Technology/Coast oem, studies women in the workplace and specifically in leadership roles. He questions whether the gender of the leader influences one's attitudes regarding leadership effectiveness and if so to what extent biases may disadvantage women in leadership roles.

Margaret A. Goralski, *Quinnipiac University,* Editor-in Chief
Charles A. Lubbers, *University of South Dakota,* Associate Editor

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

The *Quarterly Review of Business Disciplines (QRBD)* is published by the International Academy of Business Disciplines in February, May, August, and November.

Manuscript Guidelines/Comments. *QRBD* is a blind peer-reviewed journal that provides publication of articles in all areas of business and the social sciences that affect business. The Journal welcomes the submission of manuscripts that meet the general criteria of significance and business excellence. Manuscripts should address real-world phenomena that highlight research that is interesting and different – innovative papers that begin or continue a line of inquiry that integrate across disciplines, as well as, those that are disciplinary. The Journal is interested in papers that are constructive in nature and suggest how established theories or understandings of issues in business can be positively revised, adapted, or extended through new perspectives and insights. Manuscripts that do not make a theoretical contribution to business studies or that have no relevance to the domain of business should not be sent to *QRBD.* Submissions to *QRBD* must follow the journal's Style Guide for Authors, including length, formatting, and references. Poorly structured or written papers will be returned to the authors promptly. Manuscript length is approximately 16 – 20 single-spaced pages. Acceptance rate is 25-28%.

Description. The *Quarterly Review of Business Disciplines* is a quarterly publication of the International Academy of Business Disciplines (IABD); a worldwide, non-profit organization established to foster and promote education in all of the functional and support disciplines of business. The objectives of *QRBD* and IABD are to stimulate learning and understanding and to exchange information, ideas, and research studies from around the world. The Academy provides a unique global forum for professionals and faculty in business, communications, and other social science fields to discuss and publish papers of common interest that overlap career, political, and national boundaries. *QRBD* and IABD create an environment to advance learning, teaching, and research, and the practice of all functional areas of business. *Quarterly Review of Business Disciplines* is published to promote cutting edge research in all of the functional areas of business.

Submission Procedure. An electronic version of the manuscript must be submitted in MS Word to the Editor-in-Chief, Dr. Margaret A. Goralski at Margaret.Goralski@Quinnipiac.edu. Upon completion of a review by expert scholars who serve on the *QRBD* Editorial Review Board, the first author will be informed of acceptance or rejection of the paper within a one to two-month timeframe from the submission date. If the paper is accepted, the first author will receive a formal letter of acceptance along with the *QRBD* Style Guide for Authors. IABD members and authors who participate in the IABD annual conference are given first priority as a matter of courtesy. For additional information, please visit www.iabd.org.

Subscription. The annual subscription price for QRBD is US$100 plus postage and handling. Single issue price is $35 per issue plus postage and handling.

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

MISSION STATEMENT

The organization designated as the International Academy of Business Disciplines is a worldwide, non-profit organization, established to foster and promote education in all of the functional and support disciplines of business.

WWW.IABD.ORG

The Quarterly Review of Business Disciplines (QRBD) is listed in *Cabell's Directory of Publishing Opportunities.*

QRBD - QUARTERLY REVIEW OF BUSINESS DISCIPLINES

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES

VOLUME 7 NUMBER 3 NOVEMBER 2020

ISSN 2329-5163 (online)
ISSN 2334-0169 (print)

CONTENTS

ARTICLES

Technology Paradoxes, Regulatory Focus, and Trust 189
Liqiong Deng

Towards Effective Customer Data Visualization: Data-Driven Documents (D3.JS) vs. Google Charts 207
Cindi Smatt, Renée Pratt, Tamirat Abegaz, Allie Dobbs

Natural Language Processing (NLP) Applications in Patient Care: A Systematic Analysis 223
Zixing Shen

Divided and Crosscutting Blogosphere Political Dialogues: Revisiting the Spiral of Silence and the Bandwagon Effect 245
Hyun Jung Yun

Woman or Leader First? Gender Bias in the Perception of Female Leader Effectiveness 267
Jamie Birdwell

This page has been intentionally left blank.

TECHNOLOGY PARADOXES, REGULATORY FOCUS AND TRUST

Liqiong Deng, University of West Georgia

ABSTRACT

This research focuses on the paradoxical nature of user experience with information technology (IT) and draws on the Regulatory Focus (RF) theory and desirability-feasibility framework to examine how the diametric anchors of IT paradoxes (fulfills needs / creates needs, new / obsolete, control / chaos, engaging / disengaging, freedom / enslavement, assimilation / isolation, efficiency / inefficiency, competence / incompetence) influence users' perceptions of IT performance ambiguity, their cognitive and emotional trust in IT and continuance intention to use IT. According to the desirability-feasibility framework, this research classifies the IT paradoxes into two categories – the functional desirability paradoxes and the usage feasibility paradoxes. The RF theory distinguishes between two motivational orientations – promotion focus and prevention focus. Individuals with a promotion focus are driven by the need for attainment of positive outcomes, oriented toward the maximization of gains and advancement, and likely to adopt high-level construals focused on the functional desirability of IT. In contrast, individuals with a prevention focus are driven by the need for avoidance of negative outcomes, oriented toward the minimization of losses and safety, and likely to adopt low-level construals concerned with the usage feasibility of IT. Accordingly, this research proposes that individuals' regulatory foci determine the differential effects of the diametric anchors of IT functional desirability paradoxes and usage feasibility paradoxes on users' perceived IT performance ambiguity as well as their cognitive trust and emotional trust in IT, which consequently influence their continued IT use intention.

Keywords: Technology Paradoxes, Promotion Focus, Prevention Focus, Cognitive Trust, Emotional Trust, Continuance Intention

INTRODUCTION

Nowadays, as information technology (IT) has become increasingly prevalent in all aspects of people's daily life, people often find their daily experiences with IT to be ambivalent (Johnson, Bardhi & Dunn, 2008). On the one hand, people enjoy the benefits of new IT; but on the other hand, they are often confused with the complex features of IT. Users' experiences with IT may be paradoxical (Jarvenpaa & Lang, 2005), characterized by conflicting emotions where users are confronted with both positive and negative aspects of technology usage. The notion of technology paradox is not new. In social science, Winner (1994) argues that the same technology that creates positive feelings of intelligence and efficacy can also elicit feelings of stupidity and ineptitude (Winner, 1994). In marketing research, Mick and Fournier describe eight central technology paradoxes – control/chaos, freedom/enslavement, new/obsolete, competence/incompetence, efficiency/inefficiency, fulfills/creates needs, and engaging/disengaging (Mick & Fournier, 1998). In information systems (IS) research, Orlikowski (1991) and Chinn (2001) discuss the paradoxical nature of IT. Jarvenpaa and Lang propose eight paradoxes of mobile technology –

empowerment/enslavement, independence/dependence, fulfills/creates needs, competence/incompetence, planning/improvisation, engaging/disengaging, and public/private (Jarvenpaa & Lang, 2005).

IT paradoxes produce conflicting evaluations of IT, elicit mixed emotions (known as emotional ambivalence) toward IT (Mick & Fournier, 1998), and inhibit users' ability to fully evaluate the performance of IT, thus resulting in perceived performance ambiguity of IT and uncertainty concerning whether IT can be relied on in the future (Johnson, Bardhi & Dunn, 2008). In response to perceived uncertainty and performance ambiguity of IT, users' trust in IT will play a central role in determining users' continuance intention to use IT. Trust involves a leap of faith beyond what is supported by limited available information despite uncertainties or risks (Lewis & Weigert, 1985; Luhmann, 1979). As a result of IT paradoxes, users' inability to fully determine the performance of IT makes it necessary for them to draw on their trusting intention toward IT in assessing their continued IT usage intention (Johnson, Bardhi & Dunn, 2008).

While prior research has provided valuable insights into the impacts of technology paradoxes on users' emotional, cognitive and behavioral responses (Assigbetse, 2019; Chae & Yeum, 2010; Johnson, Bardhi & Dunn, 2008; Lee, 2016; Mick & Fournier, 1998; Zhuang, Hsu, Brewer & Xiao, 2013), however, no insight has been offered on the relative salience of different diametric anchors of technology paradoxes to users. Recognizing the complex characteristics of IT paradoxes, this paper attempts to explore how the effects of IT paradoxes on perceived IT performance ambiguity, IT trust and IT continuance intention may differ by individuals from the perspective of Regulatory Focus (RF) theory (Higgins, 1997; 1998). The RF theory distinguishes between two major motivational orientations – promotion focus and prevention focus (Higgins, 1997; 1998). Individuals with a promotion focus are driven by the need for attainment of positive outcomes, and are thereby oriented toward the maximization of gains, such as advancement and accomplishment. In contrast, individuals with a prevention focus are driven by the need for safety and avoidance of negative outcomes, and hence are oriented toward the minimization of losses and security. IT paradoxes involve both positive and negative aspects of technology usage, which may lead to gains and losses. The RF Theory has significant implications for IT paradoxes because an individual's RF guides his/her information processing and decision making, thereby influencing which information he/she specifically seeks out, pays attention to and retains when experiencing IT paradoxes. A better understanding of the role of motivational orientations in shaping individuals' perceptions of and attentions to various diametric anchors of IT paradoxes will provide important insights into the differentiated influences of diametric anchors of IT paradoxes and make it possible to assess their relative weights in determining different individuals' perceived IT performance ambiguity, trust in IT and IT continuance intention.

THEORETICAL BACKGROUND

Paradoxes of Technology

"Paradox maintains that something is both X and not-X at the same time" (Mick & Fournier, 1998, p. 125). The idea of technology paradox has been noted in different disciplines. In marketing literature, Mick and Fournier (1998) indicate that consumers' response to technology is paradoxical and propose eight paradoxes of technology: control/chaos, freedom/enslavement,

new/obsolete, competence/incompetence, efficiency/inefficiency, fulfills/creates needs, assimilation/isolation, and engaging/disengaging (P1-P8 shown in Table 1). They define these paradoxes around the idea that "polar opposite conditions can simultaneously exist, or at least can be potentiated, in the same thing" (Mick & Fournier, 1998, p. 124).

Table 1. Definitions of Technology Paradoxes Proposed by Mick and Fournier (1998)

Technology Paradox	Definition
P1: Control/Chaos	Technology can facilitate regulation or order, and technology can lead to upheaval or disorder.
P2: Freedom/Enslavement	Technology can facilitate independence or fewer restrictions, and technology can lead to dependence or more restrictions.
P3: New/Obsolete	New technologies provide the user with the most recently developed benefits of scientific knowledge, and new technologies are already or soon to be outmoded as they reach the marketplace.
P4: Competence/Incompetence	Technology can facilitate feelings of intelligence or efficacy, and technology can lead to feelings of ignorance and ineptitude.
P5: Efficiency/Inefficiency	Technology can facilitate less effort or time spent in certain activities, and technology can lead to more effort or time in certain activities.
P6: Fulfills/Creates Needs	Technology can facilitate the fulfillment of needs or desires, and technology can lead to development or awareness of needs or desires previously unrealized.
P7: Assimilation/Isolation	Technology can facilitate human togetherness, and technology can lead to human separation.
P8: Engaging/Disengaging	Technology can facilitate involvement, flow, or activity, and technology can lead to disconnection, disruption, or passivity.

Adopted from Mick and Fournier (1998)

Mick and Fournier's eight technology paradoxes can be applied to describe the paradoxical user experience of IT. The control/chaos paradox of IT suggests that IT provides users with a sense of control over the processes or outcomes of user activities, whereas human or technology errors that have immediate negative consequences may lead to feelings of chaos and helplessness. The freedom/enslavement paradox of IT considers that IT enables user activities independent of time and space yet leads to users' dependence on IT for task performance. The new/obsolete paradox

of IT suggests that the new ITs provide the benefits of the latest technological developments while simultaneously approaching obsolescence and constantly being replaced by augmented technologies. The competence/incompetence paradox of IT represents the ability of IT to facilitate feelings of efficacy versus its ability to induce feelings of incompetence. IT enables users to acquire new competencies or improve their existing abilities, yet the technological complexity of IT may also make users feel incompetent and unable to understand how the technology works. The efficiency/inefficiency paradox of IT considers that IT can improve efficiency by reducing the time and effort required to perform certain tasks but IT users may end up spending more time and effort due to their unfamiliarity with the technology or the IT's difficult-to-use features. The fulfills/creates needs paradox of IT occurs when the use of IT not only satisfies user needs but also creates additional needs such as frequent upgrades, additional gadgets, and complex knowledge. The assimilation/isolation paradox of IT refers to the ability to simultaneously facilitate human togetherness and human separation. IT may assimilate people who are apart through electronic channels yet increase isolation between people in close proximity by reducing the necessity of interpersonal interactions. The engaging/disengaging paradox of IT involves the ability of IT to facilitate involvement, flow or activity versus its ability to lead to disconnection, disruption or passivity. For example, IT allows users to engage in a wide variety of virtual communication activities. However, IT can also cause people to become less involved in and disengage from face-to-face activities, leading to detrimental social interactions.

The RF Theory

The RF theory distinguishes between two major motivational orientations – promotion focus and prevention focus – that guide individuals' goal pursuit behaviors (Higgins, 1997; 1998). The promotion focus driven by the need nurturance concerning an ideal self (the kind of person an individual would like to be) and thus are related to attainment of positive outcomes, such as advancement, accomplishment and aspirations. In contrast, the prevention focus are driven by the need for safety concerning an ought self (the kind of person an individual ought to be) and are related to avoidance of negative outcomes and fulfillment of responsibilities, duties, and obligations. Individuals with a promotion focus are sensitive to positive outcomes. They consider gains as success and non-gain as failure and regulate their attentions, perceptions and behaviors toward maximization of gains. Individuals with a prevention focus are sensitive to negative outcomes. They regard non-loss as success and loss as failure and regulate their attentions, perceptions and behaviors toward security and minimization of losses (Higgins & Tykocinski, 1992; Shah, Higgins & Friedman, 1998). These two distinct motivational orientation states can be either enduring personality characters or situationally induced by environmental cues (Forster, Higgins & Bianco, 2003; Forster, Higgins & Idson, 1998; Shah & Higgins, 1997). Irrespective of being dispositional or situationally induced, individuals' regulatory foci guide their preferences, information processing, decision making, and behaviors (Dholakia, Gopinath, Bagozzi & Nataraajan, 2006; Pham & Avnet, 2004; Sengupta & Zhou, 2007; Vellido, Lisboa & Meehan, 2000).

Individuals' regulatory foci prompt them to adopt strategies and engage in activities that are consistent with their regulatory orientations. Specifically, individuals utilize an eagerness strategic means to pursue a promotion goal but adopt a vigilance strategic means to fulfill a prevention goal (Crowe & Higgins, 1997). Let us consider a signal detection situation where individuals decide

whether an action is worth pursuing (Tanner & Swets, 1954; Trope & Liberman, 1996). There are four possible outcomes of each signal-detection trial: 1) a hit – deciding to take a correct action, 2) a miss – deciding not to take a correct action, 3) a correct rejection – deciding not to take a wrong action, and 4) a false alarm – deciding to take a wrong action. Since a promotion focus is concerned with the pursuit of gains and advancements, it entails the eagerness strategy to ensure hits and avoid misses (i.e., a loss of an opportunity for accomplishment). In contrast, since a prevention focus is concerned with safety and avoidance of failures, it involves the vigilance strategy to seek correct rejections and ensure against false alarms (i.e., making a mistake). It has been found that the states of eagerness value speed over accuracy and thus promote the reliance on affect-based heuristics that tend to be faster and less effortful, while the states of vigilance emphasize accuracy over speed and thus encourage the reliance on analytical reasoning that tend to be deliberative and accurate (Friedman & Forster, 2001, 2000; Cornwell & Higgins, 2016). In line with this reasoning, it can be suggested that promotion-induced eagerness is associated with the use of intuitive feelings in judgement and decision-making whereas prevention-induced vigilance is related to the use of deliberative reasons in judgement and decision-making (Cornwell & Higgins, 2016).

Regulatory Fit and Desirability-Feasibility Framework

The phenomenon that individuals prefer the means of goal pursuit that match and sustain their regulatory focus is termed as regulatory fit (Higgins, 2000; 2005; 2006). The fit between an individual's means of goal pursuit and regulatory orientation creates a subjective experience of "feeling right" that in turn increases his/her motivational engagement and enhances the perceived value of the goal pursuit (Avnet & Higgins, 2006; Wang & Lee, 2006). Several studies have applied the concept of regulatory fit to examine selective information processing and evaluation (Lee & Aaker, 2004; Pham & Avnet 2004; Pham & Higgins, 2005; Wang & Lee, 2006). Due to human being's limited processing capacity, selectivity of information is often necessary (Payne, Bettman & Johnson., 1992). Prior research on selective information processing has demonstrated that individuals tend to selectively process and reply on motivation-consistent information (Hart, Albarracin, Eagly, Brechan, Lindberg, Lee & Merrill, 2009). To the extent that motivational forces influence selective information processes (Hart et al., 2009), regulatory focus plays a key role in directing people's attention to information that fits their regulatory orientation (Wang & Lee, 2006). In other words, people are more likely to selectively pay attention to information that addresses their regulatory concerns. Since people experiencing regulatory fit are more motivated in their goal pursuit activities (Idson, Liberman & Higgins, 2000), individuals actively seeking information congruent with their regulatory orientations will be more motivated to elaborate on the information pertinent to their respective regulatory focus and, in turn, be more likely to be influenced by that information. More specifically, promotion-focused individuals will selectively pursue and elaborate on information that addresses concerns about growth and advancement, whereas prevention-focused individuals will seek out and elaborate on information that addresses concerns about safety and security (Wang & Lee, 2006).

Prior research on regulatory fit has demonstrated that promotion focus versus prevention focus is associated with change versus stability (Liberman, Idson, Camacho & Higgins, 1999), distant versus proximal temporal perspective (Pennington & Roese, 2003), abstract versus concrete mental representations (Lee, Keller & Sternthal, 2010), and desirability versus feasibility considerations (Liberman & Trope, 1998). The distinction between feasibility and desirability of

goal-directed actions corresponds to the distinction between means and ends of actions (Gollwitzer & Moskowitz, 1996; Liberman & Trope, 1998). Specifically, desirability refers to the value of an action's end state reflecting the superordinate why aspects of an action, whereas feasibility refers to the ease or difficulty of reaching an action's end reflecting the subordinate how aspects of an action (Carver & Scheier, 1990, 1999; Liberman & Trope, 1998; Vallacher & Wegner, 1987). Feasibility versus desirability is also an important dimension of level of mental construals (Eyal, Liberman, Trope & Walther, 2004). High-level construals entail developing abstract and global conceptualization, whereas low-level construals involve constructing concrete and local conceptualization (Trope, Liberman & Wakslak, 2007). Therefore, high-level construals represent attitude objects or events in terms of their abstract, essential features (Liberman, Sagristano & Trope, 2002). In contrast, low-level construals represent attitude objects or events in terms of their concrete, incidental features (Liberman, Sagristano & Trope, 2002). Desirability considerations constitute high-level construals of actions, and feasibility considerations constitute low-level construals of actions, because the why aspects of an action are more abstract and general and better convey the action's meaning than the more specific how details of the action (Vallacher & Wegner, 1987; 1989). Since prevention focused individuals vigilantly steer away from negative outcomes and losses, they would be more detail-oriented and use low-level construals to learn about all available information, especially the details that may jeopardize the success of the action. Hence, the feasibility concerns that reflect the "how aspects" of actions would be more prominent to promotion focused people. They would focus on the local details of the means in considerations and adopt concrete processing of action alternatives by which they can increase the chance to avoid negative outcomes. For promotion-focused individuals who eagerly strive to gain desirable outcomes and benefits, the desirability concerns reflecting the "why aspects" of actions would be more salient. They tend to go beyond the scrutiny of local details pertaining to the means and use high-level construals to engage in broader, more abstract consideration of the general value of the end state and why it should be attained.

An examination of the definitions of Mick and Fournier's (1998) technology paradoxes reveals that that some of the paradoxes are concerned with the desirability of using technology while others are associated with the feasibility to use technology. For example, the control/chaos, freedom/enslavement, new/obsolete, fulfills/creates needs, assimilation/isolation and engaging/disengaging paradoxes pertain to the benefits/drawbacks brought by the technology functionalities (Mishra, 2011). The use of a technology can either bring order to and control over user activities or create upheaval (control/chaos); facilitate user independence or promote dependence on technology (freedom/enslavement); realize the benefits of new technologies or cause the fear that the technology is becoming outdated and losing its usability (new/obsolete); provide a solution for an existing problem or make one aware of previously unnoticed needs (fulfills/creates needs); facilitate users' connection with others or lead to users' isolation (assimilation/isolation); facilitate users' engagement in activities or cause disruption in activities (engaging/disengaging). On the other hand, the technology paradoxes of competence/incompetence and efficiency/inefficiency are concerned with the extent to which it is easy or difficult to operate the technology (Mishra, 2011). The competence/incompetence paradox indicates users' superior versus inferior capabilities in performing certain tasks through the use of technology. The paradox of efficiency/inefficiency involves users' perceptions of time and effort spent in performing certain tasks using the technology. Both of these technology paradoxes largely reflect users' abilities to easily and efficiently operate the technology in task performance, which

determines the ease or difficulty to achieve the technology benefits. Thus, this research adopts a desirability-feasibility framework to integrate and classify the diametric anchors of technology paradoxes. The control/chaos, freedom/enslavement, new/obsolete, fulfills/creates needs, assimilation/isolation and engaging/disengaging paradoxes are categorized as the functional desirability paradoxes, and the competence/incompetence and efficiency/inefficiency paradoxes are classified as the usage feasibility paradoxes.

Trust in IT

Trust in IT is an important IS success factor that has received much attention in IS research (McKnight, 2005; McKnight, Choudhury & Kacmar, 2002). Trust in IT reflects one's willingness to depend on or be vulnerable to IT (McKnight, 2005). Research has shown that trust in IT is related to intention to adopt or continue using the technology (Reid & Levy, 2008; Wu, Zhao, Zhu, Tan & Zheng, 2011). IT paradoxes produce conflicting evaluations of IT and perceptions of IT performance ambiguity, and hence result in uncertainty regarding the future usage of IT (Johnson, Bardhi & Dunn, 2008). Thus, trust in IT will play a critical role in determining users' continued IT usage intention when users experience IT paradoxes, because trust involves a certain level of willingness to accept uncertainties and take risks.

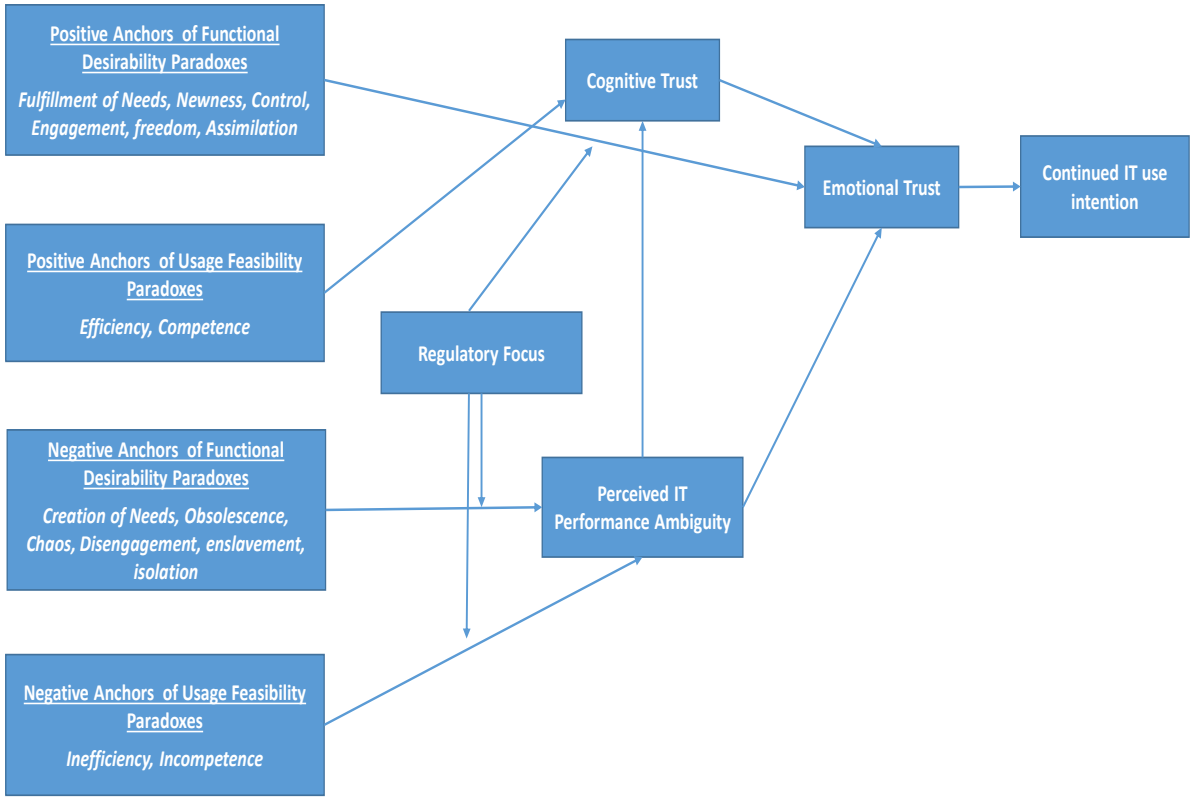
It has been suggested that trust is a multidimensional construct, consisting of two dimensions – cognitive trust and emotional trust, as trust decisions usually involve both reasoning and feeling (McAllister, 1995; Komiak & Benbasat, 2006). It is important to distinguish between the cognitive and emotional dimensions of trust, because each dimension involves different antecedents and developmental processes (McAllister, 1995). On the one hand, cognitive trust is developed through the rational evaluation of characteristics of IT, such as functionality, reliability and etc.; and on the other hand, emotional trust is formed based on the feelings of concern and attraction (Komiak & Benbasat, 2006). Cognitive trust in IT is a belief that an IT has the attributes necessary to perform as expected in a given situation in which negative consequences are possible (McKnight, Carter, Thatcher & Clay, 2011). Emotional trust in IT refers to a user's feelings of security and comfort about relying on an IT (Komiak & Benbasat, 2006). Prior research suggests that while cognitive trust serves as a foundation for the development of emotional trust, cognitive trust may no longer be needed once a high level of emotional trust has developed (McAllister, 1995). Thus, although cognitive trust in IT is a critical element in the initial adoption of IT, emotional trust in IT can become more influential in sustaining continued usage of IT (Nah & Davis, 2002).

THEORETICAL FRAMEWORK AND PROPOSITIONS

Drawing on the technology paradox research, the RF theory, and the research on trust in the IS field, this paper proposes a theoretical framework of the effects of IT paradoxes along the RF perspective (Figure 1). The framework classifies the IT paradoxes into two categories – the functional desirability paradoxes and the usage feasibility paradoxes and focuses on their differential effects on users' perceptions of IT performance ambiguity, their cognitive and emotional trust in IT and continuance intention to use IT depending on users' RF orientations. According to the RF theory, I argue that individuals will be sensitive to the factors relevant to their regulatory foci. The diametric anchors of IT functional desirability paradoxes (control/chaos, freedom/enslavement, new/obsolete, fulfills/creates needs, assimilation/isolation, and

engaging/disengaging) give rise to eager stimulation relevant to the promotion focus, whereas the diametric anchors of IT usage feasibility paradoxes (competence/incompetence and efficiency/inefficiency) induce vigilant simulation relevant to the prevention focus. To the extent that regulatory focus acts as a filter for individuals to process information selectively, promotion focused individuals and prevention focused individuals would respond differently to the diametric anchors of IT functional desirability paradoxes and usage feasibility paradoxes.

Figure 1. Theoretical Framework of the Effects of IT Paradoxes along the RF Perspective



IT paradoxes present users with both positive experiences (control, freedom, newness, fulfillment of needs, assimilation, engagement, competence and efficiency) and negative ones (chaos, enslavement, obsolescence, creation of needs, isolation, disengagement, incompetence and inefficiency). The conflicting experiences constitute stressors that entail ambivalent evaluations of IT. More specifically, on the one hand, the positive anchors of IT paradoxes (control, freedom, newness, fulfillment of needs, assimilation, engagement, competence and efficiency) increase users' trust in IT by demonstrating IT's ability to bring benefits and desired outcomes; and on the other hand, the negative anchors of IT paradoxes (chaos, enslavement, obsolescence, creation of needs, isolation, disengagement, incompetence and inefficiency) can elicit feelings of frustration, stress and anxiety in users and inhibit users' ability to fully evaluate the performance of IT, thus resulting in perceived performance ambiguity of IT (Johnson, Bardhi & Dunn, 2008).

With regard to users' experience with the negative anchors of IT paradoxes, the negative anchors of IT functional desirability paradoxes (chaos, enslavement, obsolescence, creation of needs,

isolation and disengagement) could be more accessible to individuals with a promotion focus, whereas the negative anchors of IT usage feasibility paradoxes (incompetence and inefficiency) could be more salient to individuals with a prevention focus. The promotion focus propels people to focus on the functional desirability of IT, whereas the prevention focus makes people more worried about the feasibility of using IT and concentrate on the usage feasibility of IT. Therefore, the effects of the negative anchors of IT functional desirability and usage feasibility paradoxes on users' perceptions of IT performance ambiguity may vary depending on their regulatory foci. Perceived performance ambiguity of IT will in turn negatively influence IT users' trust in IT, because being unable to fully determine the performance of IT will undermine IT users' cognitive and emotional trust in IT. Hence, the following hypotheses can be proposed.

> Hypothesis 1: The negative anchors of IT functional desirability paradoxes (chaos, enslavement, obsolescence, creation of needs, isolation and disengagement) have a stronger positive influence on perceived performance ambiguity of IT for promotion-focused users than for prevention-focused users.
>
> Hypothesis 2: The negative anchors of IT usage feasibility paradoxes (incompetence and inefficiency) have a stronger positive influence on perceived performance ambiguity of IT for prevention-focused users than for promotion-focused users.
>
> Hypothesis 3: Perceived performance ambiguity of IT will be negatively associated with cognitive trust in IT.
>
> Hypothesis 4: Perceived performance ambiguity of IT will be negatively associated with emotional trust in IT.

Regulatory fit occurs when an individual's means of goal pursuit aligns with his/her regulatory focus (Higgins, 2000). In case of such alignment, the value of the pursued goal will increase, separate from the goal's worth based on expected benefit and cost outcomes (Higgins, 2006). The positive anchors of IT functional desirability paradoxes (control, freedom, newness, fulfillment of needs, assimilation and engagement) have higher fit for individuals in a promotion focus than they do for those in a prevention focus, because desirable IT functions maintain eagerness but reduce vigilance. The positive anchors of IT usage feasibility paradoxes (competence and efficiency) have higher fit for individuals in a prevention focus than they do for those in a promotion focus, because the ability to easily and efficiently use IT is critical to avoiding failures and ensuring error-free IT usage, and thus serves the need for vigilance rather than eagerness.

The experience of regulatory fit makes people "feel right" about what they are doing, and thus strengthens their engagement in the goal pursuit process and amplifies their dominant preferences (Camacho, Higgins & Luger, 2003; Higgins, 2006; Higgins, Idson, Freitas, Spiegel & Molden, 2003; Lee & Higgins, 2009). Individuals of promotion regulatory fit make judgments based on feelings rather than reasons (Avnet and Higgins, 2006). Thus, in a promotion regulatory fit state where the positive anchors of IT functional desirability paradoxes align with the promotion focus, IT users tend to have positive feelings toward relying on IT and develop emotional trust in IT. On the contrary, individuals experiencing prevention regulatory fit base their judgments on reasons rather than feelings (Avnet and Higgins, 2006). So, in a prevention regulatory fit state

where the positive anchors of IT usage feasibility paradoxes fit with the prevention focus, IT users are more likely to develop cognitive trust in IT based on good reasons. Although the positive anchors of IT usage feasibility paradoxes may appeal less to the maximal goals activated by a promotion focus oriented toward advancement and achievement than those of IT functional desirability paradoxes, the positive outcomes associated with the competence and efficiency in using IT can still produce positive evaluation of IT, and thus lead promotion-focused IT users to develop cognitive trust in IT. Compared to the positive anchors of IT usage feasibility paradoxes, the positive anchors of the functional desirability paradoxes of IT are less accessible to the prevention-focused individuals with the minimal goal oriented toward security. In addition, trust involves a willingness to accept vulnerability and take risks, which appears largely incompatible with a concern for safety and security activated by a prevention focus. Thus, it seems reasonable to expect that the positive anchors of the functional desirability paradoxes of IT are largely unrelated to IT trust for prevention-focused IT users. Therefore, the following hypotheses are suggested.

> Hypothesis 5: The positive anchors of IT functional desirability paradoxes (control, freedom, newness, fulfillment of needs, assimilation and engagement) will be positively associated with emotional trust in IT in promotion-focused users.
>
> Hypothesis 6: The positive anchors of IT usage feasibility paradoxes (competence and efficiency) will be positively associated with cognitive trust in IT in prevention-focused users.
>
> Hypothesis 7: The positive anchors of IT usage feasibility paradoxes (competence and efficiency) will be positively associated with cognitive trust in IT in promotion-focused users.
>
> Hypothesis 8: The positive anchors of IT functional desirability paradoxes (control, freedom, newness, fulfillment of needs, assimilation and engagement) will be unrelated to cognitive trust or emotional trust in IT in prevention-focused users.

Cognitive trust and emotional trust in IT not only constitute psychological coping responses to the positive anchors of IT paradoxes and the perceived performance ambiguity of IT, but they also mediate the effects of perceived IT performance ambiguity and positive anchors of IT paradoxes on continued IT usage intention. Since trust involves a leap of faith beyond what is supported by available information (Lewis & Weigert, 1985; Luhmann, 1979), it is an especially critical prerequisite to determining continuance intention when people are unable to fully determine the performance of IT. Prior research suggests emotional trust plays an important role beyond cognitive trust in determining people's intention to use IT. Emotional trust was found to fully mediate the impact of cognitive trust on the intention to adopt online recommendation agent and on the re-patronage intention (Komiak & Benbasat, 2006; Wu, Su & Wei, 2016). Thus, the following hypotheses can be proposed.

> Hypothesis 9: Cognitive trust in IT will be positively associated with emotional trust in IT.

Hypothesis 10: Emotional trust in IT will be positively associated with continued intention to use IT.

RESEARCH DESIGN

Sample and Data Collection

To test the proposed research model and its associated propositions, I will use mobile banking services as the target technology. Mobile banking services are considered appropriate for this study because they exemplify consumer-oriented contemporary IT and Mick and Fournier's eight technology paradoxes can operate in the mobile banking environment. Thus, a web-based survey will be conducted to investigate the presence of paradoxes associated with the use of mobile banking services and their effects on the perceived performance ambiguity of mobile banking services and user's cognitive trust, emotional trust and continued usage intention along the perspective of regulatory focus. The participants for this study will be the faculty, staff and students at a southeast university in the USA. To solicit participation, an email invitation will be sent to all faculty, staff and students at the university. Interested participants may click on the link in the email invitation to be directed to the survey website. A screening question will be included at the beginning of the survey to determine whether the respondent has used mobile banking services. The survey website will be designed in such a way that only those who have used mobile banking services will be able to proceed with the survey. To encourage participation, prizes (Amazon.com Gift Cards) will be provided by means of a lucky draw. The survey will not reveal the research purpose of identifying the presence of technology paradoxes to the participants. Instead, it will merely mention that the study addresses the use of mobile banking services. Respondents will be asked to respond to all survey questions related to the presence of technology paradoxes, regulatory focus, perceived performance ambiguity, cognitive trust, emotional trust, continuance intention to use mobile banking services, and etc.

Since data will be collected from a sample of university faculty, staff and students, and users of mobile banking services tend to be younger and more tech-savvy, the results of the proposed study may not be applicable to other populations. However, all participants are users of mobile banking services, and the proposed study is designed to study their psychological and behavioral responses that are considered primitive and instinctive for all human beings, the use of faculty, staff and student sample should not present a serious threat to the validity of this study.

Measures

The survey instrument will be developed by incorporating and adapting existing valid and reliable scales where appropriate. The measurement scales of IT paradoxes will be developed to capture the presence of the technological paradoxes identified by Mick and Fournier (1988), Johnson, Bardhi amd Dunn (2008) and Jarvenpaa and Lang (2005). For each IT paradox, two questions representing the antagonism associated with that paradox will be asked. All the questions related to IT paradoxes will be interspersed in such a way that the opposing evaluations relating to a certain paradox would not be placed together or in close proximity, in order to disguise the purpose of the research. Furthermore, during the data collection process, the word paradox will not be used nor will any other term that may suggest the participants should tell their paradoxical or ambiguous

experiences arising from the use of mobile banking services. The promotion focus and prevention focus will be measured using Lockwood, Jordan and Kunda's (2002) measurement scales. The measurement scale of perceived performance ambiguity of mobile banking services will be adapted from the work of Johnson, Bardhi amd Dunn (Johnson, Bardhi & Dunn, 2008). The cognitive trust and emotional trust in mobile banking services will be measured using Komiak and Benbasat's measurement scales (Komiak & Benbasat, 2006). The measure of continued intention to use mobile banking services will be derived from the prior work on IT continuance (Bhattacherjee & Premkumar, 2004; Thong, Hong & Tam, 2006). Additionally, the participants' age and prior experience with mobile banking services will be measured as control variables.

CONCLUSIONS

This research proposes a theoretical framework of the effects of IT paradoxes along the perspective of regulatory focus. Drawing on the technology paradox research, the regulatory focus theory, and the IT trust research, the framework suggests that IT users' regulatory foci (promotion focus vs. prevention focus) moderate the impacts of IT paradoxes on users' perceived IT performance ambiguity and cognitive and emotional trust in IT, which in turn mediates the relationship between IT paradoxes and people's continued IT usage intention. The major contributions of this paper are as follows. First, this research recognizes the complex nature of IT paradoxes and utilizes the desirability-feasibility framework to categorize the IT paradoxes into the functional desirability paradoxes and the usage feasibility paradoxes. It proposes the differentiated effects of the diametric anchors of IT functional desirability paradoxes and usage feasibility paradoxes along the perspective of regulatory focus. This is the first attempt to investigate the effects of IT paradoxes from a motivational perspective. Second, this research adds to the RF theory (Higgins, 1997) by developing a model of the moderating effects of regulatory foci on the influences of IT paradoxes on users' perceived IT performance ambiguity as well as their cognitive trust and emotional trust in IT, which consequently influence their continued IT use intention. Third, the proposed theoretical framework examines simultaneously the effects of both positive and negative anchors of IT paradoxes and makes it possible to assess their relative weights in determining individuals' intentions of continued IT usage. It also further expands beyond the traditional technology acceptance research in explaining IT continuance intention. Finally, this paper also provides guidance for managers to better predict and influence IT users' perceptions and technology usage behaviors from the experience of IT paradoxes. Understanding the mediating effects of IT trust and the moderating role of regulatory foci in the relationships between IT paradoxes and IT continuance intention will help managers to counter the adverse effects of IT paradoxes by reducing perceptions of IT performance ambiguity and building cognitive and emotional trust in IT. For example, companies may be able to reduce perceived IT performance ambiguity and facilitate IT trust by taking measures (e.g., managing customer expectations, offering online tutorials and help, and etc.) to make consumers better recognize the desirability and feasibility of using IT.

REFERENCES

Assigbetse, L. (2019). *Joyful or joyless: The paradoxical effects of mobile applications on users' affective well-being and coping strategies* [Master thesis, Radboud University]. Educational Repository. Retrieved from https://theses.ubn.ru.nl/bitstream/handle/123456789/8836/Assigbetse%2c_Lilly_1.pdf?sequence=1

Avnet, T., & Higgins, E. T. (2006). How regulatory fit affects value in consumer choices and opinions. *Journal of Marketing Research*, *43*, 1–10.

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test. *MIS Quarterly*, 28(2), 229–254.

Camacho, C. J., Higgins, E. T., & Luger, L. (2003). Moral value transfer from regulatory fit: What feels right is right and what feels wrong is wrong. *Journal of Personality and Social Psychology*, *84*(3), 498–510.

Carver, C. S., & Scheier, M. F. (1990). Principles of self-regulation. In E. T. Higgins, and R. M. Sorrentino (Eds.). *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 3–52). New York, NY: Guilford Press.

Carver, C. S., & Scheier, M. F. (1999). Themes and issues in the self-regulation of behavior. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 12, pp. 1–106). Mahwah, NJ: Erlbaum.

Caspi, P. V., Olekalns, M., & Druckman, D. (2017). After the fall: Regulatory focus, trust and negotiators' responses to a crisis. *Journal of Trust Research*, *7*(1), 51–70.

Chae, M., & Yeum, D. (2010). The impact of mobile technology paradox perception and personal risk-taking behaviors on mobile technology adoption. *International Journal of Management Science*, *16*(2), 115–138.

Chinn, S. S. (2001, March-April). The technology paradox. *Industrial Management*, 25–27.

Cornwell, J. F. M., & Higgins, E. T. (2016). Eager feelings and vigilant reasons: Regulatory focus differences in judging moral wrongs. *Journal of Experimental Psychology: General*, *145*(3), 338–355.

Crowe, E., & Higgins, E.T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior & Human Decision Processes*, *69*, 117–132.

Dholakia, U.M., Gopinath, M., Bagozzi, R.P., & Nataraajan, R. (2006). The role of self-regulatory focus in the experience and self-control of desire for temptations. *Journal of Consumer Psychology*, *16*(2), 163–175.

Förster, J., Higgins, E.T., & Bianco, A.T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns*? Organizational Behavior and Human Decision Processes*, *90*, 148–164.

Förster, J., Higgins, E. T., & Idson, L. C. (1998). Approach and avoidance strength during goal attainment: Regulatory focus and the "goal looms larger" effect. *Journal of Personality & Social Psychology*, *75*, 1115–1131.

Friedman R. S., & Förster, J. (2001). The effects of promotion and prevention cues on creativity. *Journal of Personality and Social Psychology*, *81*(6), 1001–1013.

Friedman R. S., & Förster, J. (2000). The effects of approach and avoidance motor actions on the elements of creative insight. *Journal of Personality and Social Psychology*, *79*, 477–492.

Eyal, T., Liberman, N., Trope, Y., & Walther, E. (2004). The pros and cons of temporally near and distant action. *Journal of Personality and Social Psychology*, *86*(6), 781–95.

Gollwitzer, P. M., & Moskowitz, G. B. (1996). Goal effects on action and cognition. In E. T. Higgins, and A. W. Kruglanski (Eds.). *Social psychology. Handbook of basic principles* (pp. 361–399). New York, NY: The Guilford Press.

Hart, W., Albarracin, D., Eagly, A. H., Brechan, I., Lindberg, M., Lee, K., & Merrill. L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*(4), 555–588.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*, 1280–1300.

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1–46). San Diego, CA: Academic Press.

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, *55*, 1217–1230.

Higgins, E. T. (2005). Value from regulatory fit. *Current Directions in Psychological Science*, *14*, 209–213.

Higgins, E. T. (2006). Value from hedonic experience and engagement. *Psychological Review*, *113*, 439–460.

Higgins, E. T., Idson, L. C., Freitas, A. L., Spiegel, S., & Molden, D. C. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology*, *84*(6), 1140–1153.

Higgins, E. T., & Tykocinski, O. (1992). Self-discrepancies and biographical memory: Personality and cognition at the level of psychological situation. *Personality & Social Psychology Bulletin*, *18*, 527–535.

Idson, L. C., Liberman, N., & Higgins, E. T. (2000). Distinguishing gains from nonlosses and losses from nongains: A regulatory focus perspective on hedonic intensity. *Journal of Experimental Social Psychology*, *36*(3), 252–274.

Jarvenpaa, S. L., & Lang, K. R. (2005). Managing the paradoxes of mobile technology. *Information Systems Management*, *22*, 7–23.

Johnson, D. S., Bardhi, F., & Dunn, D. T. (2008). Understanding how technology paradoxes affect customer satisfaction with self-service technology: The role of performance ambiguity and trust in technology. *Psychology & Marketing*, *25*(5), 416–443.

Keller, J., Mayo, R., Greifeneder, R., & Pfattheicher, S. (2015). Regulatory focus and generalized trust: the impact of prevention-focused self-regulation on trusting others. *Frontiers in Psychology, 6*, 254.

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly*, *30*(4), 941–960.

Lee, A. Y., & Aaker, J. L. (2004). Bringing the frame into focus: The influence of regulatory fit on processing fluency and persuasion. *Journal of Personality and Social Psychology*, *86*, 205–218.

Lee, A. Y., Keller, P. M., & Sternthal, B. (2010). Value from regulatory construal fit: The persuasive impact of fit between consumer goals and message concreteness. *Journal of Consumer Research*, *36*, 735–47.

Lee, A. Y., & Higgins, E. T. (2009). The persuasive power of regulatory fit. In M. Wänke (Ed.), *The social psychology of consumer behavior* (pp. 319–333). New York, NY: Psychology Press.

Lee, W. (2016). When the future technology is now: Paradoxical attitudes of consumer and evaluation of IoT service. *International Journal of Smart Home*, *10*(6), 115–126.

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*, 967–985.

Liberman, N., Idson, L. C., Camacho, C. J., & Higgins, E. T. (1999). Promotion and prevention choices between stability and change. *Journal of Personality and Social Psychology*, *77*, 1135–1145.

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology*, *38*, 523–34.

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of Temporal Construal Theory. *Journal of Personality and Social Psychology*, *75*, 5–18.

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology*, *83*, 854–864.

Luhmann, N. (1979). *Trust and power*. Chichester, NY: Wiley.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59.

McKnight, D. H. (2005). Trust in information technology. In G. B. Davis (Eds.), The Blackwell *Encyclopedia of Management*. *Vol. 7 Management Information Systems* (pp. 329–331). Malden, MA: Blackwell.

McKnight, D. G., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 12–32.

McKnight, H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-Commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

Mick, D. G., & Fournier, S. (1998). Paradoxes of technology: Consumer cognizance, emotions, and coping strategies. *Journal of Consumer Research*, *25*, 123–143.

Mishra, A. (2011). *A study of cognitive processing and inhibitions of adopters and non-adopters of technology-based products* [Doctoral dissertation, University of Arizona]. UA Campus Repository. Retrieved from https://repository.arizona.edu/bitstream/handle/10150/202990/azu_etd_11812_sip1_m.pdf?sequence=1

Nah, F. & Davis, S. (2002). HCI research issues in e-commerce. *Journal of Electronic Commerce Research*, *3*(3), 98–113.

Orlikowski, W. J. (1991). Integrated information environment or matrix of control? The contradictory implications of information technology. *Accounting, Management, and Information Technology*, *1*, 9–42.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*, 87–131.

Pennington, G. L., & Roese, N. J. (2003). Regulatory focus and temporal distance. *Journal of Experimental Social Psychology*, *39*, 563–76.

Pham, M., & Avnet, T. (2004). Ideals and Oughts and the Reliance on Affect versus Substance in Persuasion. *Journal of Consumer Research*, *30*(4), 503–518.

Pham, M. T., & Higgins, E. T. (2005). Promotion and prevention in consumer decision making: State of the art and theoretical propositions. In S. Ratneshwar, and D. G. MICK (Eds.). *Inside consumption: consumer motives, goals and desires* (pp. 8–43). London and New York: Routledge.

Reid, M., & Levy, Y. (2008). Integrating trust and computer self-efficacy with TAM: An empirical assessment of customers' acceptance of banking information systems (BIS) in Jamaica. *Journal of Internet Banking and Commerce*, *12*(3), 1–18.

Sengupta, J., & Zhou, R. (2007). Understanding impulsive eaters' choice behaviors: The motivational influences of regulatory focus. *Journal of Marketing Research*, *44*(2), 297–308.

Shah J., & Higgins, E. T. (1997). Expectancy X value effects: Regulatory focus as determinant of magnitude and direction. *Journal of Personality and Social Psychology*, *73*, 447–458.

Shah, J., Higgins, E. T., & Friedman, R. S. (1998). Performance incentives and means: How regulatory focus influences goal attainment. *Journal of Personality and Social Psychology*, *74*, 285–293.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.

Thong, J. Y. L., Hong, S., & Tam, K. Y. (2006). The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of Human-Computer Studies*, *64*, 799–810.

Trope, Y., & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In E. T. Higgins and A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239–270). New York, NY: Guilford.

Trope, Y., Liberman, N., & Wakslak, C. (2007). Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *Journal of Consumer Psychology*, *17*(2), 83–95.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they are doing? Action identification and human behavior. *Psychological Review*, *94*, 3–15.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, *57*, 660–671.

Vellido, A., Lisboa, P. J. G., & Meehan, K. (2000). Quantitative characterization and prediction of on-line purchasing behavior: A latent variable approach. *International Journal of Electronic Commerce*, *4*(4), 83–10.

Wang, J., & Lee, A. Y. (2006). The role of regulatory focus in preference construction. Journal of *Marketing Research*, *43*, 28–38.

Winner, L. (1994). Three paradoxes of the information age. In G. Bender and T. Druckey (Eds.). *Culture on the Brink: Ideologies of Technology* (pp. 191–197). Seattle, WA: Bay.

Wu, K., Zhao, Y., Zhu, Q., Tan, X., & Zheng, H. (2011). A meta-analysis on the impact of trust on technology acceptance model: Investigation of moderating influence of subject and context type. *International Journal of Information Management*, *31*, 572–581.

Wu, X, Su, L., & Wei, S. (2016). Research on the relationship between store environmental cues and re-patronage behavior based on emotional trust. *Business Management Journal*, *38*(08), 98–108.

Zhuang, W., Hsu, M. K., Brewer, K. L., & Xiao, Q. (2013). Paradoxes of social networking sites: an empirical analysis. *Management Research Review*, *36*(1), 33–49.

TOWARDS EFFECTIVE CUSTOMER DATA VISUALIZATION: DATA-DRIVEN DOCUMENTS (D3.JS) VS. GOOGLE CHARTS

Cindi Smatt, University of North Georgia

Renée Pratt, University of North Georgia

Tamirat Abegaz, University of North Georgia

Allie Dobbs, University of North Georgia

ABSTRACT

Organizations today are in vital need of viewing, interacting, and interpreting vast amounts of data in quick easy design views. Management wants to visualize business growth, forecasting and trends in diagrams and dashboards. With an increase in data analytics, data visualization is a way to display information in a visual manner whether on paper, a computer screen, or a projector. Data-Driven Documents have become increasingly more important in the corporate world. It is becoming a new and popular way to share data in an interactive and creative way. With numerous tools designed to facilitate these data visualization needs, it becomes difficult to know and understand what users interpret as easy, useful, and/or efficient. We explore two popular freely available data visualization tools: D3.JS and Google Charts Tools for retail customer data. Thirty participants were recruited for the study. However, only 29 participants completed the task. A paired sample t-test was performed. The results indicated that D3.JS is more efficient than Google Charts, however Google Charts is more user friendly than D3.JS. Our research highlights the significance that different visualization tools may amplify users' levels of effectiveness, efficiency and preference based on its presentation and ease of use. Therefore, organizations should be mindful to their selection of data visualization tools and their audience.

Keywords: Data Visualization, Data-Driven Documents, Google Charts, D3.JS, Open Source

INTRODUCTION

Businesses collect and access more data today than they ever imagined. The challenge with large datasets is interpreting and presenting data in an easy and understandable form (Qin, Luo, Tang, & Li, 2020; Wang, Yang, Wang, Sherratt, & Zhang, 2020). Data analysis and presentation has proven to be an important aspect to every business (Jin, Wah, Cheng, & Wang, 2015), and thus creates a larger demand for better data analytic tools. One challenge that businesses have is that the results of most analyses are simply raw numbers and very difficult for the customer to interpret as is. Therefore, there is a demand for easier to read, interactive data results (Center, 2013), also known as data visualization. Data (or information) visualization is the representation or presentation of data/information in a visual and meaningful graphical form such as charts, diagrams, maps, plots, etc. Data visualization enables end-users to analyze more effectively and efficiently the relationships among the represented data.

Graphical interpretations become increasingly essential with the excessive amount of data available (Fu et al., 2014). Visual diagrams are useful at gaining the customer's attention because visuals are more likely to reach the customer, as it does not require an understanding of a language code but goes beyond specific experiences (Hariharan & Krithivasan, 2016). In addition to obtaining their attention, the diagram must accurately tell the story and express the purpose of data analysis. In order to present meaningful data visualization, it is critical to understand the needs and expectations of an end-user's sense of efficiency, effectiveness, and ease of use. We begin this process by first exploring the developer tools that aid in presenting data. By exploring different tools, we engage in preferences of users and highlight significant characteristics that represent effectiveness, efficiency, presentation layout and ease of use. Although research has shown preferences for types of visualizations (e.g., bar/line charts, colors, shapes, etc.) (Luo, Qin, Tang, & Li, 2018), there has been limited research investigating the preferred tool for presenting such characteristics. This paper attempts to address the gap by answering the questions: 1) Which visualization tool is more effective and efficient? and 2) Do user preferences influence data visualization tool choice?

In the following pages we explore the differences between data visualization tools, the effectiveness and efficiency of these tools, and the user preferences with respect to ease of use, friendliness and overall satisfaction when engaging in each of the graphical presentations. This research compares the effectiveness, efficiency, and user preference of two popular data visualization tools: D3.JS and Google Charts. We begin with a background discussion on data visualization and our two data visualization tools - Google Charts (Developers.google.com, 2019) and Data-Driven Documents (D3.JS) (Bostock, 2019). From there we present our research design and methodology. Specifically, we use AKDesigns Boutique dataset to do comparative analysis of the two data visualization tools. Finally, we share our results, discussion and conclusion.

RELATED WORK

Data visualization is highly beneficial in data analysis as it graphically allows the user to find interesting patterns and make interpretations of the data. However, there are many challenges with huge amounts of data; one such challenge is how to capture, store, analyze, share, search, and present it in a meaningful way (Martinuzzi, 2016). Tools traditionally used to present data are far less useful with the massive rate of growth of data. Data visualization has certainly been around for many years. It is the presentation of data in the style of graphs or images that increases the worthiness of data visualization tools. Business decision makers are able to analyze data and grasp difficult concepts or identify new patterns through data visualization (Martinuzzi, 2016). The most important feature of these tools is the interactive nature with which the viewer can manipulate the presentation of the data or how quickly the visualization adapts itself at runtime. Interactive data visualization allows for efficient manipulation of data by helping people understand complex patterns and/or trends that might otherwise go unnoticed. A data visualization tool can only be effective if it is the appropriate tool for the task and used in the right way.

There are several data visualization tools. There is no one criterion for the classification of visualization tools. However, some have classified these tools as either open source or commercial/proprietary (Hariharan & Krithivasan, 2016). One difference between open source and commercial visualization tools is that the commercial tools are typically drag and drop

applications, but open source tools require more coding with Java Script being the typical language used. This is a significant difference as drag and drop tools are easier to learn and user friendly while open source tools require expertise in coding and more development time. Some of the open source options are Data-Driven Documents (D3.JS), Datawrapper, and dygraphs. Commercial (or proprietary) examples include Tableau, Qlick Sense, Spotfire or FusionCharts. Although Google Charts (Zhu, 2012) is considered a commercial tool it is free for all to use. Commercial tools tend to be part of a bundled package for users, but for additional costs may be tailored, while open-source tools are tailor made visualizations.

Another classification for data visualization tools is the required amount of coding. Many open source tools require a demanding amount of coding language skills. Java Script is the primary coding language to create visualizations. This can be challenging as Java Script requires some expertise and a good amount of development time. This increases the user's learning curve needed for these types of visualization tools. Most commercial visualization tools require little to no coding. They use drag and drop to create visualizations. This creates an easier to learn environment and reduces time invested.

Developers who have the skills to create the appropriate visual representations of the data are greatly sought after. In an effort to address the coding skills necessary for many open source data visualization tools, there are companies creating and developing ways to educate and train developers in the desired languages and tools. For example, there is a data science training program called Metis that was designed to teach professionals how to develop data visualization by using D3.JS (Metis, 2015). This program allows Metis to convene talent in one location and teach developers and designers required skillsets. These skills are complex, but in high demand. Other types of classifications for visualization tools include output types (e.g., maps, timelines, etc.) or output formats (e.g., SVG, PNG, etc.).

While knowledge drawn from these classifications and others are important to designers and developers, the end-users are an even larger and more diversified audience with a greater need to engage and comprehend the visual presentation of data. End-users include individuals in various industries with varying levels of expertise (e.g., students, customers, personnel, managers, executives, etc.). End-users desire the ability to quickly and easily access data in graphical formatthat allows for processing of complex data and a better understanding of what is happening around them. For example, consider how difficult it would be to understand weather forecasts without a graphical representation of the content. Human-computer interaction (HCI) studies highlight the many important preferences and characteristics of end-users and various technologies (Fox & Hendler, 2011). Data visualization is capable of being both aesthetically pleasing and functional (Keahey, Rope, & Wills, 2017). In fact, it should. There are many resources that discuss the importance of having a visually appealing product and this goes for data visualization as well.

Visualizations are often used to convey information, communicate key concepts, or guide attention to a particular area. Once the developers have been trained and taught to understand the importance of a good-looking product, it is also important to make sure the developers know how to conduct usability testing to ensure their clients are happy with the product being created for them (Skov & Stage, 2012). Usability testing is important to developers because they need to know how their end user thinks and interacts with the system. They need to know what kinds of improvements need to

be made and if their visualization is unclear to the customer. The more the customer is involved with feedback, the better the end results will be.

We have chosen to investigate and evaluate Data-Driven Documents (D3.JS) and Google Charts. To provide some contrast in the type of data visualization tools available as well as their accessibility, we have chosen two popular free data visualization tools where one is categorized as open source (D3.JS) and the other is commercial (Google Charts).

D3.JS

D3.JS is a JavaScript visualization tool that offers efficient data manipulation with no extra installation processes. It supports large data sets and interactivity, such as zooming and panning. D3.JS is an example of a web-based visualization tool. This tool is more advantageous because D3.JS offers a larger and easier availability of opportunities since all it needs is a web browser to access it. D3.JS uses HTML (Hypertext Markup Language) for page content, CSS (Cascading Style Sheets) for aesthetics, JavaScript for interaction, and SVG (Scalable Vector Graphics) for vector graphics enabled by a shared representation of the page called the document object model (DOM) (Bostock, 2019; Bostock, Ogievetsky, & Heer, 2011; Jiang, Fang, Ge, & Zhou, 2007). It also offers the possibility of creating almost any type of visualization from the ground up. However, the learning curve is high for developers because there are no built-in charts and it requires a good understanding of SVG and DOM elements to take advantage of its capabilities (Bostock, 2019).

D3.JS uses the data from arrays to create HTML tables or to create different SVG charts. HTML is the standard language for developing web sites and integrates well with JavaScript, CSS, and other languages. Scalable Vector Graphics (SVG) is a format that is used in many areas such as animation, user interfaces, Web graphics and mobile applications (Jiang et al., 2007). SVG is a vector-based image format for handling two-dimensional graphics. The format consists of an XML-based file format as well as programming API for graphical applications. An XML-based file format uses the Document Object Model (DOM), which is a programming API which defines the way documents are accessed and manipulated. DOM is also used in HTML which makes it easy for SVG format and HTML format to work together. This makes the SVG format very relevant when developing web sites that are using interactive graphics (W3.org, 2017). D3.JS allows for data to be created into an HTML table or into different SVG charts.

Google Charts

Google Charts is not open sourced but is a free tool for personal and commercial use. It is a Java library running on HTML5 and SVG and aims at Android, iOS and total cross-browser compatibility, including older Internet Explorer versions supported via VML. All the charts created are interactive and some are even zoomable. Google Charts is very user friendly and their site features a really nice and comprehensive gallery where users can see the kind of visualizations and interactions they need (Developers.google.com, 2019). Google Charts does have an easier learning curve as it provides an extensive list of built-in charts but learning the variations among them would take a longer period of time. Google Charts tends to be preferred by developers who have the JavaScript experience, but novices may be troubled by the SQL syntax needed for data retrieval (Supaartagorn, 2016). According to

Suppaartagorn (2016), a solution would be to use a framework that is designed to connect with Google Charts API.

A descriptive comparison table highlights the main differences in Table 1.

Table 1. Comparison table of JavaScript Visualization Tools

Tool	Commercial/ Open Source	Cost	Coding Requirement	Learning Curve	Built in Charts
D3.JS	Open Source	Free	Heavy coding	Difficult	None
Google Charts	Commercial	Free	Drag and drop	Gradual	25+ variations

D3.JS is the most preferred data visualization tool among developers (Rodden, 2014). It is widely used for generating dynamic and interactive data visualization in a web browser. D3.JS runs fast with minimal overhead in almost any modern web browser (Bostock et al., 2011; Sack, Donohue, & Roth, 2015). Google Charts is the most widely used data visualization tool, as it is powerful, simple to use, and free of charge (Zhu, 2012). There are great resources that help to explain how to use Google Charts with R to make interactive charts that have been famous for appearing in TED talks (Gesmann & de Castillo, 2011). R is a programming language which is a freely available software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing (r-project.org). These interactive charts make it possible for the customer to view the actual changes in the data and to not have to imagine it for him or herself. Additionally, both D3.JS and Google Charts renders charts with Scalable Vector Graphics (SVG); although, Google Charts adds Vector Markup Language (VML). D3.JS is powerful Java Script library for manipulating documents based on data, charts rendered in SVG, and has nine chart types; Google Charts displays live data on your site, charts rendered in HTML 5 using SVG and VML, and has nine chart types. Both tools are described further in the next sections.

RESEARCH METHODOLOGY

As the field of information technology matures, understanding the effectiveness, efficiency, and user preferences of data visualization tools and frameworks become more important. This research compares the effectiveness, efficiency, and user preference of two popular data visualization tools: D3.JS and Google Charts. The data was captured through the development of an interface representing each data visualization tool including a task survey.

Interface Development

The two data visualization tools were chosen to interact with a dataset appropriate to develop a task for our participants. Both D3.JS and Google Charts have maps that are available for the public. However before the sample data can be connected and displayed, both maps must be adjusted and altered. The Google Charts Map is produced using free code provided by Google and allowed for the adjustment of features such as color, range of data, size. The D3.JS map required more detailed coding using JavaScript. Once the maps were altered, they were added to the overall design of the interface.

The tasked survey interface design is arranged in an accordian bootstrap layout. There are five sections in the accordion (See Figure 1). The bootstrap and JavaScript code allows the sections to

toggle between showing and hiding sections. If one section is showing, the other sections are hidden. To capture participant's related information, inputs are saved even when the sections are hidden.

Section 1 is a simplified consent form approved by IRB and informing the participant of the purpose, procedures, risks/benefits, and confidentiality of the study. Section 2 collects basic information about the participant (age and student status). Figure 1 show the survey user interface used to capture basic information. Section 3 labeled "Chart A Task" includes the Google Charts Map, the task survey, and a timer. The timer will stop when the accordion section is closed and remain paused until the tasked survey is complete. Section 4 labeled "Chart B Task" functions similarly to Section 3 with the exception that it uses the D3.JS visualization tool. There are differences in the appearance of the hover technique and data presentation between Google Charts and D3.JS. Finally, Section 5 addresses the user preferences about the two visualization tools in a series of three multiple-choice questions and two short answer questions.

Figure 1. Survey Interface to Capture Basic Information: Data Visualization Survey

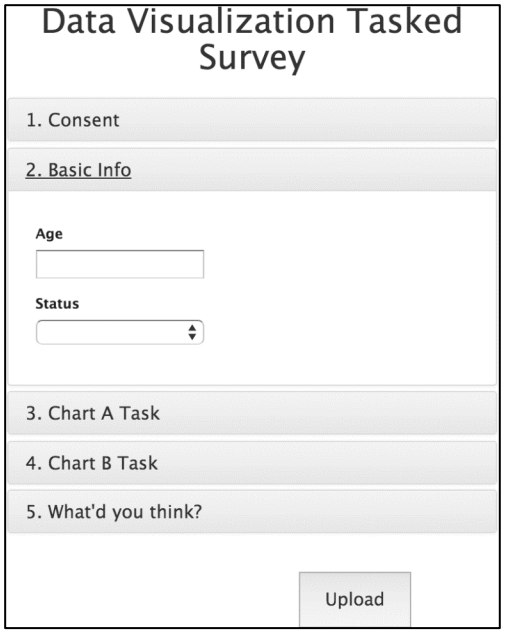


Figure 2. Chart A Task for Google Charts map

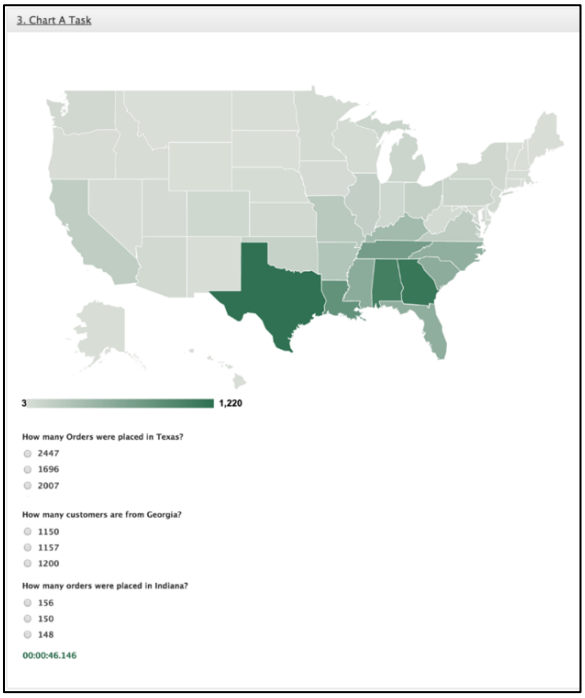
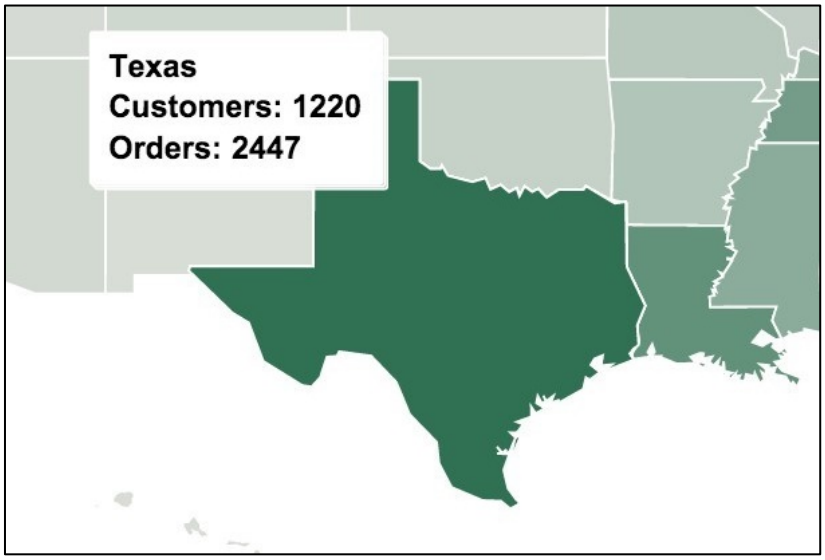


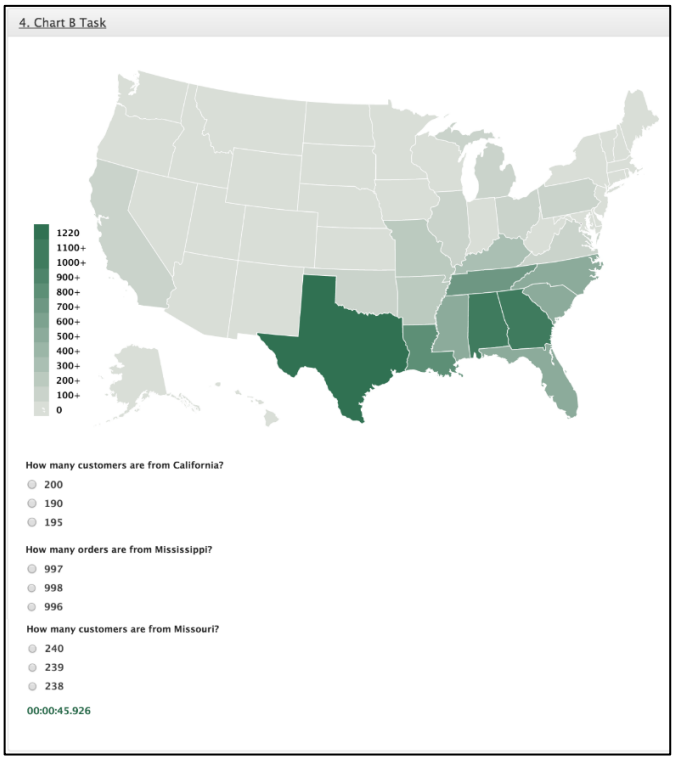
Figure 2 shows the Google Charts Map data visualization instance. The Google Charts Map is produced using the commercial but free code provided by Google. The maps allow the user to hover over each state to see how many customers and orders there have been in each state (see Figure 3).

Figure 3. The Google Charts Map's Hovered Result.



Similarly, Figure 4 shows the D3.JS map data visualization instance. The D3.JS has the capability to toggle along with its assigned region.

Figure 4. Chart B Task for D3.JS map



To provide a different look, the maps allow the user to hover over each state to see how many customers and orders there have been in each state (see Figure 5).

Figure 5. The D3.JS Map's Hovered Result

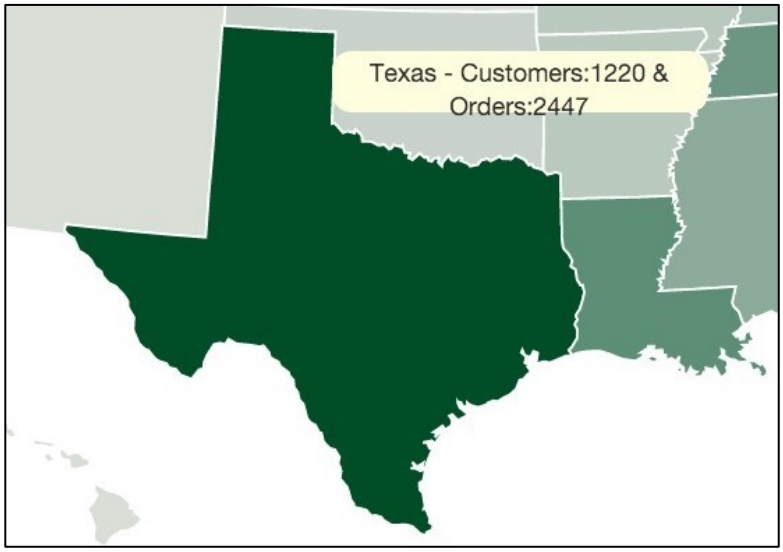
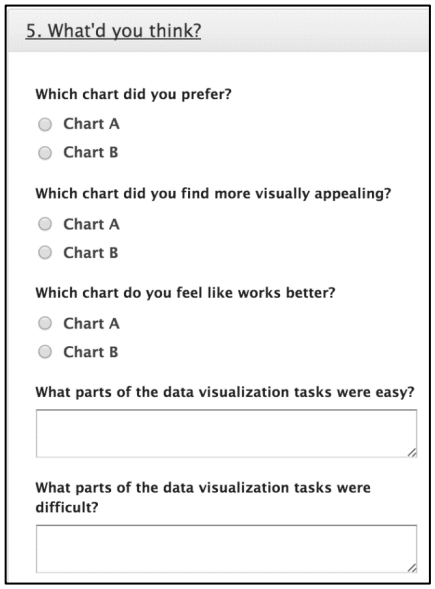


Figure 6 shows the survey used to capture the user preference information. As stated, this study attempts to evaluate the user interfaces of the two alternative visualization tools.

Figure 6. Survey for User Preference



Once the tasked survey responses are complete, the participant clicks a submit button located at the bottom of the accordion. All answers and timers are submitted to a MySQL database and a JavaScript pop-up window displays a thank you for participating message to the participant.

Experimental Procedures

Upon arrival, each participant was greeted and provided with an informed consent form. The informed consent form explained the purpose of the study, risks, benefits, and states that participation was voluntary and could be ended at any time. After completing the consent form, the participants were escorted to the computer that contained the visualizing system and provided with a copy of the task instructions. A researcher provided instructions on what to do to begin the task. The tasks were presented to the participants using pseudo random fashion across the subjects to minimize learning. For instance, if participant one received Condition 1 followed by Condition 2, then participant 2 received Condition 2 followed by 1. This pseudo random order was applied to all participants. Once the participant started, the researcher observed and recorded task completion time and number of errors.

To capture the user preference, the participants were asked three multiple-choice questions stating which chart they prefer more, which chart is more visually appealing to them, and which chart do they feel works better to them. In addition, the participants were asked to explain what parts of the data visualization tasks were easy or difficult for them. Once the participants clicked the submit button, their responses were stored in a custom database designed for this study. Later, the responses and the timestamps were imported into a MySQL table through PHP. Once the tasked survey has been submitted, a JavaScript pop-up window appeared to signal task completion. After completing the visualization task, each participant was then asked to choose their preferred subjective preference of the user interfaces. Afterward, the participant was thanked for their participation in the study.

Research Design

Participants. Participant recruitment was conducted by using telephone, email (through a class roster's emails), via Google + shared friends, using flyers, through the messaging system on the Desire2Learn (D2L) learning management system, and through verbal (word of mouth) recruitment. The target population were college students, faculty, and staff members. In total, thirty participants were recruited for this study. However, one participant was excluded from data analysis since the participant did not meet the criteria needed to be included in the research. In other words, the reason why one participant was excluded from data analysis is due to data screening (cleaning stage) as they were outliers from the rest of the groups. All the participants are young adults with ages ranging between 20 and 25. Therefore, the final results include data analysis of 29 participants.

Dataset. For this research, AKDesigns Boutique's dataset was used. This data was imported into the D3.JS and Google Charts maps. AKDesigns Boutique's data was used within the maps in the tasked survey. There are two different methods used to import the data into the maps. For the Google Charts Map, it was simpler to hard code the data into the map. AKDesigns Boutique's databases are not set-up to implement a live feed for the data to update as orders were made and customers joined. The D3.JS map used JavaScript Object Notation (JSON) to import the United States map information and a Common Separated Values (CSV) file to implement the AKDesigns Boutique data. JSON is used to allow for more customization of data visualizations (Kipp, Laa, & Cook, 2019). The CSV file was created in Excel.

RESULTS

We have conducted an IRB (institutional research board) approved usability study on the effectiveness, efficiency, and user preference of two popular data visualization tools: Google Charts and D3.JS. The performance result of one participant is an outlier because it is outside of the expected standard deviation from the mean. In other words, as stated in the research design section, one participant was excluded from data analysis is due to data screening (cleaning stage) as they were outliers from the rest of the groups. Therefore, the analysis result was conducted on 29 participants. The result is presented in terms of effectiveness, efficiency, and user satisfaction.

Effectiveness

As stated in the 'research methodology' section, the effectiveness was measured discretely as the number of successes and failures in task completion. A task was deemed successful if the participant completed the data visualization task without error. We measured the total number of errors committed as the measure for effectiveness of the data visualization tool. As stated, one of the expected outcomes of the experiment was that the participants would be able to accurately complete the data visualization task given tasks using both Google Charts (Condition-1) and D3.JS (Condition-2). The overall effectiveness result indicated that using Google Charts or D3.JS was effective. All the participants completed the task. End-users of Google Charts produced relatively more number of errors than end-users of D3.JS, 7 versus 6 respectively. However, the analysis result indicated that there is no statistical difference between the two conditions.

Efficiency

Task completion time was measured for efficiency. Since the design was a repeated measure (each participant performed both tasks), a paired sample t-test was used. The descriptive statistics and analysis are shown in Table 2 and Table 3. The mean scores for Condition-1 and Condition-2 are 43.28 and 34.66, respectively. The mean difference was approximately 8.6. Considering the p-value is less than .005, we can conclude there is significant difference in efficiency at Condition-1 and Condition-2. The descriptive statistics suggest data visualization condition could impact performance. It also suggests that Condition-2 performed better than Condition-1. However, further data analysis was required to find out the statistical significance of each of the conditions. For data visualization, Table 3 shows that there was a statistically significant difference between the two conditions ($p=0.01$, $p < 0.05$). The mean decrease in data visualization scores was 8.62 with a 95% confidence interval ranging from 3.95 to 13.30. The eta squared statistic (0.58; Cohen's $d = (34.6552 - 43.2759) / 14.92382 = 0.577647$) indicated a medium effect. The result suggests that Condition-1, took significantly longer on average to complete than Condition-2.

Table 2: Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Condition-1	43.2759	29	16.32679	3.03181
	Condition-2	34.6552	29	13.37448	2.48358

Table 3: Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Condition1 - Condition2	8.62069E0	1.22894E1	2.28209	3.94605	13.29533	3.778	28	.001

User Satisfaction

The purpose of this user preference study was used to gain knowledge of the participants' opinion of the user interface design of the visualization tools. We found that 26 of the participants expressed Google Charts was easier to use than D3.JS. In addition, they also fournd it to be more visually appealing. Hence, the final expected outcome was that, on average, the participants were more satisfied using Google Charts than D3.JS.

DISCUSSION

Key Findings and Implications

A key goal of data visualization is to enable end-users to analyze the relationships among the represented data more effectively and efficiently. Examining data visualization tools for characteristics that represent effectiveness, efficiency, and ease of use based on user preferences will highlight the significances between which tool to use when a particular characteristic is preferred. This study explores two developer tools – Google Charts and D3.JS to determine end-user preferences with respect to the presentation and representation of the data. Specifically, we address a gap in the literature for specific tool preference as well as actual user input on the preferred visual layout.

Effectiveness of a data visualization tool allows end-users to accomplish their tasks. In general, we anticipate most tools to provide effectiviness for user tasks, however we asked the question if among two data visualization tools is one more effective than the other. Our first results demonstrate that both Google Charts and D3.JS are effective. However, Google Charts is considered more effective than D3.JS. We measured the total number of errors committed by the end-users and although no statistical difference was found between the tools, participants still had fewer errors when using Google Charts. This was confirmed in a number of the comments made by participants stating things such as *"... the hover boxes in Chart A [Google Charts] were much easier to read"* and it was *"...easy to scroll over a state"*. Essentially the large text and intuitive nature of the charts, made it more effective for end-users to accomplish their task. The more clear and consistent the data visualization tool is able to present, the more effective the tool is to its end-users.

Similarly, the efficiency of a data visualization allows users to experience the quickness and speed of transforming multiple data points into a visually appealing representation. Our study found statistical significance between the two conditions ($p < 0.05$). Thus, the null hypothesis was rejected. To provide additional support, the effect size shows there is indication of a medium to large effect, with a substantial difference in the data visualization tools efficiency scores.
Finally, we found that users do have preferences towards the data visualization tool in relation to the task. In general, end-users preferred Google Charts over D3.JS because of the ease of use and aesthetic design. The font was viewed as easier to read and visibly larger. This continues to support the argument that customers not only prefer visibly appealing data representation, but have a quicker comprehension of the data. Our study continues to support the research and practical needs that key features of data visualization tools are quick adaptation of the data, useful and aestetic presentation of the data, and the end-user has preferred characteristics when engaging with the tool.

Limitations and Future Directions

This exploratory study investigated the effectiveness, efficiency, and user preference of Google Charts and D3.JS. Participants answered a series of questions based on requested tasks. As noted by Jin (2017), user's characteristics may have differing effects with respect to visualization tools. Moreover, these characteristics may direct us to understanding why certain features are preferred

over others. Hence, future research should collect and explore additional personal characteristics such as gender, race, technical expertise, and level of education.

Much of the data visualization research focuses on the comparison of specific features or the number of data points manageable by the visualization tools. Our study contributes to the literature by providing a study that explores the user view of the visualization tools and their preferred choices based on effectiveness, efficiency, and user friendliness.

Overall, while the main objective of this study was to find out whether D3.JS or Google Charts is effective in data visualization. The result indicated, on average, it took longer time for the participants to complete the task using the D3.JS option. In addition, about 90% of the participants preferred Google Charts for ease of use over D3.JS. Even though this result provides an insight for data visualization, it has several limitations. The limitation is generalizability since only thirty participants were used for the study. Diverse samples that represent the general population will be beneficial for accurate prediction.

CONCLUSION

More and more people are using data visualization tools. Google Charts and D3.JS are popular tools found to be effective and efficient. However, there is a need of user-driven reports stating the efficiency and user preference of these popular data visualization tools. This is important because it can help us understand which visualization tool could be effective under certain circumstances, as well as what characteristics are important to ease of use and user satisfaction.

We conducted a usability study on the effectiveness, efficiency, and user preferences of two popular data visualization tools. Participants were able to accurately complete the given tasks using both Google Charts and D3.JS. The research results indicated that there is a difference in ease of use between Google Charts and D3.JS. Participants believe Google Charts to be faster, more efficient, and more aesthetically pleasing than the D3.JS map even though their average times on each chart contradicts these opinions. Overall, this research highlights the significance that different visualization tools may amplify users' levels of effectiveness, efficiency and preference based on its presentation and ease of use.

REFERENCES

Bostock, M. (2019). D3. js-Data-Driven Documents.[online]. Retrieved from d3js.org

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D^3 data-driven documents. *IEEE transactions on visualization and computer graphics, 17*(12), 2301-2309.

Center, I. I. (2013). Big Data Visualization: Turning Big Data Into Big Insights. *White Paper*, 1-14. Retrieved from https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/big-data-visualization-turning-big-data-into-big-insights.pdf

Developers.google.com. (2019). Google Charts API. Retrieved from https://developers.google.com/chart

Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. *Science, 331*(6018), 705-708.

Fu, Q., Liu, W., Xue, T., Gu, H., Zhang, S., & Wang, C. (2014). *A big data processing methods for visualization.* Paper presented at the 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems.

Gesmann, M., & de Castillo, D. (2011). Using the Google visualisation API with R. *The R Journal, 3*(2), 40-44.

Hariharan, B., & Krithivasan, R. (2016). Data Visualization tools-A case study. *International Journal of Computer Science and Information Security, 14*(9), 834.

Jiang, K., Fang, Z., Ge, Y., & Zhou, Y. (2007). *Information retrieval through SVG-based vector images using an original method.* Paper presented at the IEEE International Conference on e-Business Engineering (ICEBE'07).

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research, 2*(2), 59-64.

Keahey, T. A., Rope, D. J., & Wills, G. J. (2017). Determining user preferences for data visualizations. In: Google Patents.

Kipp, M., Laa, U., & Cook, D. (2019). Connecting R with D3 for dynamic graphics, to explore multivariate data with tours. *R J., 11*(1), 245.

Luo, Y., Qin, X., Tang, N., & Li, G. (2018). *Deepeye: Towards automatic data visualization.* Paper presented at the 2018 IEEE 34th International Conference on Data Engineering (ICDE).

Martinuzzi, B. (2016). Data visualization will change the way you think about your business. In.

Metis. (2015). Data Science Training Program Launches Data Visualization With D3.js. Retrieved from https://www.thisismetis.com/?gclid=Cj0KCQjw-_j1BRDkARIsAJcfmTFx1TaDl1PYJs0HKIU_7Z_u6kR0kkwsjgOcuALKT_ZSIerP0CTv1z4aAkZ4EALw_wcB

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: A survey. *The VLDB Journal, 29*(1), 93-117.

r-project.org. (2020). The R Project for Statistical Computing. Retrieved from http://r-project.org

Rodden, K. (2014). Applying a sunburst visualization to summarize user navigation sequences. *IEEE computer graphics and applications, 34*(5), 36-40.

Sack, C. M., Donohue, R. G., & Roth, R. E. (2015). Interactive and multivariate choropleth maps with D3. *Cartographic Perspectives* (78), 57-76.

Skov, M. B., & Stage, J. (2012). Training software developers and designers to conduct usability evaluations. *Behaviour & Information Technology, 31*(4), 425-435.

Supaartagorn, C. (2016). A Framework for Web-based Data Visualization using Google charts based on MVC pattern. *Applied Science and Engineering Progress, 9*(4).

W3.org. (2017). Retrieved from https://www.w3.org/Graphics/SVG/

Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big Data Service Architecture: A Survey. *Journal of Internet Technology, 21*(2), 393-405.

Zhu, Y. (2012). Introducing google chart tools and google maps api in data visualization courses. *IEEE computer graphics and applications, 32*(6), 6-9.

This page has been intentionally left blank.

NATURAL LANGUAGE PROCESSING (NLP) APPLICATIONS IN PATIENT CARE: A SYSTEMATIC ANALYSIS

Zixing Shen, New Mexico State University

ABSTRACT

Artificial intelligence (AI) has been gaining fresh momentum with remarkable breakthroughs. AI-powered applications have been developed and deployed rapidly for various business functions across different industries. This paper focuses on natural language processing (NLP) in the context of patient care. The use of NLP in the medical field has been growing fast and drawing more and more attention. Although NLP has been dominantly used in clinical and translation research, the recent technological development and the increasing availability of patient data have provided opportunities for the direct use of NLP to patient care, the core of the medicine. This study assesses the applications of NLP in patient care. Specifically, it conducts a review of publications on how NLP is applied in patient care from 2004 to 2019. The analysis of the literature has provided interesting insights and trends, as well as gaps in the applications of NLP in patient care. This study informs the researchers and practitioners of the status quo of the NLP applications in patient care and helps stimulate research efforts that can lead to more advances in applying NLP to clinical decision support and operation.

Keywords: Natural Language Processing, NLP, NLP applications, patient care, clinical decision support

INTRODUCTION

Artificial intelligence (AI) has been gaining fresh momentum with remarkable breakthroughs. AI-powered applications have been developed and deployed rapidly for various business functions across different industries. This paper focuses on natural language processing (NLP) in the context of patient care. NLP is the automatic analysis and representation of human languages (Joshi, 1991). It is concerned with analyzing, understanding, and generating responses that ultimately make computers to interface with human languages rather than computer languages. NLP can explain a structure or a command to a computer in the natural language, written and spoken, as used by humans, translate it into a format that a computer can understand and process, and generate it back to the human user.

Over the years, NLP has become more and more sophisticated and powerful. It now has various capabilities such as content categorization, topic discovery, sentiment analysis, document summarization, and machine translation. NLP is increasingly used on large amounts of data written in plain text to get insights from texts. In addition, NLP can process languages in voice format. Examples of voice NLP applications include Xbox, Skype, and Apple's Siri. With its wide and growing usage, NLP is now included in programming languages like Python and R.

NLP has been used in the medical field for decades. Its applications in medicine have been growing fast and drawing more and more attention. Open-source NLP software tailored explicitly to the medical text, such as the clinical text analysis and knowledge extraction system (cTAKES) and the concept extraction-based text analysis system (CETAS), have been developed. NLP methods such as speech information recognition (Xiao et al., 2015), semantic labeling (Zhang et al., 2015), syntactic parsing (Sidorov et al., 2014), and negation detection (Nikfarjam et al., 2015) have been developed for medical information processing. Along with other technologies such as machine learning, NLP is applied in medical research including, but not limited to, disease classification (Reddy & Bhaskar, 2018), disease relation identification (Gu et al., 2019), gene-name recognition (Hakenberg et al., 2005), phenotype discovery (Bai et al., 2018).

Research on NLP applications in the medical field has been reviewed. Some of the reviews focus on the use of NLP in specific areas such as oncology (Yim et al., 2016), radiology (Cai et al., 2016), and chronic diseases (Sheikhalishahi et al., 2019), or in certain types of free-text data like electronic patient-authored text data (Dreisbach et al., 2019). Others synthesize the applications of NLP in biomedical research (Cohen, 2014). and the software tools used in NLP applications (Cohen et al., 2013). There are also analyses of the NLP academic research themes and scientific collaboration patterns (Chen et al., 2018). While these reviews have improved the understanding of NLP applications in the medical field, they have not provided a comprehensive picture of the NLP applications in patient care.

Patient care, at the core of the medical field, consists of testing, diagnosing, and treating patients and involves primarily doctors, nurses, clinical pharmacists, as well as patients. Although NLP has been dominantly applied in clinical and translation research, the recent technological development in NLP and the increasing availability of patient data in the format of text and image have provided opportunities for the direct use of NLP in patient care. Patient data, from various sources, such as Electronic Health Records (EHRs) and Electronic Medical Records (EMRs), as well as patients in social media, are rich in synonymy and semantically similar and related concepts. They are continuously growing in magnitude, particularly with real-time imaging and point of care devices, as well as wearable computing and mobile health technologies. The increasingly massive patient data contain a considerable amount of information valuable to clinicians and patients. However, much of such information comes in an unstructured form. NLP is thus crucial for transforming relevant unstructured information hidden in text (e.g., doctor's notes) and speech (e.g., doctor's dictates) into structured information.

Thanks to technological advances in information processing and data storage, new NLP methods and techniques, such as speech information recognition, semantic labeling, syntactic parsing, word sense disambiguation, negation detection, and temporal analysis, have emerged (Chen et al., 2018). They can meet the need for syntactic and semantic understanding of the languages, helps resolve ambiguity in language, and adds useful numeric structure to the data for many downstream applications, such as speech recognition and text analytics. As such, they enable efficient and effective analyses of vast amounts of patient data. Therefore, NLP applications can be extremely instrumental in improving clinical decision support and advancing patient care.

Given the potentials of NLP applications (Demner-Fushman et al., 2009) as well as the growing interests in AI and big data, it is of great significance to analyze the literature and understand how

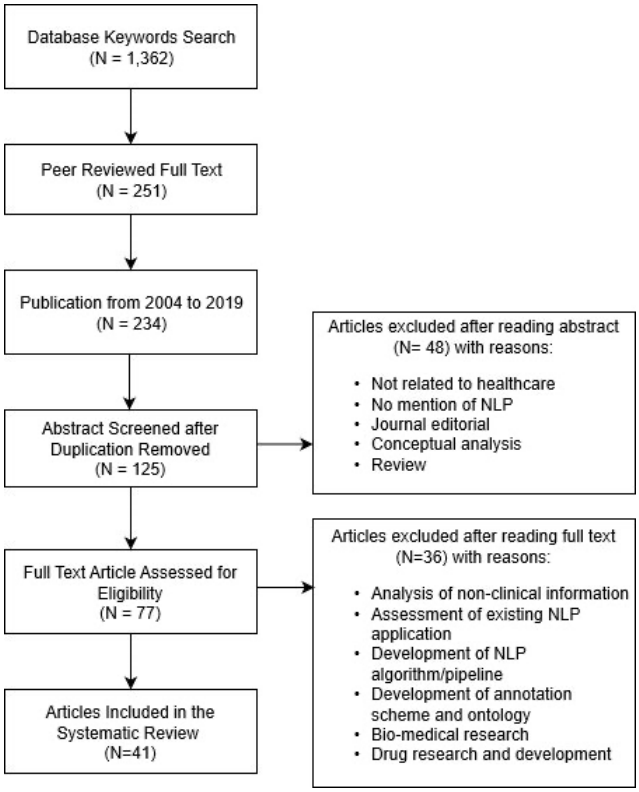
NLP has been applied in patient care. This paper answers this call. It aims to inform researchers and practitioners of the status quo of the NLP applications in patient care. This study further synthesizes and discusses the current trends and gaps across studies and proposes major directions for future research in NLP applications in patient care.

METHOD

Search Strategy and Selection Process

The PRISMA model (Liberati et al., 2009) was followed in searching the literature and selecting papers (See Figure 1). First, the EBSCO database was queried with several keywords – natural language processing, NLP, patient care, healthcare, and medicine. EBSCO is one of the largest database sources for high-quality peer-reviewed journals that publish research related to healthcare and medicine. The keyword search generated 1362 papers. Of them, 251 are peer-reviewed journal papers with full-text available in the database. Then peer-reviewed journal papers published from 2004 to 2019, the most recent one and half decades, were included. This reduced the sample size to 234. After duplicated papers were removed, 125 unique pieces remained.

Figure 1. PRISMA Flow Diagram



Next, the abstract of the 125 papers was checked. Papers that were not related to healthcare and medicine and NLP (e.g., Simon et al., 2014) were excluded. Journal editorials (e.g., Tao et al., 2017), conceptual analyses (e.g., Hope et al., 2012), and reviews (e.g., Kohane, 2011) were also removed. Forty-eight papers were eliminated after reading the abstract. Then the full text of the remaining 77 papers was examined. The following categories of articles were excluded as they did

not have any NLP application directly related to patient care: (1) analysis of medical publication text for non-clinical information (e.g., Lerchenmueller and Sorenson, 2016); (2) assessment of the effectiveness of an existing NLP method (e.g., Ferrández et al., 2012); (3) development of an NLP method (e.g., Jackson et al., 2018),; (4) development of an annotation scheme (e.g., Long et al., 2019); (5) discovery research in the biomedical field (Klann et al., 2015); (6) drug research and development (e.g., Peissig et al., 2007). Finally, a total of 41 papers (indicated with * in the reference section) with NLP applications specific to patient care were selected for coding.

Coding Procedure

A coding sheet was first developed. Special attention was paid to NLP utility (that is, what the NLP application was used for in patient care). Lexicon, extraction, and classification information was coded, as well. The software used in the extraction and the algorithm used in the classification, when applicable, were also noted. The second section of the coding sheet collected information on the data used in the NLP application, including text type, data source, and the number of texts. The last section of the coding sheet gathered information on study publication year and country affiliation. Then, each paper was read and analyzed using the coding sheet. Table 1 shows an example of the coding of a paper. The appendix lists the coding of all the 41 papers.

Table 1. Example of the Coding Sheet

Paper: Hong and colleagues, 2017	
NPL Application in Patient Care	
Clinical Utility	Screening advanced colorectal neoplasia (CRN)
Lexicon/dictionary	Custom made
Extraction	
• Method	Matching terms
• Software tool	CETAS (Concept Extraction-based Text Analysis System)
Classification	
• Approach	Statistical analysis
• Algorithm	Logistic regression
Data Used in NLP Application	
Data source	Samsung Medical Center
Text type	colonoscopy reports and pathology reports
Number of texts	49,450
Publication Year and Country	
Publication Year	2017
Country	Korea

RESULTS

Patient Care Application

Utility. A patient care application can be the use of NLP by medical professionals in the observation and treatment of actual patients and by patients in their medical journeys. It can also

be the use of NLP in the administration and management of patient care. The utilities of NLP applications in patient care were found in three areas - clinical decision support, patient care administration, and patient decision making and support, as summarized in Table 2.

Table 2. Patient Care Utility

Category	N	%
Clinical Decision Support	27	66%
Diagnose disease	9	
Detect medication/drug use	6	
Predict/identify clinical conditions and outcomes	12	
Patient Care Administration	13	32%
Quality control	3	
Patient care communication	4	
Care cost & utilization	4	
Disease prevention & control	2	
Patient Support & Decision Making	1	2%
Patient behavior and emotions		

The use of NLP for patient support and decision making was scarce (one paper only). De Silva and colleague (2018) applied NLP to investigate the interactions between clinical factors and patient emotions and behaviors in the process of diagnosis, treatment, and recovery of prostate cancer. Clinical decision support was most studied. NLP applications were tasked to provide clinical decision support in diagnosing diseases such as multiple sclerosis (Chase et al., 2017; Xia et al., 2013), influenza (Ye et al., 2017), nontuberculous mycobacterial disease (Jones et al., 2018), dementia (Shibata et al., 2018), staphylococcus aureus (Jones et al., 2012), polycystic ovary syndrome (Castro et al., 2015), depression (Parthipan et al., 2019), and diabetes (Mishra et al., 2012). NLP applications also helped clinicians in detecting medication discrepancy (Li et al., 2015), antipsychotic polypharmacy (Kadra et al., 2015), adverse drug reactions (Iqbal et al., 2015; Li et al., 2013), and identifying aspirin use (Pakhomov et al., 2010) and opioid misuse (Afshar et al., 2019). Additionally, NLP applications were effective in predicting certain medical conditions like suicidal behavior (Carson et al., 2019; Cook et al., 2016; Poulin et al., 2014; Taylor et al., 2016; Zhong et al., 2018), severe injuries after falls (Toyabe, 2012),smoking status (Regan et al., 2016), skeletal site-specific fractures (Wang, Mehrabi, et al., 2019), advanced colorectal neoplasia (Hong et al., 2017), as well as clinical outcome like patient mortality (Beeksma et al., 219; McCoy et al., 2015; Waudby-Smith et al., 2018).

NLP applications were also found useful for patient care administration in several ways. First, they improved quality control because of their ability to detect the discordance between patient self-report and documentation of symptoms in the medical record (Pakhomov et al., 2008) and assess quality in PTSD (post-traumatic stress disorder) care (Shiner et al., 2012) and imaging diagnosis (Zheng et al., 2019). They also streamlined the communication by fostering the shared meaning between clinician and patients (Balyan et al., 2019), identifying communication failures between home health nurses and physicians (Pesko et al., 2018), and providing more accurate and relevant information to patients in their searching care providers (Cook et al., 2019) and to clinicians for more tailored care (Wang, Wang, et al., 2019). Moreover, NLP applications helped with the management of clinical care cost and utilization as they predicted utilization of ADI (advanced

diagnostic imaging) (Zhang et al., 2019), identified patients at risk for readmission (Navathe et al., 2018), tracked healthcare disparities (Wieland et al., 2013) and quantified the healthcare costs and utilization for patients with binge-eating disorder (BED) (Bellows et al., 2015). Finally, NLP applications identified information for disease prevention (Workman & Stoddart, 2012) and detected concerned HIV-related messages in online forums to facilitate medical intervention (Sung et al., 2014), and therefore was useful for disease prevention and control.

Lexicon. Central to NLP is the use of standardized terminology for each concept that is of fundamental interest in a particular field. A concept is an intrinsically unique entity with an unambiguous meaning (e.g., a specific disease such as lung cancer or a symptom such as chest pain). The standardized terminologies for unique concepts are collected in lexicons. Summarized in Table 3 is the lexicon information of the reviewed literature.

Table 3. Lexicon

Category	N	%
Not reported	30	73%
Readily available (e.g., UMLS)	3	7%
Custom made	5	13%
Both	3	7%

The vast majority did not provide explicit information on the lexicon. Of the 11 papers that explicitly reported the lexicon, three used readily available lexicons - UMLS (Unified Medical Language System) (Chase et al., 2017; Li et al., 2015), SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) (Jones et al., 2018; Li et al., 2015), and RxNorm dictionary (Li et al., 2015). Customized dictionaries were either configured by referring to existing lexicons (Hong et al., 2017) or created by medical specialists (Castro et al., 2015) to meet the specific needs in processing patient data. Few studies (De Silva et al., 2018, Jones et al., 2012; Xia et al.,2013) complemented the readily available lexicons with in-house developed ones as well.

Extraction. The overall goal of NLP applications in patient care is to determine which concepts are mentioned in a medical report and in what capacity. Extraction is the first task to achieve this. In extraction, NLP analyzes the text to identify individual concepts and their modification by other terms. When extraction has been completed, each individual concept found in the text is output as a separate item in a structured format. The extraction information of the reviewed literature is summarized in Table 4. The various extraction methods can be put into two broad categories – pattern matching and linguistic analysis. Pattern matching makes use of regular expressions or sequences of characters and special symbols that explicitly define a character pattern to be searched for. For example, a list of key terms in searching the clinical notes was used to identify aspirin use (Pakhomov et al., 2010). Similarly, a regular expression approach was used to extract concepts of interest (Mishra et al., 2012).

Linguistic analyses involve using knowledge, both syntactic and semantic, to infer what concepts are mentioned and how each concept modifies other concepts. The syntactic analysis relies on the rules that control the arrangement of words in a sentence to check the text for meaningfulness. For example, an in-house tokenizer and part-of-speech tagger were developed to tokenize and parse the clinical notes (Li et al., 2015). Similarly, structural features of the clinical notes were analyzed

using sentence boundary detector, tokenizer, normalizer, part-of-speech tagger, shallow parser, etc. (Carson et al., 2019). On the other hand, semantic analysis is concerned with the knowledge regarding the different meanings of words in the context of a sentence. For example, topic modeling (Zhang et al., 2019) and neural network (Zheng et al., 2019) were used to capture semantic structures and patterns in clinical texts. Syntactic and semantic analyses were used together in one study (De Silva et al., 2018), where the syntactic analysis was used to retrieve patient demographic information and semantic analysis to extract patient-reported side effects. In addition, natural language is also based on other components such as phonetics and morphology. However, because the use of language in clinical settings is more limited than that in general text, the syntactic and semantic approaches achieve sufficient accuracy for NLP applications. This is why only one paper (Shibata et al., 2018) used morphological analysis.

As shown in Table 4, the linguistic approach was preferred over pattern matching. This may be explained by the fact that linguistic methods offer more information than pattern matching and are therefore more capable of analyzing complex concepts in patient data. The review also indicates that the syntactic analysis was the dominant linguistic approach (24 papers), while the use of semantic analysis was scant (i.e., Zhang et al., 2019; Zheng et al., 2019).

Table 4. Extraction

Category	N	%
Pattern Matching	13	32%
Linguistic Analysis	28	68%
Syntactic	24	
Semantic	2	
Hybrid	1	
Morphological analysis	1	

Presented in Table 5 is the summary of the software tools used in extraction. More than half of the papers used readily available software tools, some of which are specialized in NLP (e.g., cTAKES), while others are general-purpose tools with NLP functions (e.g., Python). The most popular NLP tool was cTAKES, followed by Python, GATE (the general architecture for text engineering), and Medline. Other readily available software tools include CETAS (Hong et al., 2017), MedLEE (Medical Language Extraction and Encoding) (Chase et al., 2017), R (Zhang et al., 2019), NLTK (Natural Language Toolkit) (Parthipan et al., 2019) among others. Customized tools were also built to fit the specific needs of NLP applications. While most tools were developed for the English language, tools for other languages were used in four papers. The software of Text Mining Studio and morphological analyzers such as Juman++ and MeCab were used in processing clinical data written in Japanese (Toyabe, 2012). Chinese segmentation tools like Jieba (Zheng et al., 2019) and CKIP (The Chinese Knowledge Information Processing) (Sung et al., 2014) were used for extraction from Chinese documents.

Table 5. Extraction Tool

Category	N	%
Not reported	5	12%
In-house developed	9	22%
Readily available	26	64%
cTAKES	5	
Python	3	
GATE	2	
Medline	2	
Others (e.g., CETAS)	14	
Both	1	2.%

Classification. Classification is performed after extraction. In classification, NLP analyzes the structured data extracted from textual data to determine whether they contain one or more desired concepts and modifiers that indicate that the textual data possess one or more specified characteristics with a given certainty (e.g., positive for a specific disease). The information on the classification approach is summarized in Table 6.

Table 6. Classification

Category	N	%
Classification not reported	1	2%
Knowledge-based rules	18	44%
Statistically inferred rules	22	54%
Regression models	7	
Bayesian models	2	
Conditional random field	2	
Word2Vec	2	
Latent Dirichlet allocations	2	
Others (e.g., random forests)	7	

Classification information was identified in all papers except one (Taylor et al., 2016). The various classification techniques fell into two broad categories – knowledge-based and statistical approach. The knowledge-based approach utilizes the real-world domain knowledge developed by domain specialists. For example, a four-step process was developed by experts for automated medication discrepancy detection (Li et al., 2015). Similarly, the decision rules were manually constructed for detecting falls (Toyabe, 2012). The statistical approach infers rules and patterns directly from data (i.e., a large corpus, like a book, a collection of sentences). Algorithms used in the statistical approach included regression models (Castro et al., 2015; Hong et al., 2017; McCoy et al., 2015; Navathe et al., 2018; Waudby-Smith et al., 2018; Xia et al., 2013; Zhang et al., 2019), Bayesian models (Chase et al., 2017; Ye et al., 2017), conditional random field (Li et al., 2013; Shiner et al., 2012), Word2Vec (Beeksma et al., 2019; De Silva et al., 2018), latent Dirichlet allocations (Afshar et al., 2019; Wang, Wang, et al., 2019), random forests (Carson et al., 2019), genetic programming (Poulin et al., 2014), genetic algorithm (Sung et al., 2014), and convolutional neural network (Zheng et al., 2019). As all these are related to machine learning, the statistical approaches in the

reviewed literature heavily used machine learning to derive insights from human languages. Machine learning can automatically learn rules and patterns but requires a large body of data to make a statistically sound inference.

As revealed in Table 6, more research relied on statistically inferred rules than knowledge-based rules in classification. The increasing popularity of the statistical approach over the knowledge-based approach can be attributed to factors like growing volumes and varieties of available medical data, computational processing that is cheaper and more powerful, and affordable data storages that enable complex mathematical calculations to deliver faster, more accurate results. In addition, compared to the statistical approach, the knowledge-based approach is more costly and time-consuming as it requires considerable manual effort to build the necessary knowledge base. Moreover, medical knowledge continually changes, and updating knowledge may also be challenging.

Data

Data source. Text data were acquired from various sources. These different data sources can be grouped into two categories – publicly available and institutional proprietary, as shown in Table 7. Examples of publicly available data sources include online support group discussion (De Silva et al., 2018), portal (Sung et al., 2014), medical publication database (Workman & Stoddart, 2012), the government agency such as Veteran Affair's Health Administration (Bellows et al., 2015; Jones et al., 2012; Jones et al., 2018; Poulin et al., 2014; Shiner et al., 2012). Intuitional proprietary data were from clinics and hospitals like Mayo Clinic (Pakhomov et al., 2008; Pakhomov et al., 2010; Wang, Mehrabi, et al., 2019; Wang, Wang, et al., 2019; Wieland et al., 2013), Shanghai Tongren Hospital (Zheng et al., 2019). One paper did not report data sources (Shibata et al., 2018). As revealed by Table 8, proprietary data were utilized more frequently than publicly available data. This was largely expected given the sensitivity of as well as the legal and regulatory issues related to clinical data.

Table 7. Data Source

Data Source	N	%
Not Reported	1	2%
Publicly available	16	39%
Intuitional Proprietary	24	59%

Text type. The types of text varied between studies. Table 8 provides a summary. The vast majority used text data from EHRs/EMRs. Some papers reported text type with such general terms as clinical notes and clinical records (Parthipan et al., 2019; Zheng et al., 2019), and others provided specific information on the unstructured data, which included physician note (Navathe et al., 2018), nurse note (Pesko et al., 2018), triage note (Zhang et al., 2019), radiology note (Wang, Mehrabi, et al., 2019), patient chart (Wieland et al., 2013), discharge summary (Mishra et al., 2012), laboratory record (Jones et al., 2012), incident report (Toyabe, 2012), imaging report and pathology report (Hong et al., 2017; Zheng et al., 2019). Text data outside EHRs and EMRs were collected and analyzed, as well. Such data included patient narrative (Shibata et al., 2018), drug label (Li et al., 2013), and the patient provided information (Pakhomov et al., 2008), text message

(Cook et al., 2016), online discussion post (De Silva et al., 2018), state licensure list (Cook et al., 2019), and medical publication (Workman & Stoddart, 2012).

Table 8. Text Type

Text Type	N	%
EHRs & EMRs	33	81%
Other (e.g., online post)	7	17%
Both	1	2%

The number of texts. The number of texts used for NLP applications ranged from 42 patient narratives (Shibata et al., 2018) to 4,795,428 online posts (De Silva et al., 2018). Table 9 summarizes the distribution of numbers of texts. Datasets smaller than 10,000 were analyzed in nearly 60% of the papers. Very few studies (Afshar et al., 2019; Balyan et al., 2019; Beeksma et al., 2019; De Silva et al., 2018; Zhang et al., 2019) worked with datasets larger than 100,000.

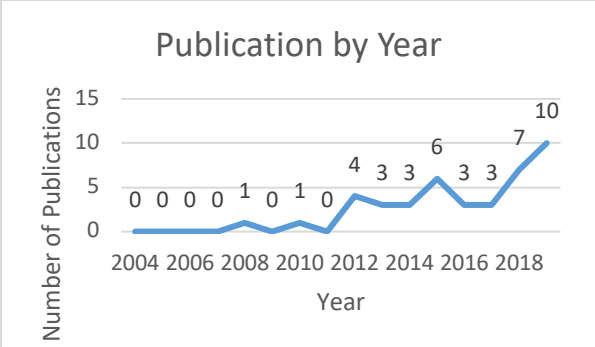
Table 9. Number of Texts

Number of Texts	N	%
N < 1,000	9	21%
1,000 < N < 10,000	15	37%
10,000 < N < 100,000	12	30%
N > 100,000	5	12%

Publication Year and Country

Publication year. As shown in Figure 2, forty studies were published after 2010 and only one (Pakhomov et al., 2008) before 2010. Despite some year to year fluctuation, the interest in NLP applications in patient care has been growing steadily over the past 15 years, especially since 2011. There was a big jump in the last three years, from 2016 to 2019.

Figure 2. Publication by Year



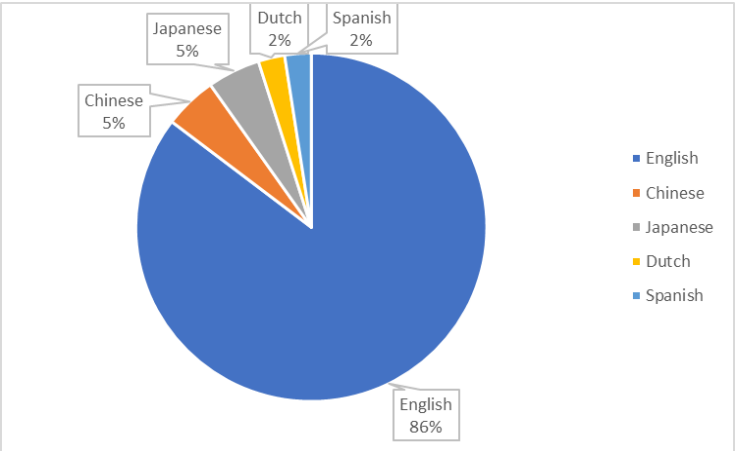
Country. As shown in Table 10, the studies were conducted in nine different countries, with the USA leading the list (29 papers). As shown in Figure 3, English was the most analyzed language (35 papers), and other languages, including Chinese (Sung et al., 2014; Zhang et al., 2019),

Japanese (Shibata et al., 2018; Toyabe, 2012), Dutch (Beeksma et al., 2019), and Spanish (Cook et al., 2016), were also studied.

Table 10. Publication by Country

Country	N	%
USA	29	72%
UK	3	8%
China	2	5%
Japan	2	5%
Australia	1	2%
Canada	1	2%
Korea	1	2%
Netherlands	1	2%
Spain	1	2%

Figure 3. Languages in NLP



DISCUSSION

NLP Application

The study has revealed the paucity of NLP applications for patient support and decision-making. The amount of online patient-related texts from social media posts and themed forums is overwhelming and could lead to a significant increase in NLP applications targeting patient support and decision-making. One barrier could be the lack of lexicons for patient-authored text data. Patients use everyday language, not technical, medical terms as clinical professionals. The lexicons for the language used in formal settings (hospitals and clinics) do not work for the language used in informal communities (Twitter and online discussion groups). As such, future research is needed to develop lexicons to help generate patient vocabularies, match them with standard terminologies, and annotate social medial and web communities with medical terms.

Most NLP applications have been used for clinical decision support. Upon further examination of these studies, it is found that NLP was employed for passive clinical decision support that requires input by the user to generate output. Future research can investigate how to move NLP applications from passive to active decision support. Active NLP decision support can push patient-specific information to users. The findings from many studies in the reviewed papers can be capitalized to build active NLP systems. For example, the automatic identification of adverse drug reactions (Li et al., 2013) can become the input of a clinical decision support system that alerts prescribers in real-time to medication problems. Likewise, information on suicidal ideation (Cook et al., 2016) generated from textual data can inform timely clinical interventions to prevent suicide. The use case for patient care administration can go beyond what the existing literature has found. For example, NLP's ability to review medical records more quickly and thoroughly than traditional, manual programs can dramatically increase efficiency and accuracy in patient care documentation. Future NLP research on patient care documentation is needed to help realize the full potential of NLP in transforming patient care operations.

The study has shown the dominance of syntactic representation of text as well as the lack of semantic-based NLP in clinical care. The syntactic analysis, though powerful, can process only information that it can "see" in the text. Such limitation can be overcome in the semantic analysis, which can identify a cascade of concepts related to the words in the text. Therefore, the semantic analysis can support more complex NLP tasks such as word-sense disambiguity, textual entailment, and semantic role labeling (Cambria and White, 2014), all of which are very relevant in understanding the patient data. The semantic analysis warrants more future research to harvest more valuable information from unstructured clinical texts.

Readily available extraction tools, many of which are open-sourced, are widely used in NLP applications in patient care. These general-purpose tools cover a broad set of needs and produce outcomes that are easy to interpret. On the other hand, the phrase structure used by these tools (for example, "colon and lung cancer" may be identified as "colon" and "lung cancer" rather than "colon cancer" and "lung cancer") inherently reduces the possibility to extract complex medical information from text. The in-house developed tools are successful in addressing specific needs, but they have a relatively narrow focus tuned highly to the data at particular institutions, and therefore, low in transferability (Kreimeyer et al., 2017). How to provide versatile NLP flatforms with available pipelines for many specific clinical tasks, simple or complex, is the methodology challenge to be addressed to move NLP systems toward full-scale acceptance and routine use in patient care settings.

Statistical NLP has become more popular than rule-based NLP. A variety of machine learning algorithms were adopted in the reviewed studies. However, only one paper (Zheng et al., 2019) made use of the method of deep learning, the most exciting recent development in machine learning. Given the promise of deep learning for text processing (LeCun et al., 2015), how to apply deep learning to solve clinical care problems is another avenue for future research.

Data

This review has revealed that the availability of public datasets remains limited. The need remains for data availability and accessibility that would increase participation in NLP for patient care. Furthermore, the review has revealed that most existing research has worked on small numbers of texts. As previously mentioned, statistical NLP nowadays relies heavily on machine learning algorithms, which require enormous amounts of data to make valid statistical inferences. It is clear that future research needs to work on more sizable data for more robust models and more powerful outcomes. Finally, all papers in this review analyzed textual data. That is, voice data are not studied at all. The recent breakthroughs in speech recognition make voice NLP valuable to patient care. For example, patient care providers can implement chatbots to automate some of their communications with their patients to improve their operational efficiency. There is also a promise in the use of speech NLP for information entry at the bedside. In short, future research needs to consider the impact of voice data on NLP applications in patient care.

Publication Year and Country

The review has shown a surge of research publications in the last decade, which is aligned with the rapid developments of artificial intelligence technologies in recent years. As AI technologies have been witnessing tremendous and rapid advancements in recent years, NLP applications will become more sophisticated and make greater inroads into patient care. More research will continue to investigate the profound impacts of NLP applications on patient care. It would be interesting to see more research on patient care NLP applications from research groups outside the USA and to develop more NLP tools for languages other than English.

Limitations

The rate of publication in the field of patient care NLP applications is fast increasing, and literature-based reviews like this one may have difficulty in keeping up with new developments. Nevertheless, this review provides a general scope of NLP applications in patient care. The study has limitations as well. First, as only journal papers with full-text available were analyzed, it is possible that the information from the journal papers which were not accessible at the time of the review was not gathered. In addition, this review excluded conference papers. As a result, this review may miss the latest development in the dynamic and fast-paced research field. Though the study is representative of patient NLP applications in the last 15 years, its findings may not capture every aspect of the area. A further review may want to include more databases and more types of papers to get a more comprehensive understanding of the applications of NLP in patient care.

CONCLUSION

As healthcare goes digital, NLP has become indispensable in medicine. Although NLP has been dominantly applied in clinical and translation research, usually through phenotyping, recent developments in NLP methods and techniques provide opportunities to represent patient care knowledge and drive clinical decision support and operation. The direct use of NLP is making inroads into clinical decision support and patient care. This paper presents a systematic analysis of research on NLP applications in patient care in the last 15 years. It has provided encouraging

results that NLP is improving the delivery and management of patient care when compared with traditional methods. This paper helps information systems (IS) researchers understand the status quo of NLP in the domain of patient care. Equally important, it highlights important areas of NLP-powered research to move the field forward. NLP applications in patient care are promising, and much more systematic research is needed for the practical or commercial implementation of NLP in patient care.

REFERENCES

Afshar, M., Joyce, C., Dligach, D., Sharma, B., Kania, R., Xie, M., Swope, K., Salisbury-Afshar, E., & Karnik, N. S. (2019). Subtypes in patients with opioid misuse: A prognostic enrichment strategy using Electronic Health Record data in hospitalized patients. *PLoS ONE*, *14*(7), e0219717.

Bai, T., Chanda, A. K., Egleston, B. L., & Vucetic, S. (2018). EHR phenotyping via jointly embedding medical concepts and words into a unified vector space. *BMC Medical Informatics and Decision*, *18*(S 4), article no.123.

Balyan, R., Crossley, S. A., Brown, W., Karter, A. J., McNamara, D. S., Liu, J. Y., Lyles, C. R., & Schillinger, D. (2019). Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPSE study. *PLOS ONE*, *14*(2), e0212488.

Beeksma, M., Verberne, S., van den Bosch, A., Das, E., Hendrickx, I., & Groenewoud, S. (2019). Predicting life expectancy with a long short-term memory recurrent neural network using Electronic Medical Records. *BMC Medical Informatics and Decision Making*, *19*(1), article no. 36.

Bellows, B. K., DuVall, S. L., Kamauu, A. W. C., Supina, D., Babcock, T., & LaFleur, J. (2015). Healthcare costs and resource utilization of patients with binge-eating disorder and eating disorder not otherwise specified in the Department of Veterans Affairs: Healthcare costs and utilization in binge eating. *International Journal of Eating Disorders, 48*(8), 1082-1091.

Cai, T., T Giannopoulos, A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K., Rybicki, F., & Mitsouras, D. (2016). Natural language processing technologies in radiology research and clinical applications. *Radiographics*, *36*(1), 176-191.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, *9*(2), 48-57.

Carson, N. J., Mullin, B., Sanchez, M. J., Lu, F., Yang, K., Menezes, M., & Cook, B. L. (2019). Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of Electronic Health Records. *PLOS ONE*, *14*(2), e0211116.

Castro, V., Shen, Y., Yu, S., Finan, S., Pau, C. T., Gainer, V., Keefe, C. C., Savova, G., Murphy, S. N., Cai, T., & Welt, C. K. (2015). Identification of subjects with polycystic ovary syndrome using Electronic Health Records. *Reproductive Biology and Endocrinology, 13*(1), article no. 116.

Chase, H. S., Mitrani, L. R., Lu, G. G., & Fulgieri, D. J. (2017). Early recognition of multiple sclerosis using natural language processing of the Electronic Health Record. *BMC Medical Informatics and Decision Making, 17*(1), article no. 24.

Chen, X., Xie, H., Wang, F. L., Liu, Z., Xu, J., & Hao, T. (2018). A bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making*, *18*(s1), article no. 14.

Cohen, K. B. (2014). Chapter 6: Biomedical natural language processing and text mining, in I. N. Sarkar (Ed) *Methods in Biomedical Informatics: A Pragmatic Approach* (pp.141-177). Academic Press.

Cohen, R., Elhadad, M., & Birk, O. (2013). Analysis of free online physician advice services. *PLoS ONE*, *8*(3), e59963.

Cook, B. L., Progovac, A. M., Chen, P., Mullin, B., Hou, S., & Baca-Garcia, E. (2016). Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Computational and Mathematical Methods in Medicine*, 1-8.

Cook, M. J., Yao, L., & Wang, X. (2019). Facilitating accurate health provider directories using natural language processing. *BMC Medical Informatics and Decision Making*, *19*(S3), article no. 80.

De Silva, D., Ranasinghe, W., Bandaragoda, T., Adikari, A., Mills, N., Iddamalgoda, L., Alahakoon, D., Lawrentschuk, N., Persad, R., Osipov, E., Gray, R., & Bolton, D. (2018). machine learning to support social media empowered patients in cancer care and cancer treatment decisions. *PLOS ONE, 13*(10), e0205855.

Demner-Fushman, D., Chapman, W.W., & McDonald C. J. (2009). What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics*, *42*(5), 760-772.

Dreisbach, C, Koleck, TA, Bourne, P.E., & Bakken, S. (2019). A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International Journal of Medical Informatics*, *125*, 37-46.

Ferrández, O., South, B. R., Shen, S., Friedlin, F. J., Samore, M. H., & Meystre, S. M. (2012). Evaluating current automatic de-identification methods with veteran's health administration clinical documents. *BMC Medical Research Methodology, 12*(1), article no. 109.

Gu, J, Sun, F, Qian, L, & Zhou, G. (2019). Chemical-induced disease relation extraction via attention-based distant supervision. *BMC Bioinformatics*, *20*, article no. 403.

Hakenberg, J, Bickel, S, Plake, C, Brefeld, U., Zahn, H., Faulstich, L., Leser, U., & Scheffer, T. (2005). Systematic feature evaluation for gene name recognition. *BMC Bioinformatics*, *6*, article no. S9.

Hong, S. N., Son, H. J., Choi, S. K., Chang, D. K., Kim, Y.-H., Jung, S.-H., & Rhee, P.-L. (2017). A Prediction model for advanced colorectal neoplasia in an asymptomatic screening population. *PLOS ONE, 12*(8), e0181040.

Hope, C. J., Garvin, J. H., & Sauer, B. C. (2012). Information extraction from narrative data. *American Journal of Health-System Pharmacy*, *69*(6), 455-461.

Iqbal, E., Mallah, R., Jackson, R. G., Ball, M., Ibrahim, Z. M., Broadbent, M., Dzahini, O., Stewart, R., Johnston, C., & Dobson, R. J. B. (2015). identification of adverse drug events from free-text electronic patient records and information in a large mental health case register. *PLOS ONE, 10*(8), e0134208.

Jackson, R., Kartoglu, I., Stringer, C., Gorrell, G., Roberts, A., Song, X., Wu, H., Agrawal, A., Lui, K., Groza, T., Lewsley, D., Northwood, D., Folarin, A., Stewart, R., & Dobson, R. (2018). CogStack - Experiences of deploying integrated information retrieval and extraction services in a large national health service foundation trust hospital. *BMC Medical Informatics and Decision Making, 18*(1), article no. 47.

Jones, M. M., DuVall, S. L., Spuhl, J., Samore, M. H., Nielson, C., & Rubin, M. (2012). Identification of methicillin-resistant staphylococcus aureus within the nation's Veterans Affairs medical centers using natural language processing. *BMC Medical Informatics and Decision Making, 12*(1), article no. 34.

Jones, M. M., Winthrop, K. L., Nelson, S. D., Duvall, S. L., Patterson, O. V., Nechodom, K. E., Findley, K. E., Radonovich, L. J., Samore, M. H., & Fennelly, K. P. (2018). Epidemiology of nontuberculous mycobacterial infections in the U.S. Veterans Health Administration. *PLOS ONE, 13*(6), e0197976.

Joshi, A. (1991). Natural Language Processing. *Science*, 253(5025), 1242-1249.

Kadra, G., Stewart, R., Shetty, H., Jackson, R. G., Greenwood, M. A., Roberts, A., Chang, C.-K., MacCabe, J. H., & Hayes, R. D. (2015). Extracting antipsychotic polypharmacy data from Electronic Health Records: Developing and evaluating a novel process. *BMC Psychiatry, 15*(1), article no. 166.

Klann, J. G., Phillips, L. C., Turchin, A., Weiler, S., Mandl, K. D., & Murphy, S. N. (2015). A numerical similarity approach for using retired current procedural terminology (CPT) codes for electronic phenotyping in the scalable collaborative infrastructure for a learning health system (SCILHS). *BMC Medical Informatics and Decision Making, 15*(1), article no. 104.

Kohane, I. S. (2011). Using Electronic Health Records to Drive Discovery in Disease Genomics. *Nature Reviews Genetics*, *12*(6), 417-428.

Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., Forshee, R., Walderhaug, M., & Botsis, T. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *Journal of Biomedical Informatics*, *73*, 14-29.

LeCun, Y., Bengio, Y., and Hinton, G. (2015, May). *Nature*, 521, 436-444.

Lerchenmueller, M. J., & Sorenson, O. (2016). Author disambiguation in PubMed: Evidence on the precision and recall of authority among NIH-funded scientists. *PLOS ONE*, *11*(7), e0158731.

Li, Q., Deleger, L., Lingren, T., Zhai, H., Kaiser, M., Stoutenborough, L., Jegga, A. G., Cohen, K. B., & Solti, I. (2013). Mining FDA drug labels for medical conditions. *BMC Medical Informatics and Decision Making, 13*(1), article no. 53.

Li, Q., Spooner, S. A., Kaiser, M., Lingren, N., Robbins, J., Lingren, T., Tang, H., Solti, I., & Ni, Y. (2015). An end-to-end hybrid algorithm for automated medication discrepancy detection. *BMC Medical Informatics and Decision Making, 15*(1), article no. 37.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine*, 339: b2700.

Long, H., Zhu Y., Jia L., Gao B., Liu J., Liu L., & Herre H. (2019). An ontological framework for the formalization, organization, and usage of TCM-knowledge. *BMC Medical Informatics and Decision Making*, *19*(S2), article no. 53.

McCoy, T. H., Castro, V. M., Cagan, A., Roberson, A. M., Kohane, I. S., & Perlis, R. H. (2015). Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: An Electronic Health Record study. *PLOS ONE, 10*(8), e0136341.

Mishra, N. K., Son, R. Y., & Arnzen, J. J. (2012). Towards automatic diabetes case detection and ABCS protocol compliance assessment. *Clinical Medicine & Research*, *10*(3), 106-121.

Navathe, A. S., Zong, F., Lie, V. J., Chang, F. Y., Sordo, M., Topaz, M., Navathe, S. B., & Rocha, R. A. (2018). Hospital readmission and social risk factors identified from physician notes. *Health Services Research*, *53*(2), 1110-1136.

Névéol A, & Zweigenbaum P. (2016). Clinical natural language processing in 2015: Leveraging the variety of texts of clinical interest. *Yearbook of Medical Informatics*, *10*(1), 234-239.

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of American Medical Informatics Association*, *22*(3), 671-81.

Pakhomov, S. V., Jacobsen, S. J., Chute, C. G., & Roger, V. L. (2008). Agreement between patient-reported symptoms and their documentation in the medical record. *The American Journal of Managed Care, 14*(8), 530-539.

Pakhomov, S., Shah, N., Hanson, P., Balasubramaniam, S., & Smith, S. (2010). Automated processing of Electronic Medical Records is a reliable method of determining aspirin use in populations at risk for cardiovascular events. *Journal of Innovation in Health Informatics, 18*(2), 125-133.

Parthipan, A., Banerjee, I., Humphreys, K., Asch, S. M., Curtin, C., Carroll, I., & Hernandez-Boussard, T. (2019). Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. *PLOS ONE, 14*(2), e0210575.

Peissig, P., Sirohi, E., Berg, R.L., Brown-Switzer, C., Ghebranious, N., McCarty, C. A., & Wilke, R.A. (2007). Construction of atorvastatin dose-response relationships using data from a large population-based DNA Biobank. *Basic & Clinical Pharmacology & Toxicology*, *100*, 286-288.

Pesko, M. F., Gerber, L. M., Peng, T. R., and Press, M. J. (2018). Home health care: Nurse – physician communication, patient severity, and hospital readmission. *Health Services Research*, *53*(2):1008-1024.

Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goertzel, B., Watts, B., Flashman, L., & McAllister, T. (2014). Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE, 9*(1), e85733.

Reddy, E. M., & Bhaskar, P. (2018). Able machine learning method for classifying disease-treatment semantic relations from bio-medical sentences. *International Journal of Recent Research Aspects*, *5*(1), 223-226.

Regan, S., Meigs, J. B., Grinspoon, S. K., & Triant, V. A. (2016). Determinants of smoking and quitting in HIV-infected individuals. *PLOS ONE, 11*(4), e0153103.

Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: Systematic review. *JMIR Medical Informatics*, *7*(2), e12239.

Shibata, D., Ito, K., Nagai, H., Okahisa, T., Kinoshita, A., & Aramaki, E. (2018). Idea density in Japanese for the early detection of dementia based on narrative speech. *PLOS ONE, 13*(12), e0208418.

Shiner, B., D'Avolio, L. W., Nguyen, T. M., Zayed, M. H., Watts, B. V., & Fiore, L. (2012). Automated classification of psychotherapy note text: Implications for quality assessment in PTSD care: NLP of PTSD psychotherapy notes. *Journal of Evaluation in Clinical Practice, 18*(3), 698-701.

Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic n-grams as machine learning features for natural language processing. *Expert Systems with Applications*, *41*(3), 853-860.

Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., & Adini, B. (2014). Twitter in the crossfire - The use of social media in the Westgate Mall terror attack in Kenya. *PLoS ONE*, *9*(8), e104136.

Sung, R.-J., Chiu, C., Chiu, N.-H., & Hsiao, C.-H. (2014). Online detection of concerned HIV-Related messages in web forums. *AIDS Care*, *26*(3), 337-342.

Tao, C., Gong, Y., Xu, H., & Zhao, Z. (2017). Introduction: The International Conference on Intelligent Biology and Medicine (ICIBM) 2016: Special focus on medical informatics and big data. *BMC Medical Informatics and Decision Making*, *17*(S2), article no. 77.

Taylor, C. L., van Ravesteyn, L. M., van denBerg, M. P. L., Stewart, R. J., & Howard, L. M. (2016). The prevalence and correlates of self-harm in pregnant women with psychotic disorder and bipolar disorder. *Archives of Women's Mental Health, 19*(5), 909-915.

Toyabe, S. (2012). Detecting inpatient falls by using natural language processing of Electronic Medical Records. *BMC Health Services Research, 12*(1), article no. 448.

Wang, L., Wang, Y., Shen, F., Rastegar-Mojarad, M., & Liu, H. (2019). Discovering associations between problem list and practice setting. *BMC Medical Informatics and Decision Making*, *19*(S3), article no. 69.

Wang, Y., Mehrabi, S., Sohn, S., Atkinson, E. J., Amin, S., & Liu, H. (2019). Natural language processing of radiology reports for identification of skeletal site-specific fractures. *BMC Medical Informatics and Decision Making*, *19*(S3), article no. 73.

Waudby-Smith, I. E. R., Tran, N., Dubin, J. A., & Lee, J. (2018). Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. *PLOS ONE, 13*(6), e0198687.

Wieland, M. L., Wu, S. T., Kaggal, V. C., & Yawn, B. P. (2013). Tracking health disparities through natural-language processing. *American Journal of Public Health, 103*(3), 448-449.

Workman, T. E., & Stoddart, J. M. (2012). Rethinking information delivery: Using a natural language processing application for point-of-care data discovery. *Journal of the Medical Library Association: JMLA, 100*(2), 113-120.

Xia, Z., Secor, E., Chibnik, L. B., Bove, R. M., Cheng, S., Chitnis, T., Cagan, A., Gainer, V. S., Chen, P. J., Liao, K. P., Shaw, S. Y., Ananthakrishnan, A. N., Szolovits, P., Weiner, H. L., Karlson, E. W., Murphy, S. N., Savova, G. K., Cai, T., Churchill, S. E., Plenge, R. M., Kohane, I. S., & De Jager, P. L. (2013). Modeling disease severity in multiple sclerosis using Electronic Health Records. *PLoS ONE, 8*(11), e78927.

Xiao, B, Imel, Z. E., Georgiou, P. G., Atkins, D. C., Narayanan, S. S. (2015). Rate my therapist: Automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLoS One*, *10*(12), e0143055.

Ye, Y., Wagner, M. M., Cooper, G. F., Ferraro, J. P., Su, H., Gesteland, P. H., Haug, P. J., Millett, N. E., Aronis, J. M., Nowalk, A. J., Ruiz, V. M., López Pineda, A., Shi, L., Van Bree, R., Ginter, T., & Tsui, F. (2017). A study of the transferability of influenza case detection systems between two large healthcare systems. *PLOS ONE, 12*(4), e0174970.

Yim, W, Yetisgen M, Harris W.P., & Kwan S.W. (2016). Natural language processing in oncology: A review. JAMA Oncology, *2*, 1-8.

Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Chokshi, F. H., & Schrager, J. D. (2019). Advanced diagnostic imaging utilization during emergency department visits in the United States: A predictive modeling study for emergency department triage. *PLOS ONE*, *14*(4), e0214905.

Zhang, Y. Y., Tang, B. Z., Jiang, M., Wang, J. Q., & Xu, H. (2015). Domain adaptation for semantic role labeling of clinical text. *Journal of American Medical Informatics Association*, *22*(5), 967-79.

Zheng, T., Gao, Y., Wang, F., Fan, C., Fu, X., Li, M., Zhang, Y., Zhang, S., & Ma, H. (2019). Detection of medical text semantic similarity based on convolutional neural network. *BMC Medical Informatics and Decision Making*, *19*(1), article no. 156.

Zhong, Q.-Y., Karlson, E. W., Gelaye, B., Finan, S., Avillach, P., Smoller, J. W., Cai, T., & Williams, M. A. (2018). Screening pregnant women for suicidal behavior in Electronic Medical Records: Diagnostic codes vs. clinical notes processed by natural language processing. *BMC Medical Informatics and Decision Making, 18*(1), article no. 30.

Appendix: Summary of coding

	Application in Patient Care					Data			Publication	
Paper	**Clinical Utility**	**Lexicon**	**Extraction**	**Extraction Tool**	**Classification /Algorithm**	**Source**	**Type**	**No. of Texts**	**Year**	**Country**
Afshar et al. (2019)	detect opioid misuse	not reported	syntactic analysis	cTAKES	statistical - latent Dirichlet allocation	an urban tertiary academic center	clinical note	422,147	2019	USA
Balyan et al. (2019)	foster shared meaning	not reported	syntactic analysis	TAALES, TAACO, etc.	statistical - support vector machines, etc.	KPNC Diabetes Registry	secure message	283,216	2019	USA
Beeksma et al. (2019)	predict patient mortality	not reported	pattern matching	in-house developed	statistical - Word2Vec	FaMe-net repository	Consultation document	149,314	2019	Nether-land
Bellows et al. (2015)	quantify costs & utilization for binge-eating disorder patients	not reported	pattern matching	not reported	knowledge-based	A VA clinic	clinical note	1,487	2015	USA
Carson et al. (2019)	predict suicidal behavior	not reported	syntactic analysis	Invenio	statistical - random forest	a community health system	clinical note	9,415	2019	USA
Castro et al. (2015)	diagnose polycystic ovary syndrome	custom made	syntactic analysis	cTAKES	statistical - regression	Partners Healthcare Research Patients Data Registry	clinic note	640	2015	USA
Chase et al. (2017)	diagnose multiple sclerosis	UMLS	syntactic analysis	MedLEE	statistical - Bayesian	Columbia University Medical Center	clinical note	2,999	2017	USA
Cook et al. (2016)	predict suicidal behavior	not reported	syntactic analysis	in-house developed	statistical - LIBLINEAR	a hospital system in Madrid, Spain	text message	1,453	2016	Spain
Cook et al. (2019)	provide provider information to patients	not reported	pattern matching	in-house developed	knowledge-based	Connecticut eLicensing website, the National Plan Provider Enumeration System	state licensure list	7,408	2019	USA
De Silva et al. (2018)[0]	understand patient behavior and emotion	both	hybrid	Python	statistical - Word2Vec	online support group	online discussion post	4,795,428	2018	Australia
Hong et al. (2017)	predict advanced colorectal neoplasia	custom made	pattern matching	CETAS	statistical - regression	Samsung Medical Center	Colonoscopy, pathology report	49,450	2017	Korea
Iqbal et al. (2015)	detect adverse drug reactions	not reported	pattern matching	in-house developed	statistical - Java Annotation Patterns Engine	South London and Maudsley (SLaM)	clinical record	15,908	2015	UK
Jones et al. (2012)	diagnose staphylococcus aureus	both	pattern matching	in-house developed	knowledge-based	VA network of databases	microbiology record	68,427	2012	USA
Jones et al. (2018)	diagnose Nontuberculous mycobacterial disease	SNOMED CT	pattern matching	not reported	knowledge-based	VA Corporate Data Warehouse	laboratory data	6,031	2018	USA
Kadra et al. (2015)	detect antipsychotic polypharmacy	not reported	syntactic analysis	in-house developed	knowledge-based	South London and Maudsley (SLaM)	diagnostic data	7,201	2015	UK
Li et al. (2013)	detect adverse drug reactions	not reported	syntactic analysis	cTAKES	statistical - conditional random field	Cincinnati Children's Hospital	drug label	38,071	2013	USA
Li et al. (2015)	detect medication discrepancy	UMLS, SNOMED CT, RxNorm	syntactic analysis	both	knowledge-based	Cincinnati Children's Hospital	clinical note, discharge prescription list	975	2015	USA
McCoy et al. (2015)	predict patient mortality	not reported	syntactic analysis	Python	statistical - regression	a large New England health system	discharge note	23,343	2015	USA
Mishra et al. (2012)	diagnose diabetes	not reported	pattern matching	ConText	knowledge-based	i2b2 data set	discharge summary	889	2012	USA
Navathe et al. (2018)	identify readmission risk	custom made	syntactic analysis	MTERMS	statistical - regression	Partners Healthcare System	physician note	93,606	2018	USA
Pakhomov et al. (2008)	detect the discordance between self-report and documented symptoms	not reported	pattern matching	in-house developed	knowledge-based	Mayo Clinic	clinical note, patient form	1,119	2008	USA
Pakhomov et al. (2010)	detect aspirin use	not reported	pattern matching	Perl	knowledge-based	Mayo Clinic	clinical note	499	2010	USA

Parthipan et al. (2019)	diagnose depression	not reported	syntactic analysis	NLTK	knowledge-based	Stanford University Hospitals and Clinical	clinical note	41,713	2019	USA
Pesko et al. (2018)	identify communication failures	custom made	syntactic analysis	not reported	knowledge-based	Medicare Provider Analysis and Review file, VNSNY electronic medical records	nurse note	2,680	2018	USA
Poulin et al. (2014)	predict suicidal behavior	not reported	syntactic analysis	not reported	statistical - genetic programming	VA medical records	clinical note	210	2014	USA
Regan et al. (2016)	identify smoking status	not reported	syntactic analysis	GATE	knowledge-based	Partners Healthcare System	clinic note	3,487	2016	USA
Shibata et al. (2018)	diagnose dementia	not reported	morpho-logical	Juman++, MeCab	knowledge-based	Not reported	patient narrative	42	2018	Japan
Shiner et al. (2012)	assess quality in PTSD care	not reported	syntactic analysis	in-house developed	statistical - conditional random field	A VA clinic	mental health note	221	2012	USA
Sung et al. (2014)	facilitate intervention for HIV	not reported	syntactic analysis	CKIP	statistical - genetic algorithm	Yahoo! Chinese-based portal	online post	2,083	2014	China
Taylor et al. (2016)	predict suicidal behavior	not reported	syntactic analysis	GATE	not reported	South London and Maudsley (SLaM)	clinical record	420	2016	UK
Toyabe (2012)	predict injuries after falls	not reported	syntactic analysis	Text Mining Studio	knowledge-based	Niigata University Hospital, Japan	incident report	2,590	2012	Japan
Wang, Mehrabi et al. (2019)	identify skeletal site-specific fractures	not reported	syntactic analysis	Medline	statistical - latent Dirichlet allocation	Mayo clinic	radiology note	2,356	2019	USA
Wang, Wang et al. (2019)	provide more information to clinicians	not reported	pattern matching	Med-Tagger	knowledge-based	Mayo clinic	clinical note	64,250	2019	USA
Waudby-Smith et al. (2018)	predict patient mortality	not reported	syntactic analysis	Python	statistical - regression	MIMIC-III	nurse note	27,477	2018	Canada
Wieland et al. (2013)	tracked healthcare disparities	not reported	pattern matching	not reported	knowledge-based	Mayo Clinic	patient chart	5,782	2013	USA
Workman & Stoddart (2012)	identified information for disease prevention	not reported	syntactic analysis	Medline	knowledge-based	PubMed	PubMed citation	3,276	2012	USA
Xia et al. (2018)	diagnose multiple sclerosis	hybrid	syntactic analysis	cTAKES	statistical - regression	Partners Healthcare System	clinical note	595	2013	USA
Ye et al. (2017)	diagnose influenza	not reported	pattern matching	in-house developed	statistical - Bayesian	University of Pittsburgh Medical Center, Intermountain Healthcare in Utah	clinical note	88,693	2017	USA
Zhang et al. (2019)	predict the utilization of advanced diagnostic imaging	not reported	semantic analysis	R	statistical - regression	US National Hospital Ambulatory Medical Care Survey data	triage note	139,150	2019	USA
Zheng et al. (2019)	assess imaging diagnosis quality	not reported	semantic analysis	Jieba	statistical - convolutional neural network	Shanghai Tongren Hospital, China	imaging and pathology report	16,354	2019	China
Zhong et al. (2019)	predict suicidal behavior	custom made	syntactic analysis	cTAKES	knowledge-based	Partners Healthcare System	clinical note	1,120	2018	USA

DIVIDED AND CROSSCUTTING BLOGOSPHERE POLITICAL DIALOGUES: REVISITING THE SPIRAL OF SILENCE AND THE BANDWAGON EFFECT

Hyun Jung Yun, Texas State University

ABSTRACT

This study revisited two traditional public opinion theories, the Spiral of Silence and the Bandwagon effect, and proposed modified versions of the reverse Spiral of Silence as an ad-hoc approach and the digressing Bandwagon effect as a post-hoc approach to examine the patterns of political opinion formation and distribution in the blogosphere. The study analyzed textual-based interactive weblog dialogues and found that political weblogs fed crosscutting views that used to be suppressed in real-world politics. There were differing degrees of suppressing Spiral of Silence and digressing Bandwagon effects across the different types of electoral, general political and non-political weblogs through the different venues of liberal, neutral, and conservative political dialogues. Politically liberal bloggers in electoral weblogs were most likely to present crosscutting views among the various types of partisan bloggers, nullifying the Spiral of Silence and Bandwagon effects. This study identified the importance of the theoretical justifications of, and alterations to, the classical approaches in the new communication and information era.

Keywords: The Spiral of Silence; The Bandwagon Effect; Political Weblog; Crosscutting Political Dialogue; Partisan Blogger

Political communication online has become a part of the inevitable political dialogue culture brought by new media technology that is instant, efficient, expressive, and individualistic (Bennett & Pfetsch, 2018; Dahlgren, 2005; Kaye & Johnson, 2002). Online political discourse has constantly changed over the last two decades, and no plausible theory yet explains the new communication dynamics thoroughly (Blumler, 2015; Mann & Stewart, 2000; Zúñiga, Jung, & Valenzuela, 2012). Rather, classical and traditional theories have been falsified in their applications to online communication studies due to tenacious core assumptions that are disrupted by new and unexpected online communication circumstances that are different from traditional offline communication (Chadwick, 2017; Walgrave & Aelst, 2006).

Among multiple platforms of online political communication, weblog dialogues are most likely to resemble the traditional offline political communication in terms of communication styles, patterns, and effects (Davis, 2011). Therefore, weblogs are a good place to start and investigate the applicability of classical theories in the new setting of online political discourse. Bloggers are news creators, distributors, and commentators. In the new technology communication era, political bloggers have become one of the main sources of political information (Davis, 2011; McKenna & Pole, 2007). However, heuristic academic approaches have not kept pace with the rapidly emerging political communication by the increasing number of political weblog participants. Similar to the limited academic understanding and the theoretical and methodological applications in general online communication research, the blogosphere has been randomly and sparsely explored (Bennett & Pfetsch, 2018; Sim & Hew, 2010).

The current study proposed different versions of the two traditional public discourse theories, the Spiral of Silence and the Bandwagon effect, in order to weigh their adaptability to the new political communication era. This study analyzed the patterns and dynamics of different partisan political weblogs and found evidence that the classical theories were not fully capable of providing reliable tools to explain the new types of political dialogue in the blogosphere, and thus emphasized the malleable modifications of the traditional theories to be better suited to the newly emerging virtual public sphere.

THEORETICAL UNDERPINNING

Weblogs as Crosscutting Public Squares

Although there has been an increasing volume of descriptive studies about political communication via the Internet or social media, in general, more specific, analytical, and theory-oriented research on the different types of social media are not yet fully examined and the findings are mixed (Blumler, 2015; Sim & Hew, 2010; Wallsten, 2008). For instance, the channel of political blogs has become one of the prominent sources of political information and news and has provided a unique atmosphere and various features for political dialogues; however, it has been understood as just one kind of social media since the earliest studies on the subject (Laschever, 2017; Leccese, 2009; Wallsten, 2008). Therefore, this study specifically looks into the dynamics of political weblogs in their practical functions and theoretical meanings in current American political communication.

As a part of social media and a form of information channels, weblogs also have the online discourse characteristics of anonymity, autonomy, and tolerance, discussed by previous studies (Kushin & Kitchener, 2009; Meraz, 2009; Yun, 2014). Under the circumstances, bloggers' communication traits are often defined as personalization, honest opinion expressions, audience-oriented dialogue processes, and fragmented discourse (Wall, 2005). *Anonymity*. Bloggers are free from many constraints that are often imposed on in-person discussions, partially due to the anonymous weblog environment (Swigger, 2012). As a result, they tend to be unnecessarily honest and raw (Lasica, 2001), and are more likely to infuse views outside of traditional norms and ideas (Davis, 2011). Some argue that bloggers are more likely to be the anonymous public rather than named journalists and are more likely to be information transmitters rather than information sources (Davis, 2011). *Autonomy*. Blogosphere is where bloggers document and express personal beliefs, values, and opinions in public to share with other bloggers and readers (Davis, 2011; Leccese, 2009). In other words, online communication via weblogs is individualistic self-expression (Herring, Scheidt, Bonus, & Wright, 2004). Therefore, weblogs are often very individualistic and highly opinionated public spheres (Lasica, 2001). *Tolerance*. Under a strong individualistic, anonymous, and fragmented weblog culture, bloggers are highly tolerant of the different personal views expressed by other bloggers (Davis, 2011; Lasica, 2001). Bloggers generally pre-assume a higher-level of tolerance in their networking sites and are more willing to embrace opposing or deviating views (Lawrence, Sides, & Farrell, 2010).

Weblogs as general political venues. Unlike other social media, political weblogs cover a wide range of ideas and have various levels of participation. The scope drawn by political weblogs can cover a variety of topics ranging from general political issues and agendas to very specific policy

programs and politicians, in various degrees of depth and intensity (Chuenchom, 2019; Tremayne, 2007). Bloggers, often referred to as "new journalists" or "black market journalists" (Davis, 2011), provide alternative political information, having much more freedom and fewer obligations than traditional journalists (Skube, 2007). Therefore, political weblogs, where the external status quo pressures are lifted and are less bounded by societal norms and expectations, tend to be more independently, anonymously, and tolerantly operated than traditional media channels or non-political social weblogs where sociocultural norms and reciprocal expectations are still heavily laden. As a result, bloggers in such circumstances are even more expressive about their own values and beliefs (Papacharissi, 2004, 2010).

Political weblogs are more likely to drive partisan dialogue, and the full spectrum of polarized political views from the very far-left to the very far-right are presented freely and independently in the blogosphere. Partisan political bloggers engage more proactively in all levels of political participation, from political discussion, campaign activities, and voting to political demonstrations and violence (Adamic & Glance, 2005; Lawrence et al., 2010). Moreover, different political groups affiliated with a specific party and ideology tend to show different patterns in their political participation via weblogs. For example, political dialogue among liberals tends to be more diverse and open, engaging broader demographics and covering broader political issues, while conservative dialogue typically focuses on specific political issues discussed by more homogenous groups of people (Adamic & Glance, 2005; Best & Krueger, 2005; Laschever, 2017).

Weblogs as electoral campaign venues. The anonymous, independent, and tolerant weblog environments tend to be more encouraging for election-oriented weblogs where normative arguments are no longer valid, and preferences are independently accessed. Therefore, political participants via electoral weblogs are even more expressive of their own beliefs and ideas (Carlson, 2007; Huddy, Mason, & Aarøe, 2015).

The public dialogue space in election-specific weblogs can turn into even more partisan-driven communication venues since political discussion about electoral issues are mostly initiated from partisan standpoints (Huddy et al., 2015). Especially, during election years and campaign seasons, political weblogs operated by politically more motivated partisan groups and individuals are more likely to get heightened attention, and politically intentional individuals, like political candidates and their campaign staff, tend to dominate the virtual political weblog dialogues (Vatrapu, Robertson, Dissanayake, & Jeedigunta, 2008).

Moreover, as weblog discourse is intensified by partisanship, the anonymous, autonomous, and tolerant online sphere starts to expand in different directions and degrees. For instance, in liberal online spaces where broader issues are presented and more diverse individuals from broader ranges of demographics are interacting, the political bloggers participate more freely and vocally, (Laschever, 2017; Lawrence et al., 2010) cultivating more active crosscutting views than in other partisan weblogs (Papacharissi, 2004; Tremayne, 2007).

Weblogs as social venues. In contrasts to political weblogs, social weblogs are more implicitly generated and more steadily evolved (Tremayne, 2007). Social bloggers who have been culturally bounded together for longer periods of time and with deeper relationships can be

relatively less anonymous, less autonomous, and less tolerant. On socially, culturally, and personally evolved issues, these bloggers are more likely to rely on each other, share more personal information, and mutually respect their shared values (Baker & Moore, 2008). Therefore, it can be implied that social bloggers are relatively more supportive of and agreeable to the presented issues and themes within their social blogospheres than political bloggers in political blogospheres.

Revisiting the Spiral of Silence in Blogosphere Dialogue

Traditional theoretical parameters of the Spiral of Silence. Since the theory of the Spiral of Silence was introduced by Noelle-Neumann (1974), a great number of studies have adopted the theory to explain the patterns of political expression in public political discourse. The theory argues that people are vulnerable to others' predominant views since humans' natural social sense prevents people from disagreeing with others and thus avoid being alone. This theoretical approach generously embraces the parameter of interdisciplinary understanding of public opinion at multiple levels, covering from the micro-level of socio-psychological mechanisms of "fear of isolation" and "social pressure," and intermediate-levels of "socio-political sense of majority public opinion," to the macro-level of "public expression." The theory provides a multi-angular lens that reflects humans' natural social psychology in daily political communication (Noelle-Neumann, 1993).

The classical theory, however, has faced critical difficulties in explaining political expression in evolving political dialogue in online environment because the fundamental assumptions and solid core argument of the theory are not easily adaptable to and utilized in the new technology communication era.

New communication parameters falsifying the Spiral of Silence. Online political communication has added a new circumstance that makes it hard for the mainstream communication theories and perspectives to explain the changes since those approaches were initially introduced to understand the old-fashioned, in-person and one-way communication. The most challenging elements of cyber political communication are the three components of anonymity, autonomy, and tolerance (Swigger, 2012; Yun, 2014). The three elements substantially weaken the core assumptions of the traditionally predominant theories like the Spiral of Silence that are grounded on socially and physically bounded communication.

The Spiral of Silence theory has been highly validated by the three levels of communication components: the socio-psychological elements of social pressure and fear of isolation; the intermediate social sense of dominant views; and the political consequence as public expression. These are the core assessments of the theoretical approach within socio-physically bounded interactive communications (Noelle-Neumann, 1974, 1993). However, in socially unrestrained and physically unbounded online zones, these fascinating theoretical components have lost their explanatory power and failed to provide valid instructions. The core theoretical components are more seriously falsified in the online atmosphere of weblogs, especially in political weblogs, where socio-psychological senses are significantly diminished due to the partisan-driven thin virtual walls that abate any normative or dominant argument in the loosely divided public sphere.

First, social pressure or fear of isolation is no longer a consistent or persistent condition for political expression because of the anonymous online atmosphere. Although the anonymity of online communication via weblogs is not fully guaranteed, it still creates the atmosphere that lowers the pressure of social expectation of self. As the level of the anonymity increases, social pressure decreases, and thus the sense of being isolated decreases (Yun, 2014).

In weblogs, especially political weblogs, bloggers can obtain anonymity if they choose to. Political participants in online forums in the blogosphere have a higher degree of anonymity than in traditional offline political meetings. Moreover, these political bloggers tend to be more anonymous compared to non-political social bloggers who are often more steadily and exclusively networked among interacting bloggers based on more extended and deeper relationships (Hsu & Lin, 2008). Politically liberal online spaces, where more diverse individuals with broader backgrounds and issues engage, tend to have much more anonymity, therefore have much less social pressure and consciousness than the conservative blogosphere where often more consistent and familiar partisan groups interact (Hargittai, Gallo, & Kane, 2008; Laschever, 2017).

Second, in online political dialogue, the sense of dominant views, the "quasi-statistical sense" according to Noelle-Neumann (1974), becomes ambiguous in that the autonomous component of online political communication feeds an individualistic political environment and triggers independent political individuals to speak, and thus confuses the separation between the majority views and the minority voices (Kushin & Kitchener, 2009; Yun, 2014). Therefore, a predominant opinion is less likely to be formed in an autonomous online environment, and political participants in a more autonomous environment are less likely to be exposed to, and constrained by, one dominant group or one majority view (Holt, 2004; Witschge, 2004). In addition, the patterns of political discourse depend on multiple mediating factors, such as group sizes, participants' demographics, types of issues, and socio-political cultures. However, in online formats, all these mediating factors are unclear and are often not even relevant (Davis, 1999). As a result, the pseudo sense of whether the presented political views online are dominant public opinions is unclear due to a lack of cues for understanding the mediating factors. For instance, a seemingly dominant view in a small online group discussion could in reality be a minority view in disguise rather than a general public view.

The autonomous and independent online atmosphere is a relatively more common environment in political weblogs where there are multiple venues and types of political discussions, and it provides a more diverse but isolated place for independent individuals to express their own political views, and thus cultivates crosscutting views. These tendencies can be more ubiquitous in politically neutral weblogs where there is often no particularly predominant view presented from moderate individuals who do not have, or relate to, any specific political affiliation or issue (Koop & Jansen, 2009).

Third, political expression in online turns to more crosscutting dialogues due to a tolerant virtual atmosphere. Online environment lowers political and socioeconomic huddles for political participants who are used being under-presented and isolated (Yun, 2014). Therefore, political mavericks and social minorities are more likely to participate and raise their voices in online political discussions under less pressure from predominating views and ideas (Brenner, 2012;

Witschge, 2004). As a result, in online political dialogue, people are often sharing less popular opinions and tolerating the different views of others (Davis, 1999; Kim, 2011; Wojcieszak & Mutz, 2009). As the level of tolerance increases, political voices against pre-perceived or predominant views are more likely to be expressed (Kushin & Kitchener, 2009).

In American politics, political tolerance tends to be higher from political liberals who tend to have broader political issues and views coming from diverse demographic groups rather than from conservatives who have more specific political issues and directional approaches and consist of politically and socioeconomically more cohesive groups (Carney, Jost, Gosling, & Potter, 2008). The pattern is directly reflected in weblog political discourse. Like real-world liberals, the politically liberal bloggers are often socio-economically and politically more diverse and are more tolerant to exchanging different and opposing views than their politically conservative counterparts who tend to be more cohesive in their interactions and discourses among their like-minded bloggers (Hargittai et al., 2008; Laschever, 2017; Nam, Jost, & Van Bavel, 2013; Rainie & Smith, 2012).

Alternative justification of the Spiral of Silence. The characteristics of online communication – anonymity, autonomy, and tolerance – significantly weaken the theoretical justification for the Spiral of Silence, particularly in its applications to weblogs. However, the accumulated theoretical supremacy of the theory cannot be ignored even in the new technology communication era. Therefore, rather than being stuck by the powerful but invalidated theoretical components of the Spiral of Silence in online venues, the current study proposes the reverse effect of the Spiral of Silence as an ad-hoc tool to examine blogosphere dialogues. In other words, as a pre-assumed theoretical perspective, the modified version of the Spiral of Silence can be used to understand the reverse phenomena of the Spiral of Silence in the initial process of opinion formation in the virtual sphere. The reverse Spiral of Silence can explain how different types of political weblog dialogues reduce the degree of social pressure and the sense of oppression by dominant views and encourage political expression as the levels of virtual anonymity, autonomy, and tolerance increase. As an ad-hoc approach in exploring online political communication, the modified model of the reverse Spiral of Silence can guide us in understanding the early stage of bloggers' socio-psychological mindsets when bloggers enter the blogosphere political discussion, and their dialogue begins to mature.

Revisiting the Bandwagon Effect in Blogosphere Dialogue

Traditional theoretical parameters of the Bandwagon effect. Since the very first practical success of a Bandwagon political campaign by Dan Rice in 1848 (Carlyon, 2001), the metaphor of bandwagon has been used extensively in professional politics and theoretical academia. While the Spiral of Silence is initiated from a socio-cognitive understanding in the beginning of opinion formation and process, the Bandwagon effect is more of a behavioral approach that occurs in the later stage of opinion formation. Joining the majority and getting on a more popular and bigger bandwagon are the main properties of the theory (Henshel & Johnston, 1987).

However, the Bandwagon effect is a theoretically ambiguous concept and a methodologically difficult measure (Klapper, 1964). Previous studies used the Bandwagon effect with different operational definitions and measures depending on the purposes of the studies. The amorphous

theoretical parameters of the Bandwagon effect can help the theory be more applicable to emerging communication research in the new technology era but can also weaken the theoretical power of predictability and generalizability.

New communication parameters falsifying the Bandwagon effect. As discussed in the Spiral of Silence's application in weblog research, the cyberspace components of anonymity, autonomy, and tolerance (Swigger, 2012; Yun, 2014) can also weaken the core theoretical mechanisms of the Bandwagon effect which are behavioral consequences of joining the majority.

First, the sense of conformity by joining the majority is no longer sufficient motivation in online due to the absence of a physical gathering space in an anonymous cyberspace (Wood & Smith, 2004). In the weblog sphere where anonymous individuals come across at random times (Davis, 2011; Wall, 2005), bloggers have less desire or need to be a part of the visible majority group. However, in a more specific, issue-focused weblog with more exclusive bloggers, a visible bandwagon tends to form, and the bloggers have stronger desires to share the established view and be a part of the dominant group. For instance, social weblogs, where bloggers have greater social bonds and networks based on self-interested themes, are more likely to show the pattern of the Bandwagon effect. Moreover, specific issue-oriented weblogs are more likely to show the pattern than broader and general weblogs (Baker & Moore, 2008; Tremayne, 2007).

Second, often there is no dominant view or apparent bandwagon to join in blogosphere since the autonomous online environment prevents a majority opinion from forming. Independent online visitors are less likely to stay as a group and are more likely to move freely around, depending on their interests and time constraints (Lasica, 2001; Wall, 2005). Therefore, unlike the Bandwagon effect in the real world, predominant or leading bandwagons are unlikely to form, and are not very visible, desirable, or appreciated in blogosphere.

In weblogs, individual bloggers have their own vehicles and venues to deliver their views to others. Political bloggers with different political ideologies independently interact in equally accessible cyberspace. However, the degree of autonomy varies across the different ideological spectrums. From the accumulated research on the patterns of different partisan groups, we can imply that political conservatives are more agreeable and supportive of people or views within their political boundaries since their political target topics tend to be specific and Republican party supporters are more loyal to their party and candidates than politically liberal or moderate people who have broader interests in random issues and are loosely affiliated to their party (Hargittai et al., 2007; Nam et al., 2013). Therefore, there tends to be more homogeneous consensus and agreement among the conservative Republican bloggers than the liberal Democrat counterparts (Laschever, 2017). Moreover, bloggers who interact with each other regarding non-political social topics, such as personal interests and hobbies, and virtually know each other are more agreeable to, and supportive of, suggested ideas and preferences within their social boundaries (Baker & Moore, 2008). Those socially networked bloggers, who are more tightly bounded and have less autonomy, may still form a more visible like-minded group and operate as a bandwagon in the blogosphere.

Third, tolerant, unlimited online spaces create multiple bandwagons, rather than one single dominant one. In a democratic and open society like the U.S., people respect the diversity and

differences of others across online and offline boundaries. The level of tolerance, however, is higher online where there are unlimited spaces that individuals can share without much conflict (Foust, 2017). As a result, online citizens are more willing to accept different views and groups, allowing for multiple bandwagons to coexist in unlimited spaces.

Therefore, there are multiple bandwagons co-occurring with few concerns of clashing interests in the highly tolerant blogosphere. Around each bandwagon, navigating within or across different weblogs, bloggers easily hop in and out of different bandwagons. The trends can be more true for political bloggers who gather for timely political issues on political weblogs where differences are openly tolerated, and less true for non-political social bloggers who mingle for more consistent social or personal interests where differences are often silently tolerated (Baker & Moore, 2008; Marlow, 2004). For the reasons of social forbearance and political freedom, people are more respectable and accepting of societal diversities but more expressive about their political views. In other words, political bloggers tend to be more expressive about different views and move around more easily across different bandwagons than social bloggers who are unlikely to attack or disagree with each other about social or cultural topics within their stable bandwagon group. The openly expressive atmosphere becomes even more apparent when it is about elections, or regarding political choices and preference rather than about general political norms, values, or traditions (Carlson, 2007; Huddy et al., 2015). Bloggers become more overt by moving from social blogs to political blogs, and by moving from general political blogs to more specific electoral blogs. It suggests that we are more likely to observe multiple loose bandwagons by crosscutting political bloggers and more tightly bonded bandwagons by covertly agreeing socially networked bloggers.

Alternative justification of the Bandwagon effect. Unlike the reverse effect of the Spiral of Silence, but with the same rationale of online anonymity, autonomy, and tolerance, weblogs provide good conditions to feed dynamic Bandwagon effects depending on the types of blogs in blogosphere. Each type of weblog has a unique circumstance since the blogosphere is an unbounded place for timely, like-minded people to selectively consume online information based on their own interests (Ancu & Cozma, 2009). Therefore, bloggers are more likely to search for, and observe like-minded views, and they are still more likely to agree with people within a click of a weblog's boundary across different topics, issues, and bloggers. These temporary and loose ties of cybernetic groups among bloggers make the gradual agreements bigger and stronger as multiple weblog dialogues evolve.

This study, therefore, values the Bandwagon effects as a more post-hoc tool to explain the later stage of weblog discourses after bloggers form visible opinions. The study also investigates how the formed bandwagons increase in size and become more cohesive as bloggers observe the degrees of agreement within the engaged weblogs. In blogosphere, the Bandwagon effect provides a more legitimate tool to examine the final stage of opinion formation rather than being a pre-assumed ad-hoc approach due to its behavior-oriented theoretical perspective. Therefore, the Bandwagon effect is a more compromising theoretical perspective as a post-hoc approach to explore the later stages of the public dialogue process in the new evolving political discourse environment in cyberspace.

RESEARCH HYPOTHESES

The anonymous, autonomous, and tolerant atmosphere of online communication can lift socio-psychological pressure and help people feel less vulnerable to the majority opinion of others, and thus significantly handicap the theory of the Spiral of Silence (Yun, 2014). As an ad-hoc approach in blogosphere, the opposing effects of the Spiral of Silence can be posited. Bloggers in a less value-laden and more preference-oriented blogosphere, such as political blogs, especially electoral weblogs, tend to be freer from the socio-psychological pressure of the conscious agreement. For reasons involving ideological and partisan traits, liberal bloggers with broader demographic backgrounds and issues can be much less constrained by the dominant views than conservative bloggers who are relatively more consistent and cohesive in their issues and perspectives.

> H1.1: Election-related weblog dialogues are more likely to suppress the effect of the Spiral of Silence than general political weblog dialogues.
>
> H1.2: General political weblog dialogues are more likely to suppress the effect of the Spiral of Silence than non-political social weblog dialogues.
>
> H1.3: Liberal political weblog dialogues are more likely to suppress the effect of the Spiral of Silence than conservative political weblog dialogues.

Previous studies have observed that there are less or more Bandwagon effects among bloggers depending on the sizes and types of online political groups (Tan, Goswami, Chan, & Zhong, 2005). Due to the variations in political and social traits among bloggers in the anonymous, autonomous, and tolerant weblog atmosphere, the Bandwagon effect, as a behavioral post-hoc approach, is restrained or present in varying degrees across different weblog dialogues. In the absence of a physical presence, the more behavior-oriented blogospheres, such as election-driven political weblogs, are more likely to repress the Bandwagon effect. In addition, due to the ideological patterns and traits exhibited in Republicans, who tend to be relatively more homogenous compared to loosely tied and diverse Democrats, the Bandwagon effect in the conservative weblog discourse can still appear and be stronger than in the liberal or neutral political blogosphere discourse.

> H2.1: Election-related weblog dialogues are more likely to digress from the Bandwagon effect than general political weblog dialogues.
>
> H2.2: General political weblog dialogues are more likely to digress from the Bandwagon effect than non-political social weblog dialogues.
>
> H2.3: Liberal political weblog dialogues are more likely to digress from the Bandwagon effect than conservative political weblog dialogues.

METHODS

Sample Collection and Coding Process

This study explored textual-based interactive weblog dialogues including the subsequent comments resulting from those blog articles and posts across general political, election-related, and non-political social blogs in order to observe the patterns and processes of opinion formation and distribution in the blogosphere. The samples of weblog articles and comments were randomly collected over both election and non-election years, from February 15, 2015 to April 28, 2016, from the top 100 political and social blogs sorted by Feedspot, for a more comprehensive comparison among election-related posts, non-electoral general political posts, and non-political social posts, in order to have better control over mediating factors such as different political atmospheres and election effects.

In analyzing general political and election-specific weblogs, the researcher intended to observe political weblogs and the subsequent comments along the ideological spectrum from liberal, independent, and conservative blogs to find the patterns and trends of online public opinion formation and distribution among those ideologically varying political bloggers. For the purpose of the research and analysis, this study limited the sample weblog articles to those that had at least five subsequent comments reacting to the weblog articles, and selected only the first five comments across general political, electoral-specific, and non-political weblogs as well as through liberal, moderate, and conservative venues, in order to observe more consistent patterns and processes of opinion formation and distribution in the blogosphere and to minimize non-essential variations that can occur extensively across the different types of blog articles and comments.

Twelve college students went through a two-hour training session on sample selections, archiving methods, and coding processes, including a Q & A session. After, the training session, they randomly selected liberal, neutral, and conservative political and non-political weblog articles based on political content, bias, and tone across different types and partisanship of weblogs from the top 100 Feedspot political and social blogs, and then saved the first five comments from each of the selected sample articles during the collection period. The unit of analysis is a weblog article with the first five comments posted after the initial article. At the end of the sample collection period, the student coders coded ten example weblog articles and the resulting comments, and they achieved the Cronbach's Alpha intercoder reliability of .82.

Out of the 801 randomly selected weblog articles and the resulting comments, 124 were liberal electoral articles, 122 were neutral electoral articles, 131 were conservative electoral articles, 103 were liberal general political articles, 113 were neutral general political articles, 103 were conservative general political articles, and 105 were non-political articles.

Measures

Type of weblog. The initial analysis layer of weblogs was categorized into the following types: general political, election-specific, and non-political social weblogs based on the content of the blog articles. The second layer of the weblogs under general political and electoral weblogs was categorized as follows: liberal, neutral, and conservative weblogs based on partisan tone and bias.

Level of agreement. The level of agreement between each of the first five commenters to the initial weblog articles and each of the previous comments were measured on a scale of seven;

from 1 – "strongly disagree" to 7 – "strongly agree." The averages of the total fair-wise agreements across the different types of weblogs were accessed. For the gradual changes in the level of agreement, both the average of agreement of each the 1st, 2nd, 3rd, 4th, and 5th commenters, and the accumulated agreements from the 1st commenters up to the 2nd, 3rd, 4th, and 5th commenters across the different types of weblogs were computed.

Disagreement and crosscutting. For additional measures of disagreement against dominant views, the total numbers of the first five commenters who made attacking or conflicting statements to previous bloggers and commenters were counted in a dichotomous manner where 0 was a "no" and 1 was a "yes." The average and percentage of the attacking statements across the different types of weblogs were compared.

Commenting structure. Several other measures were assessed to understand the subsequent commenting patterns of bloggers in the different types of weblogs. At the sample collection stage, *the total number of the responding comments* were counted to see how the average number of comments varied across the different types of weblogs. *The average commenting timeline* for the first five comments was also retrieved to see how long it took for the first five active and intentional commenters to respond, depending on the types of weblogs they were responding on. In addition, *the average length of the comments in words* for those first five commenters were counted by the types of weblogs. The varying degrees in *tone for subsequent comments* to the different types of weblog articles/posts were measured on a scale of seven, where 1 was "very negative" and 7 was "very positive." Lastly, any *personal cues* from bloggers that were observed and nuanced in the statements or in any other forms, such as pictures and nick names, were coded in a dichotomous manner where 0 was "no cue" and 1 was "specific personal cue," for the list of socio-economic and political demographics; gender, political party, religion, education, occupation, age, ethnicity, and geographic location. The average number of personal cues for the first five commenters was used to see how those bloggers in the different types of weblogs were revealing and sharing their personal demographic information in their blogosphere dialogues.

RESULTS

Genetic Patterns of Weblog Dialogue

Utilizing the measures described above, this study found different patterns of dialogue across the different types of weblogs. Regarding the *average number of total subsequent comments*, election-specific (M=455.21, SE=78.32) and general political (M=419.76, SE=86.06) weblog articles seemed to generate greater amounts of dialogues than non-political blog posts (M=37.45, SE=147.85) among the subsequent commenters. More specifically, the moderate general political blogs (M=964.38, SE=215.83) were most likely to feed bloggers' opinionated discussions and reactions to others' views ($F[6, 767]=6.045$, $p \leq .001$) (See Table 1).

However, politically conservative bloggers posted their comments on engaged blogs faster than other partisan bloggers. The *average timeline* to get the first five subsequent comments were about 13.47 hours (SE=10.35) in the conservative general political blogs and 26.86 hours (SE=9.02) in the conservative electoral blogs. It took a much longer time for commenters to post on both electoral (M=46.65, SE=9.22) and general political blogs (M=49.46, SE=10.46) in a

neutral atmosphere, and the longest time on the general political blogs in a liberal atmosphere (M=73.86, SE=10.35) (F[6, 686]=3.749, p ≤ .001) (See Table 1).

In terms of the *length of comments*, non-political bloggers who were often more likely to be interested in socioeconomic topics rather than political issues, tended to be lengthier in their replies to the views of previous commenters compared to political commenters. The average word count for the first five commenters on non-political weblogs was 189.34 (SE=57.09) while the average lengths were 63.74 (SE=2.95) words in general political blogs and 49.13 (SE=1.93) words in more specific electoral blogs (F[6, 772]=6.273, p ≤ .001). In reviewing the *tone* of the weblog dialogue, non-political social blog discourses (M=4.70, SE=.17) were more supportive and positive than political weblog discourses (M=2.93, SE=.05 in electoral weblogs; M=2.66, SE=.07 in general political weblogs) (F[6, 756]=41.357, p ≤ .001) (See Table 1).

Various *demographic cues* were either implicitly hinted at or explicitly presented in weblog dialogues. The demographics were observed in all types of weblogs, but there were different degrees of personal information shared across the different types of weblogs. The general demographics, such as bloggers' gender, political party, religious stand, level of education, and occupation were present more frequently but to varying degrees across different weblogs. However, the demographics that are more vulnerable at the personal level such as age, ethnic background, and physical location were less likely and less frequently presented across the different types of weblogs. Gender was the most frequently presented demographic cue in weblog dialogue; it was the most commonly shared information on election-focused weblogs, especially by politically liberal bloggers (M=2.56, SE=.19) (F[6, 788]=7.086, p ≤ .001). Generally, conservative bloggers in general political weblogs were most likely to reveal their political party affiliations (M=2.25, SE=.20) (F[6, 788]=22.718, p ≤ .001) and religious affiliations (M=0.47, SE=.11) (F[6, 788]=6.562, p ≤ .001) than other partisan and non-political bloggers. In election-specific but ideologically neutral weblogs, bloggers were more likely to refer to their level of education (M=0.26, SE=.07) (F[6, 788]=3.760, p ≤ .001). Unlike political bloggers who were interested in political exchanges of ideas and values, non-political social bloggers who intended to gain more socioeconomic benefits, such as networking and marketing, were more likely to share their occupations with others within engaged weblogs (M=0.24, SE=.06) (F[6, 788]=7.020, p ≤ .001). However, bloggers were still uncomfortable disclosing their specific ages, ethnic backgrounds, and personal geographic locations, and these demographics were less likely to be shared in the blogosphere and were not distinctively different across the different types of weblogs (see Table 2).

Theoretical Patterns of Weblog Dialogue

As the effects of the Spiral of Silence and the Bandwagon unfolded, the current study found the digressing effects of the classical theories in political, and especially in electoral weblogs due to the higher degrees of tolerant, autonomous and anonymous dialogue environments. The tendency was more apparent among the liberal bloggers who often engaged in broader political issues than the conservative bloggers who often engaged in a more specific scope and set of issues.

The ad-hoc approach of the reverse Spiral of Silence seemed to work in the initial stage of online opinion formation in the blogosphere. The politically liberal, especially electoral atmosphere

sprouted different views of bloggers and weakened the effect of the Spiral of Silence. ANOVA and Scheffe Post-Hoc tests confirmed that bloggers and commenters in election-related weblogs (M=3.98, SE=.05) were less likely to be vulnerable to others' political views and thus more likely to be expressive about their disagreement to others and more openly share crosscutting viewpoints than bloggers and commenters in general political (M=4.95, SE=.07) and non-political (M=5.19, SE=.13) social blogospheres. In other words, contrasting to the argument of the Spiral of Silence, electoral weblog posts and comments were significantly more incongruent to dominant views, and the reverse effect of the theory was stronger in the electoral blogosphere discussions compared to non-electoral political ($p \leq .001$) or non-political social ($p \leq .001$) weblog dialogues ($F[2, 719]=78.056$, $p \leq .001$) (see Table 3).

When analyzing different partisan and ideological weblog dialogues, ANOVA and Scheffe Post-Hoc tests showed that the pattern was consistent. Liberal bloggers and commenters in electoral weblogs (M=3.83, SE=.09) were the least agreeable amongst themselves ($p \leq .001$), and conservative bloggers and commenters in general political blogs (M=5.39, SE=.12) were the most agreeable amongst themselves ($p \leq .001$) compared to bloggers and commenters in other types and partisanship of the weblogs (M=4.54, SE=.05) ($F[6, 715]=33.840$, $p \leq .001$) (see Table 3). The finding implies that political participation in political discussions via weblogs were less likely to be vulnerable to the political atmosphere and the views of the majority, and thus tended to be highly expressive, unlike what the Spiral of Silence theory argued in face-to-face real-world political discussions. As an ad-hoc approach, the suppressed Spiral of Silence effect seemed to be more apparent for political bloggers and commenters in the liberal electoral blogosphere.

Furthermore, the patterns of crosscutting opinion expressions across the different types of weblogs were reconfirmed when comparing the more explicit opposite statements/posts to dominant views related to the original weblog articles. The disagreeing minority views, negative comments, or attack statements against the majority voices were more frequently observed in electoral weblogs (M=1.88, SE=.08) than general political weblogs (M=1.48, SE=.09), and the contradicting crosscutting expressions were more prevalent in political weblogs than in the dialogue of non-political social weblogs (M=1.02, SE=.14) ($F[2, 785]=13.477$, $p \leq .001$). Among different ideological blogospheres, those opposite and negative voices against the dominant views were most likely to be observed in ideologically liberal electoral weblogs (M=2.13, SE=.14), while those negative disagreement and attack statements were least present in conservative political weblogs (M=1.04, SE=.13) ($F[6, 781]=7.563$, $p \leq .001$) (see Table 3).

Among the last commenters (5^{th}) who perceived the views of the majority by observing previous commenters (up to 4^{th} commenters) on the original articles, about 40% of the 5^{th} commenters in election-specific blog discussions made negative attacks against the majority views of the original weblog article bloggers or previous commenters, while 28% of the commenters in non-electoral general political blogs, and 23% of the non-political social blog commenters did ($\chi^2=17.046$, $p \leq .001$). Politically liberal commenters in election-specific weblogs (48%), and neutral commenters on both electoral (38%) and general political (38%) weblogs, were most likely to oppose and attack the view of the majority openly and directly. The conservative commenters in general political blogs were least likely to express disagreeing views to the predominant opinions (16%) ($\chi^2=34.297$, $p \leq .001$) (see Table 3). The finding implies that

bloggers tend to disagree with each other, express views against others, and attack different views where either a liberal atmosphere or no dominant majority view by neutral bloggers is present. In other words, unlike in the Spiral of Silence that posits little opinion deviance in face-to-face political expressions, a considerable number of online political participants expressed incongruent political views against dominant views on weblogs in the early stage of opinion formation. The altering theoretical tendency was stronger for election-focused liberal weblogs than for general conservative political weblogs and was more apparent for political weblogs than non-political social weblogs.

When looking into the gradual process of joining online discourse and forming opinions as the weblog dialogues progressed, the study found gradually incrementing Bandwagon effects across the different types of weblogs. As bloggers perceived more agreement among comments and replies, they were more likely to agree with the overall view of the weblog dialogue. However, the degree of the Bandwagon effect, as a post-hoc approach, depended on the types of weblogs and their bloggers.

As shown in Table 4, the level of the correlation for each level of accumulated agreement by additional commenters had gradually increased. Following commenters perceived higher levels of accumulated agreements, and thus gradually agreed more with the overall views of the previous posts and comments, and finally were more confident in joining the dominant views ($p \leq .001$) (see Table 4). According to the ANOVA and Scheffe Post-Hoc tests, the Bandwagon effect varied and was more evident in non-political social blog discourse rather than in political blog discourse and was more evident in general political blog discourse than in election-related blog discourse ($p \leq .001$). The gradual agreements among the first five commenters on non-political blog posts had the highest standardized correlation value (r) of .81 and mean (M) of 5.19 (SE=.15), the level of the accumulated agreements among the general political blog commenters was relatively moderate (M=4.95, SE=.08, r=.56), and the value for the commenters to election-specific blog posts was the lowest (M=3.98, SE=.06, r=.50). Among political bloggers across different partisan lines, political commenters on liberal electoral blogs (M=3.83, SE=.12 r=.34) tended to have the lowest level of gradual agreement amongst themselves, and commenters on conservative political blogs were likely to have the highest level of gradual agreement amongst themselves (M=5.39, SE=.15 r=.57) ($F[6, 715]=33.839$, $p \leq .001$). In other words, the Bandwagon effect was more likely to digress in liberal dialogue within electoral weblogs than in other partisan discourses in other types of blogs (see Table 4). In conclusion, the reverse effects of the Spiral of Silence and the digressing Bandwagon effects occurred in the blogosphere, and the deviating theoretical trends by cyberspace crosscutting views were more likely to be present in liberal electoral weblogs than conservative general political or non-political social weblogs.

DISCUSSION

Political discourse via weblogs has been burgeoning and highly interactive with real-world politics. However, academic research on the effects and consequences of blogosphere politics is still far behind and has not kept pace with the practical side of real-world utilization. Therefore, early studies including the current study in the new era of political communication are struggling

to explore online political discourse without any established theoretical applications and reliable approaches.

The current study observed the varying effects of the different types of weblog dialogues and suggested modified theoretical approaches revisiting the classical theories and concepts that mainstream studies have repeatedly tested in the field of communication. As argued above, the study found evidence that the traditional theoretical approaches of the Spiral of Silence and Bandwagon effects are not able to fully explain the dynamics of political discourse in the anonymous, autonomous, and tolerant blogosphere, and needed to be altered in order to adapt to the changes in political dialogue and the rapid evolution of communication technology.

As predicted in the hypothesized patterns of blogosphere dialogue, applied by various pieces of previous bodies of work on online communication, the more preference-oriented and less norm-laden election-focused political weblogs, especially in a broader and more horizontally operating left-leaning atmosphere, provide more active crosscutting dialogue spaces compared to the more culturally-driven and value-laden social blogs or general political blogs, especially operated among more cohesively and hierarchically running conservative counterparts. In liberal electoral weblogs, the political netizens are less likely to be restrained by the predominant views and are more likely to be expressive of their own views and preferences under the lighter socio-psychological and physical pressures. Despite the different degrees of theoretical deviations depending on the types of weblogs, the evidence indicates digressing Bandwagon effects and suppressing Spiral of Silence effects in blogosphere politics.

By adopting modified classical theories and understanding different levels and patterns of online opinion exchanges, formations, and distributions, weblog feeders and consumers can utilize weblogs in a more efficient way and make weblogs a more positive, constructive, and egalitarian political dialogue sphere; something that real-world politics have not achieved, especially for the politically disadvantaged people in our society.

For effective and beneficial research, the socio-economic, political, and psychological meanings of weblogs need to be multidimensionally considered in combination with new technological capabilities, utilizing various methodological approaches with extensive data sources (Savant, Bhattacharyya, & Kim, 2016; Tan et al., 2005). This does not mean that academia needs completely new theories or approaches for online political communication. Rather, the future studies are encouraged to explore more possibilities for rigorous modifications of the verified classical theories that are still invaluable in their insights and applications, and continuously test them for more adaptive measurements in new communication and politics.

REFERENCE

Adamic, L. A., & Glance, N. (2005, August). *The political blogosphere and the 2004 US election: Divided they blog.* Paper presented at the 3rd International Workshop on Link Discovery, Chicago, Illinois.

Ancu, M., & Cozma, R. (2009). MySpace politics: Uses and gratifications of befriending candidates. *Journal of Broadcasting and Electronic Media, 53*(4), 567–583.

Baker, J. R., & Moore, S. M. (2008). Blogging as a social tool: A psychosocial examination of the effects of blogging. *CyberPsychology & Behavior, 11*(6), 747–749.

Bennett, W. L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of communication, 68*(2), 243–253.

Best, S. J., & Krueger, B. S. (2005). Analyzing the representativeness of Internet political participation. *Political Behavior, 27*(2), 183–216.

Blumler, J. G. (2015). Core theories of political communication: Foundational and freshly minted. *Communication Theory*, *25*(4), 426–438.

Brenner, J. (2012). Pew Internet: Social networking (full detail). *Pew Research Center*. Retrieved from http://pewinternet.org/Commentary/2012/March/Pew-Internet-Social-Networking-full-detail.aspx

Carlson, M. (2007). Blogs and journalistic authority: The role of blogs in US election day 2004 coverage. *Journalism Studies, 8*(2), 264–279.

Carlyon, D. (2001). *Dan Rice, the most famous man you've never heard of*. New York, NY: Public Affairs.

Carney, D. R., Jost, J. T., Gosling, S. D., & Potter, J. (2008). The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind. *Political Psychology, 29*(6), 807–840.

Chadwick, A. (2017). *The hybrid media system: Politics and power*. Oxford, UK: Oxford University Press.

Chuenchom, S. (2019). The Political Weblogs and Political Communication. *Journal of Information Science*, *37*(4), 72–91.

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political communication, 22*(2), 147–162.

Davis, R. (1999). *The web of politics: The Internet's impact on the American political system.* Oxford, UK: Oxford University Press.

Davis, R. (2011). A symbiotic relationship: Bloggers and journalists. In D. A. Graber (6th Ed.), *Media Power in Politics* (pp. 293–302). Washington, DC: CQ Press.

Foust, J. (2017). *Online journalism: principles and practices of news for the Web*. Abingdon, UK: Taylor & Francis.

Hargittai, E., Gallo, J., & Kane, M. (2008). Cross-ideological discussions among conservative and liberal bloggers. *Public Choice, 134*(1), 67–86.

Henshel, R. L., & Johnston, W. (1987). The emergence of bandwagon effects: A theory. *The Sociological Quarterly, 28*(4), 493–511.

Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004, January). *Bridging the gap: A genre analysis of weblogs*. Paper presented at the 37th annual Hawaii International Conference, Big Island, Hawaii.

Holt, R. (2004). *Dialogue on the Internet: Language, civic identity, and computer–mediated communication.* Westport, CT: Praeger Publishers.

Hsu, C. L., & Lin, J. C. C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & management, 45*(1), 65–74.

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review, 109*(1), 1–17.

Kaye, B. K., & Johnson, T. J. (2002). Online and in the know: Uses and gratifications of the web for political information. *Journal of Broadcasting & Electronic Media, 46*(1), 54–71.

Kim, Y. (2011). The contribution of social network sites to exposure to political difference: The relationships among SNSs, online political messaging, and exposure to cross-cutting perspectives. *Computers in Human Behavior, 27*(2), 971–977.

Klapper, J. T. (1964). *Bandwagon: A review of the literature.* In: Office of Social Research. New York, NY: Columbia Broadcasting Systems.

Koop, R., & Jansen, H. J. (2009). Political blogs and blogrolls in Canada: Forums for democratic deliberation? *Social Science Computer Review, 27*(2), 155–173.

Kushin, M. J., & Kitchener, K. (2009). Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday, 14*(11). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2645/2350

Laschever, E. (2017). Are They Not Worthy? How Partisan Political Blogs Legitimize the Tea Party Movement and Occupy Wall Street. *Sociological Forum, 32*(2), 359–380.

Lasica, J. D. (2001). Weblogs: A new source of news. *Online Journalism Review*. Retrieved from http://www.jdlasica.com/2001/05/31/weblogs-a-new-source-of-news/

Lawrence, E., Sides, J., & Farrell, H. (2010). Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. *Perspectives on Politics 8*(1), 141–157.

Leccese, M. (2009). Online information sources of political blogs. *Journalism & Mass Communication Quarterly, 86*(3), 578–593.

Mann, C., & Stewart, F. (2000). *Internet communication and qualitative research: A handbook for researching online*. London, UK: Sage.

Marlow, C. (2004, May). *Audience, structure and authority in the weblog community*. Paper presented at the 54th International Communication Association Conference, New Orleans, LA.

McKenna, L., & Pole, A. (2008). What do bloggers do: an average day on an average political blog. *Public Choice, 134*(1), 97–108.

Meraz, S. (2009). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication, 14*(3), 682–707.

Nam, H. H., Jost, J. T., & Van Bavel, J. J. (2013). Not for All the Tea in China! Political Ideology and the Avoidance of Dissonance-Arousing Situations. *PLOS ONE, 8*(4), e59837.

Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication, 24*(2), 43–51.

Noelle-Neumann, E. (1993). *The spiral of silence: Public opinion-our social skin*. Chicago, IL: The University of Chicago Press.

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New media & society, 6*(2), 259–283.

Papacharissi, Z. (2010). *A networked self: Identity, community, and culture on social network sites*. New York, NY: Routledge.

Rainie, L., & Smith, A. (2012). Politics on social networking sites. *Pew Research Center's Internet & American Life Project*. Retrieved from http://pewinternet.org/~/media//Files/Reports/2012/PIP_PoliticalLifeonSocialNetworkingSites.pdf

Savant, P. D., Bhattacharyya, D., & Kim, T. H. (2016). Hadoop based weblog analysis: a review. *International Journal of Software Engineering and its Applications*, *10*(6), 13–30.

Sim, J. W. S., & Hew, K. F. (2010). The use of weblogs in higher education settings: A review of empirical research. *Educational Research Review, 5*(2), 151–163.

Skube, M. (2007, August 19). Blogs: All the noise that fits. *Los Angeles Times*. Retrieved from http://www.latimes.com/la-op-skube19aug19-story.html

Swigger, N. (2012). The online citizen: Is social media changing citizen's beliefs about democratic values? *Political Behavior, 35*(3), 589–603.

Tan, C. H., Goswami, S., Chan, Y. P., & Zhong, Y. Q. (2005, August). *Conceptual evaluation of weblog as a computer mediated communication application*. Paper presented at the 11th Americans Conference on Information Systems, Omaha, NE.

Tremayne, M. (2007). *Blogging, citizenship, and the future of media*. New York, NY: Routledge.

Vatrapu, R., Robertson, S., Dissanayake, W., & Jeedigunta, A. (2008). Are Political Weblogs Public Spheres or Partisan Spheres? A Virtual Ethnographic Study of Online Participations and Implications for Civic Participation in the Internet Age. *International Reports on Socio-Informatics, 5*(1), 7–26.

Walgrave, S., & Aelst, P. V. (2006). The contingency of the mass media's political agenda setting power: Toward a preliminary theory. *Journal of communication, 56*(1), 88–109.

Wall, M. (2005). Blogs of war: Weblogs as news. *Journalism, 6*(2), 153–172.

Wallsten, K. (2008). Political blogs: Transmission belts, soapboxes, mobilizers, or conversation starters? *Journal of Information Technology & Politics, 4*(3), 19–40.

Witschge, T. (2004). Online deliberation: Possibilities of the Internet for deliberative democracy. In P. M. Shane (Ed.), *Democracy online: The prospects for political renewal through the Internet* (pp. 109–122). New York, NY: Routledge.

Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication, 59*(1), 40–56.

Wood, A. F., & Smith, M. J. (2004). *Online communication: Linking technology, identity and culture*. New York, NY: Routledge.

Yun, H. J. (2014). The spirals of newly transcending political voices: Social media purify the atmosphere of political dialogues in cyberspace. In J. Hendricks and D. Schill (Eds.), *Presidential Campaigning and Social Media* (pp.154–170). New York, NY: Oxford University Press.

Zúñiga, G. H., Jung, N., & Valenzuela, S. (2012). Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication, 17*(3), 319–336.

Table 1. Genetic Patterns of Weblog Dialogue

Types of Weblogs***	# of Total Comments***	Commenting Timeline***	Length of Comments***	Tone of Comments***
Electoral Blogs	M=455.21 (SE=78.32)	M=33.63 (SE=5.37)	M=49.13 (SE=1.93)	M=2.93 (SE=.05)
Liberal	M=310.28 (SE=56.76)	M=27.53 (SE=9.34)	M=52.41 (SE=3.19)	M=2.99 (SE=.08)
Neutral	M=326.71 (SE=64.77)	M=46.65 (SE=9.22)	M=46.16 (SE=2.93)	M=3.03 (SE=.10)
Conservative	M=710.56 (SE=215.83)	M=26.86 (SE=9.02)	M=48.73 (SE=3.80)	M=2.78 (SE=.19)
General Political Blogs	M=419.76 (SE=86.06)	M=45.57 (SE=6.07)	M=63.74 (SE=2.95)	M=2.66 (SE=.07)
Liberal	M=158.63 (SE=30.85)	M=73.86 (SE=10.35)	M=45.95 (SE=2.71)	M=2.63 (SE=.12)
Neutral	M=964.38 (SE=215.83)	M=49.46 (SE=10.46)	M=80.21 (SE=6.55)	M=2.79 (SE=.09)
Conservative	M=61.26 (SE=12.62)	M=13.47 (SE=10.35)	M=62.57 (SE=4.05)	M=2.54 (SE=.13)
Non-Political Blogs	M=37.45 (SE=147.85)	M=36.12 (SE=12.09)	M=189.34 (SE=57.09)	M=4.70 (SE=.17)
	F[6, 767]=6.045, $p \leq .001$	F[6, 686]=3.749, $p \leq .001$	F[6, 772]=6.273, $p \leq .001$	F[6, 756]=41.357, $p \leq .001$

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

+ M=Mean (SE=Standard Error)

+ Units: Commenting Timeline in Hour; Length of Comments in Word; Min.=1 & Max.=7 in Tone of Comments

Table 2. Demographic Cues in Weblog Dialogue

Types of Weblogs***	Gender***	Party***	Religion***	Education***	Occupation***	Age	Ethnicity	Geographic Location
Electoral Blogs	M=2.22 (SE=.11)	M=2.11 (SE=.09)	M=0.10 (SE=.02)	M=0.22 (SE=.04)	M=0.05 (SE=.01)	M=0.02 (SE=.01)	M=0.02 (SE=.01)	M=0.11 (SE=.03)
Liberal	M=2.56 (SE=.19)	M=2.07 (SE=.15)	M=0.10 (SE=.04)	M=0.20 (SE=.05)	M=0.05 (SE=.03)	M=0.02 (SE=.02)	M=0.02 (SE=.01)	M=0.15 (SE=.06)
Neutral	M=2.21 (SE=.19)	M=2.03 (SE=.17)	M=0.14 (SE=.04)	M=0.26 (SE=.07)	M=0.06 (SE=.03)	M=0.01 (SE=.01)	M=0.02 (SE=.01)	M=0.08 (SE=.04)
Conservative	M=1.90 (SE=.17)	M=2.22 (SE=.17)	M=0.07 (SE=.03)	M=0.19 (SE=.06)	M=0.05 (SE=.03)	M=0.03 (SE=.02)	M=0.02 (SE=.01)	M=0.09 (SE=.04)
General Political Blogs	M=1.40 (SE=.10)	M=1.75 (SE=.11)	M=0.24 (SE=.04)	M=0.08 (SE=.02)	M=0.01 (SE=.01)	M=0.01 (SE=.01)	M=0.01 (SE=.00)	M=0.04 (SE=.02)
Liberal	M=1.66 (SE=.18)	M=1.86 (SE=.18)	M=0.14 (SE=.06)	M=0.02 (SE=.01)	M=0.01 (SE=.01)	M=0.01 (SE=.01)	M=0.01 (SE=.01)	M=0.03 (SE=.02)
Neutral	M=1.41 (SE=.16)	M=1.20 (SE=.16)	M=0.13 (SE=.06)	M=0.15 (SE=.05)	M=0.01 (SE=.01)	M=0.00 (SE=.00)	M=0.00 (SE=.00)	M=0.06 (SE=.05)
Conservative	M=1.13 (SE=.16)	M=2.25 (SE=.20)	M=0.47 (SE=.11)	M=0.05 (SE=.02)	M=0.01 (SE=.01)	M=0.03 (SE=.03)	M=0.01 (SE=.01)	M=0.04 (SE=.03)
Non-Political Blogs	M=1.88 (SE=.21)	M=0.05 (SE=.03)	M=0.04 (SE=.03)	M=0.04 (SE=.02)	M=0.24 (SE=.06)	M=0.03 (SE=.02)	M=0.02 (SE=.01)	M=0.07 (SE=.03)
	F[6, 788]=7.086, $p \leq .001$	F[6, 788]=22.718, $p \leq .001$	F[6, 788]=6.562, $p \leq .001$	F[6, 788]=3.760, $p \leq .001$	F[6, 788]=7.020, $p \leq .001$	F[6, 792]=0.616, $p \leq .718$	F[6, 792]=0.583, $p \leq .744$	F[6, 788]=0.997, $p \leq .426$

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

+ M=Mean (SE=Standard Error): Min.=0 & Max.=5

Table 3. Digressing Agreements in Weblog Dialogue

Types of Weblogs* **	**Level of Agreement*** **	**Level of Opposing View*** **	**% of Attack Statement*** **
Electoral Blogs	M=3.98 (SE=.05)	M=1.88 (SE=.08)	40.4%
General Political Blogs	M=4.95 (SE=.07)	M=1.48 (SE=.09)	28.3%
Non-Political Blogs	M=5.19 (SE=.13)	M=1.02 (SE=.14)	22.8%
	F[2, 719]=78.056, p ≤ .001	F[2, 785]=13.477, p ≤ .001	χ^2=17.046, p ≤ .001
Electoral-Liberal	M=3.83 (SE=.09)	M=2.13 (SE=.14)	48.0%
Electoral-Neutral	M=3.92 (SE=.08)	M=1.77 (SE=.15)	38.0%
Electoral-Conservative	M=4.18 (SE=.09)	M=1.74 (SE=.14)	35.4%
General Political-Liberal	M=4.97 (SE=.12)	M=1.57 (SE=.18)	30.1%
General Political-Neutral	M=4.53 (SE=.10)	M=1.80 (SE=.15)	38.1%
General Political-Conservative	M=5.39 (SE=.12)	M=1.04 (SE=.13)	15.7%
Non-Political	M=5.19 (SE=.13)	M=1.02 (SE=.14)	22.8%
	F[6, 715]=33.840, p ≤ .001	F[6, 781]=7.563, p ≤ .001	χ^2=34.297, p ≤ .001

* p ≤ .05; ** p ≤ .01; *** p ≤ .001

+ M=Mean (SE=Standard Error): Min.=1 & Max.=7 in Level of Agreement; Min.=0 & Max.=5 in Level of Opposing View

Table 4. Varying Accumulated Agreements in Weblog Dialogue

Types of Weblogs* **	**Up to the 2nd Commenters*** **	**Up to the 3rd Commenters*** **	**Up to the 4th Commenters*** **	**Up to the 5th Commenters*** **
Electoral Blogs	M=4.02 (SE=.07), r=.29***	M=4.02 (SE=.07), r=.39***	M=3.99 (SE=.07), r=.42***	M=3.98 (SE=.06), r=.50***
Liberal	M=3.84 (SE=.12), r=.12	M=3.89 (SE=.12), r=.43***	M=3.87 (SE=.13), r=.45***	M=3.83 (SE=.12), r=.34***
Neutral	M=4.01 (SE=.13), r=.38***	M=3.96 (SE=.12), r=.48***	M=3.95 (SE=.11), r=.43***	M=3.92 (SE=.09), r=.56***
Conservative	M=4.19 (SE=.13), r=.39***	M=4.20 (SE=.13), r=.27***	M=4.15 (SE=.11), r=.38***	M=4.18 (SE=.12), r=.58***
General Political Blogs	M=4.95 (SE=.10), r=.47***	M=4.99 (SE=.09), r=.57***	M=4.92 (SE=.09), r=.55***	M=4.95 (SE=.08), r=.56***
Liberal	M=4.91 (SE=.18), r=.59***	M=5.07 (SE=.15), r=.61***	M=4.98 (SE=.16), r=.60***	M=4.97 (SE=.15), r=.52***
Neutral	M=4.62 (SE=.15), r=.42***	M=4.58 (SE=.15), r=.48***	M=4.49 (SE=.14), r=.50***	M=4.53 (SE=.12), r=.50***
Conservative	M=5.35 (SE=.20), r=.37***	M=5.34 (SE=.17), r=.57***	M=5.34 (SE=.16), r=.49***	M=5.39 (SE=.15), r=.57***
Non-Political Blogs	M=5.21 (SE=.17), r=.63***	M=5.20 (SE=.14), r=.71***	M=5.23 (SE=.15), r=.68***	M=5.19 (SE=.15), r=.81***
	F[6, 761]=16.912, p ≤ .001	F[6, 748]=23.844, p ≤ .001	F[6, 726]=28.673, p ≤ .001	F[6, 715]=33.839, p ≤ .001

* p ≤ .05; ** p ≤ .01; *** p ≤ .001

+ M=Mean (SE=Standard Error): Min.=1 & Max.=7

+ r= Accumulated correlations among commenters

This page has been intentionally left blank.

WOMAN OR LEADER FIRST? GENDER BIAS IN THE PERCEPTION OF FEMALE LEADER EFFECTIVENESS

Jamie Birdwell, Florida Institute of Technology, Coast oem

ABSTRACT

Even though more women are part of the workforce, women still face considerable obstacles in leadership. Among these obstacles, a subordinate's perception of the leader is most salient. Considerable research has been undertaken to understand the disparity of subordinate perceptive evaluations between male and female leaders. This article examines two questions: Does the gender of the leader influence the subordinates' attitudes regarding the leader's effectiveness? If so, to what extent do women suffer disadvantages from the perceptions of their subordinates? Examined through the lens of role congruity theory, this article shows that women face significant hurdles when compared to men in similar positions; however, research shows potential moderating effects such as leadership style, time, and organizational culture, calling for future research to expand the understanding of the emerging phenomenon better. Managerial implications are presented to aid firms in increasing the leadership effectiveness of women, as well as suggestions to help challenge persistent impediments to successful female leadership.

Keywords: Leadership Effectiveness, Gender Bias, Female Leadership, Role Congruity Theory, Gender Stereotypes

INTRODUCTION

Men continue to occupy significantly more leadership positions than women, even though more women are part of the workforce than men. As gender diversity has increased in employee makeup, issues related to gender bias are salient workplace problems. Compensation and leadership positions for women still trudge behind men. For example, in 2019, women earned approximately 79 cents to every dollar earned by males (Elkins, 2020), and data from the US Census Bureau's 2018 American Community Survey showed that, as compared to earnings of Caucasian men, Asian women earned 97%, Caucasian women 80%, African American women 66%, and Hispanic women 58% (Sheth, Gal, Hoff, & Ward, 2020). Further census data shows that managerial positions occupied by women are 38%, and, in Fortune 500 companies, women account for only 4.2% of chief executive positions (Waber, 2014). Given this minority-related disadvantage, women leaders tend to engage in more masculine leader behaviors (Eagly, Nater, Miller, Kaufmann, & Sczesny, 2020; Koenig & Eagly, 2014; Saint-Michel, 2018).

As organizational diversity evolves, firms need to be adept and agile to focus on gender inequities. Perceptions, attitudes, and behaviors of both genders may affect organizational culture. Men and women have disparate life experiences, gender socialization, and gender norms, and these differences influence leadership development and promotion (Kiser, 2015). For example, studies have shown that women face issues with advancement past middle management (Schnarr, 2012). Evidence suggests male executives appear to have little concern to promote women to managerial

roles, thereby perpetuating a climate of gender polarization (The Conference Board of Canada & Canada, 2013). As Kiser (2015) states, "if men control promotion and advancement to upper-level positions, and an organization does not promote a culture of gender diversity, women are left behind and do not have the same opportunities as men for promotion" (p. 599).

Much research and many frameworks have been developed to understand and explain the differences between men's and women's status in organizations. Both cultural theories and role theories may be used to improve the understanding of those differences. Cultural theories describe the collectively shared patterns of assumptions, values, and expectations that guide the firm's cognitions, emotions, and perceptions (Koland, 2016; Rosser, 2003). Hence, the culture includes the "collective, subjective logic that forms the unspoken, often unconscious subtext of social life" (Rosser, 2003, p. 74).

Like cultural theories, role theories focus attention on expectations and norms for gender-specific behaviors (Saint-Michel, 2018; Vinkenburg, van Engen, Eagly, & Johannesen-Schmidt, 2011; Zheng, Kark, & Meister, 2018) and are, in the case of this article, concerned with gender differences expressed through disparate expectations of leader behaviors (Bosak & Sczesny, 2011). Ayman (1993) contended that roles are normative and are the primary means by which social judgments are used to identify the "ideal" behavioral patterns (Becker, Ayman, & Korabik, 2002).

Consequently, role theory allows one to compare and contrast qualitatively and quantitatively the perceived norms and behaviors exhibited by men and women in leadership positions. Role theories allow for the empirical study of the influence of followers' perceived expectations on a leader's effectiveness (Rosser, 2003). This article will use role theory to explore perceived expectations of leaders' effectiveness rather than a more macro perspective consistent with cultural theories.
This article contributes to the extant literature by focusing on subordinates' role and their perception of leadership, particularly in their sensitivity, discernment, and evaluation of female leadership effectiveness. Fundamentally, this article seeks to determine the following:

> RQ 1. Does the gender of the leader influence the subordinates' attitudes regarding the leader's effectiveness?
>
> RQ 2. If so, in what way do women suffer disadvantages from the perceptions of their subordinates?

To answer these questions, this article will analyze the follower's perceptions of women leaders compared to men through the lens of role congruity theory. This analysis shows that women face significant hurdles compared to men in similar positions; however, research identifies potential moderating effects such as leadership style, time, and organizational culture.

LITERATURE REVIEW

Research has focused on a more comprehensive and better understanding of the tensions between the female gender and leadership role(s). Since the 1970s, several theories have been put forward to explain the undervaluation of women as leaders, including the lack of fit theory (Heilman, 2001),

expectation theory (Berger, Fisek, Norman, & Zelditch, 1977; Ridgeway, 2001; Ridgeway & Jacobson, 1977), the *think-manager-think-male* paradigm (Schein, 1973), and role congruity theory (Eagly & Karau, 2002). Psychological research has shown people tend to attribute the same characteristics to both men and leaders while attributing less to women and managers (Klatt, Eimler, & Krämer, 2016).

Because of a perceived incongruity between female traits and the masculinity of leadership, researchers have focused on the deficits and obstacles of women becoming leaders (Eagly & Karau, 2002; Rosette & Tost, 2010). Indeed, Festing, Knappert, and Kornau (2015) argue that one of the most significant obstacles women face is the lack of a cultural fit between female values and male-dominated leadership roles, especially in the highest level of management. Women in situations and settings that align more with the perceived female gender role tend to experience fewer obstacles and have higher evaluations (Ferguson, 2018; Ridgeway, 2001).

Extensive literature (Avolio, Walumba, & Weber, 2009; Brands, Menges, & Kilduff, 2015; Collins, Burrus, & Meyer, 2014; Newman & Butler, 2014; Post, 2015) on the dyadic relationship between leader and follower shows that this relationship has a substantial impact on job performance, job satisfaction, organizational commitment, and employee turnover (Geys, 2014). Many of the characteristics of leadership are more congruent to male, agentic qualities than female, communal traits (Díaz-García, 2010; Eagly & Karau, 2002; Garcia-Retamero & López-Zafra, 2006; Hoover, Hack, Garcia, Goodfriend, & Habashi, 2019; Lemoine, Aggarwal, & Steed, 2016); therefore, the assumption for leadership is more associated with masculinity than femininity (Garcia-Retamero & López-Zafra, 2006; Koenig & Eagly, 2014; Saint-Michel, 2018; Wang, Chiang, Tsai, Lin, & Cheng, 2013). However, Paustian-Underdahl, Walker, and Weber (2014) contend that academic study and debate between gender researchers and leadership scholars should not focus on the perceptions of gender differences between men and women. Instead, they argue, the discussion should center on how, when, and why there may be gendered differences in subordinate perceptions of the leader's effectiveness.

Foundationally, Eagly, Karau, and Johnson (1992) undertook a comprehensive meta-analysis of gender-based evaluation of leaders. Their investigation included a sample of sixty-one studies and 147 units. They found that the extant empirical literature of the time showing that women are devalued in leadership roles is significant yet divergent (Rosser, 2010). Further findings indicated that female leaders tended to be more negatively evaluated than men when they engaged in autocratic (agentic) behavior (Eagly et al., 1992; Rosser, 2010). Additionally, their findings indicated female leaders "tended to be especially devalued when they direct male subordinates" (Rosser, 2010, p. 73). Eagly et al.'s (1992) research suggested that traditional leadership stylings are perceived to be masculine and favorable for male leaders, while participatory leadership was characterized as female and less favorable.

Leadership in the United States is typically associated with masculinity (Eagly, Wood, & Diekman, 2000; Eagly & Carli, 2007; Hernandez Bark, Escartin, Schuh, and van Dick, 2016; Jogulu & Wood, 2006; Koenig & Eagly, 2014; Paustian-Underdahl et al., 2014; Sojo, Wood, Wood, & Wheeler, 2016; Wang et al., 2013; Wood & Eagly, 2015). Gender stereotypes are automatically triggered cognitively because of the perception of the gender of the person (Eagly & Carli, 2007; Haslam & Renneboog, 2011; Robinson & Lipman-Blumen, 2003). Although

contemporary literature and its descriptions of managerial roles do include some stereotypically feminine elements (being helpful, understanding) and gender-neutral stereotypes (intelligent, dedication), most qualities are as demonstrably masculine as they were in the 1950s (Carli & Eagly, 2016; Eagly & Carli, 2007; Elprana, Felfe, Stiehl, & Gatzka, 2015; Hearn, 2019; Vinkenburg et al., 2011).

Perceptions of leadership as masculine does yield a variety of effects. Since people attribute more authority to men, men tend to be more influential in groups than women. Therefore, women may try to adopt those more agentic traits that are so stringently ascribed to men to be perceived as capable of leading effectively (Carli, 2001; Hernandez Bark et al., 2016; Lemoine et al., 2016). Further, women face discrimination relative to many leadership positions, especially in male-dominated areas, where those positions are perceived to be incongruous with the female's gender roles (Eagly et al., 1995; Eagly & Karau, 2002; Heilman & Okimoto, 2007; Kalysh, Kulik, & Perera, 2016). Research (Garcia-Retamero & López-Zafra, 2006; Schein, 2007) asserts that the masculine stereotype of leadership may contribute to the segregation of the workforce whereby "men tend to monopolize positions of authority" (Garcia-Retamero & López-Zafra, 2006, p. 52). Van Engen et al. (2001) argued that how women think about themselves as potential leaders is a consequence of leadership's perceived masculinity.

Rudman, Moss-Racusin, Phelan, and Nauts (2012) and Rudman and Glick (2001) found that their followers consider women who exhibit agentic qualities to be less fit for positions that require communal traits such as interpersonal orientation. Research of Heilman and Chen (2005) contended that women are expected to be altruistic and helpful. Followers viewed the female leaders negatively if they were not, while men were seen more positively when male leaders engaged in altruistic behaviors but were not penalized when they did not. Rhee and Sigler (2015) found that subordinates evaluated female leaders more negatively for both authoritarian and participatory managerial leadership styles than males who used similar leadership styles. Further, they found that male leaders were equally preferred and perceived as effective whether their style was authoritarian or participatory.

ROLE CONGRUITY THEORY

Gender Role Stereotypes. Gender roles are "beliefs…expectations…in that they [gender roles] are normative and…describe qualities or behavioral tendencies believed desirable for each sex," and there is a perception inferred between the role and the people who are engaged in that role as possessing traits or characteristic of the needs of that specific role (Eagly & Karau, 2002, p. 574). The effect of gender roles influences organizational behavior in the way people react to leaders in terms of gender expectancies (de Klerk & Verreynne, 2017; Vinkenburg et al., 2011) and the extent to which people internalize their gender role (Hernandez Bark et al., 2016; Koenig, Eagly, Mitchell, & Ristikari, 2011). Gender roles are deeply rooted in societal expectations and norms (Bosley, 2018).

Gender role stereotypes contain both descriptive and prescriptive norms for gendered behavior, describing the expectations about who men and women are as well as who they should be (Eagly & Carli, 2007; Eagly & Karau, 2002; Ferguson, 2018; Onesto, 2017; Vinkenburg et al., 2011) and are deeply rooted in societal norms and expectations (Bosley, 2018). "Descriptive norms are the

shared beliefs about what men and women actually do, whereas prescriptive norms are the shared understandings of what men and women ought to do" (Ferguson, 2018, p. 410). Therefore, the agentic gender role describes and prescribes that men should be assertive, independent, and self-confident; however, the communal gender role describes and prescribes that women should be expressive, helpful, and nurturing. Role incongruity occurs when a woman or man performs social roles that do not align with the descriptive and prescriptive norms for their gender, whereas role congruity occurs when the two social roles align. For example, if a woman is performing a nurturing role, such as a caregiver, and fits the stereotyped behavior and expectations, she will experience role congruity. However, if a man were to perform that same role, he would suffer role incongruity because the social expectation is incompatible with the stereotypical gender role expectations.

Role Theories. Role theories focus on expectations and norms (Billing & Alvesson, 1994) and are concerned with gender differences expressed through disparate expectations and behaviors (Billing & Alvesson, 1994; Billing, 2011; Bosley, 2018). Ayman (1993) contended that roles are normative and are the primary means by which social judgments are used to identify the "ideal" behavioral patterns. Eagly and Carli (2007) maintained that people tend to classify male and female leaders according to how they associate their feelings about men and women in general. Once they have categorized a person as a leader, they frame that person into their expectations based on those associations (Eagly & Carli, 2007).

As postulated by Eagly and Karau (2002) in their seminal article, role congruity posits that the underrepresentation of women in leadership positions may result from gender role stereotypes held by followers. Fundamentally, role congruity theory assays the association between gender roles and the other social roles individuals enact (Eagly & Karau, 2002; Ferguson, 2018). Ridgeway (2013) argued that people's primary means of understanding and engaging in social behavior is primarily based on gender. Therefore, women may suffer disadvantages in leadership because of prejudice against female leaders as well as resistance when women occupy leader roles (Eagly, 2007; Peachey & Burton, 2011).

Eagly and Karau's theory also postulates that leadership stereotypes exist, and such stereotypes are the basis for expectations of leadership qualities for leaders (Vu, Duong, Barnett, and Lee (2017). Because of these stereotypes, men are seen as possessing the necessary and inherent qualities, those more "congruent," of leadership, whereas women are not perceived as having the requisites skills to lead efficiently (Vu et al., 2017). Significant research shows that all leadership roles, whether the role calls for communal "feminized" traits and skills, entail some degree of gender-role incongruity, and supervisors are anticipated to act agentically (Eagly & Karau, 2002; Heilman & Caleo, 2018; Koenig et al., 2011; Lyness & Heilman, 2006; Powell, 2011).

Role Congruity Theory. Role congruity theory is an evolutionary step from social role theory whereby perceivers infer a correspondence between their actions and their inner dispositions. Therefore, the descriptive nature of gender roles originates in others' beliefs rather than in the person demonstrating the behavior (Cenkci & Ozcelik, 2015). The supposition here is that certain traits and behavioral tendencies are perceived to be desirable for each gender, as well as the social expectations of the roles that men and women should occupy (Eagly & Karau, 2002; Lemoine et al., 2016; Peachey & Burton, 2011). Role congruity theory advanced the notion that women leaders

face potential prejudice since leadership ability is generally ascribed to men who display agentic qualities rather than women who exhibit communal (Eagly & Karau, 2002; Gloor, Morf, Paustian-Underdahl, & Backes-Gellner, 2020; Triana, Richard, & Yücel, 2017). Further, role congruity theory contends that women (1) may be disadvantaged in securing leadership positions because they are not perceived as qualified because they lack the necessary leadership skills, and (2), even if a woman occupies a leadership position, she may be unfavorably evaluated because she "may be perceived as violating the gender norm ascribed to women" (Peachey & Burton, 2011, p. 418). Empirical research has shown that the masculinity of leader role expectations are, in fact, very robust and have been found across nationalities (Cenkci & Ozcelik, 2015; Schein, 2001), in a variety of industries/services (Ko, Kotrba, & Roebuck, 2015), and by a variety of methods (Rosette & Tost, 2010).

Role congruity has profound effects on discrimination and prejudice in that leaders are perceived as legitimate if that role aligns with social expectations (Ferguson, 2018). Eagly and Karau (2002) theorized that role incongruity leads to two forms of discrimination and prejudice toward female leaders. First, because of descriptive gender norms, people have *a priori* beliefs about males and females based on stereotyped gender roles. Because agentic qualities ascribed to men closely resemble those attributed to leadership more so than do the communal attributes ascribed to women (Lemoine et al., 2016), women are more likely to experience impediments to access these roles if they engage in behavior incongruous with perceived expectations. Second, when men and women who already perform or participate in social roles that deviate from expected gender roles, they will experience prejudice and discrimination. Communal performance by men or agentic actions by women is evaluated less favorably and seen as less legitimate because of a violation of prescriptive gender norms about what men and women ought to do (Eagly & Karau, 2002; Fritz & van Knippenberg, 2017; Saint-Michel, 2018).

Research is beginning to show that role congruity may be moderated by leadership style, especially when considering that women have been more favorably evaluated than men when they display transformational leadership (Cuadrado, Navas, Molero, Ferrer, & Morales, 2012; Eagly & Carli, 2007; Koenig et al., 2011; Paustian-Underdahl et al., 2014; Peachey & Burton, 2011), as well as the gender of the followers (Gloor et al., 2020) as the perceived masculinity of leadership moderates over time (Koenig et al., 2011; Paustian-Underdahl et al., 2014). Although an in-depth exploration of leadership styles and theories is beyond the scope of this article, transformational leadership has been shown to be more congruent with the stereotypical female gender role, whereas transactional leadership is ascribed more with a masculine gender role (Brescoll, 2016; Kiser, 2015; Koenig et al., 2011; Saint-Michel, 2018). Since positive organizational outcomes have been found to be related to transformational leadership, women leaders who display transformational qualities may have a leadership advantage over men (Rosette & Tost, 2010; Saint-Michel, 2018; Wang et al., 2013). Peachey and Burton (2011), in an empirical study of perceptions of the effectiveness of male and female leaders in intercollegiate athletic directors, argued that organizational context might mediate this female leadership advantage. Their study's results provided evidence that, indeed, organizational context does seem to mediate female leadership advantage through transformational leadership. Regardless of leadership style, subordinates' gender role expectations and biases about the leader may impact the leader's perceived effectiveness (Gloor et al., 2020; Koenig et al., 2011).

PERCEPTIONS OF LEADERSHIP EFFECTIVENESS

Subordinate perceptions of the leader are salient to understanding the leader's effectiveness. Research shows that women leaders are disliked more than men who occupy similar positions and roles (Fitzsimmons, Callan, & Paulsen, 2012; Heilman & Okimoto, 2007) and face social and economic penalties (Rosette & Tost, 2010) when they enact or direct authority (Brescoll, 2011). Rosser (2003) hypothesized that perceptions are rooted in an individual's experiences with a leader's behavior, directly or indirectly, based on what a leader says and does. The contextual and cognitive factors which influence gender perception include (1) the beliefs, expectations, and motivator between the leader and the perceiver; (2) whether or not elements inform the gender schema among which are the nature of the task, the characteristics/traits of the leader, and the organizational context; (3) whether or not differential expectations are conveyed to women and men; and (4) whether these result in adaptation of the leader's self-protection (Becker et al., 2002).

Research has long held that social perceivers hold stereotypic beliefs about groups (Katz & Braly, 1933; Vinkenburg et al., 2011) and their leadership. Therefore, individuals determine whether or not they believe the leader(s) is(are) effective. Moreover, Heilman (2001) asserts women who aspire to leadership positions face a perceived lack of fit. Both the lack of fit (Heilman, 2001) and the think-manager-think-male paradigm (Groeneveld, Bakker, & Schmidt, 2020; Schein, 1973) empirically describe the barriers women may face in achieving leadership positions. Women are not believed to have the perceived requisites to assuming a leadership position as males--the same abilities, traits, and talents (Klatt et al., 2016). Role congruity theory has shown that perceptions of role incompatibility have profound adverse effects on a woman's leadership effectiveness (Bongiorno, Bain, & David, 2014; Gloor et al., 2020; Rosette & Tost, 2010; Saint-Michel, 2018), leader emergence (Lemoine et al., 2016) and perceptions of leadership style and ability (Ellemers, Rink, Derks, & Ryan, 2012; Esser, Kahrens, Mouzughi, & Eomois, 2018; Porter, 2009).

Subordinate's Leadership Expectations. Individuals/subordinates have expectations for how women and men should behave in general and how male and female leaders should act. Historical research by Lord & Maher (1993) maintained that as a leader behaves with a subordinate's expectations consistently, subordinates rate the leader as effective. Indeed, Lord and Maher's (1993) leadership categorization theory posits that individuals hold mental representations (prototypes) by which leaders should behave. A person's prototype affects many aspects: his/her attention, encoding, and schema information (Phillips & Lord, 1982), and physical features associated with sex, ethnicity, and race may activate prototypes which may affect perceiver's expectations for male or female leaders (Lord & Emrich, 2001).

Eagly et al. (1992) found that male leaders were evaluated slightly more favorable than female leaders, primarily when leaders were described using a masculine leadership style; however, current research suggests that perception of the masculinity of a leadership role may be lessening (Koenig et al., 2011; Paustian-Underdahl et al., 2014). Dominant women and women who use more assertive speech are less influential to men than women who are less dominant or assertive (Baxter, 2015), and women who express anger tend to be evaluated less favorably than men (Schaubroeck & Shao, 2012).

Echoed in Rosser's (2003) assertion that organizational context has a strong influence on leader behavior, Becker et al. (2002) found substantial support between the subordinate perception of the leader's effectiveness as a function of the leader's gender and the organizational context. Specifically, they found that being in an out of role position enhanced the follower's perception of both initiating structure and female leaders' consideration. Further, they found that high self-monitoring women leaders had lower self/subordinate discrepancies in the educational field versus industry, suggesting that high self-monitoring women attend more to organizational norms than situational cues. The authors did acknowledge the possibility that industry may be a moderating factor; however, their findings show that high self-monitoring women face less conflict with perceptive behaviors than low-monitoring women (Becker et al., 2002).

Lack of FIT. Historic literature suggested that women who aspire to leadership positions face a perceived lack of fit. Statham (1987) suggested that sex-differentiated management may exist. In her study, she found that female leaders were perceived as more task-oriented and focused on their followers. Women leaders interacted more with subordinates, and followers felt that the female leaders paid close attention to everything happening within their purview. On the other hand, she found men were perceived to be more self-focused, more distant from their followers, and emphasized their power over people and situations. Followers felt that the male leader led from a distant, "stay out of it" manner.

More currently, Paustian-Underdahl et al. (2014), in a meta-analytic study of forty-nine years of research of contextual moderators of perceptions of leadership effectiveness, presented several notable findings. First, they found the moderating effect of the relationship between a leader's role and the perception of that leader's effectiveness, observing that different results occur depending on whether self- or other-ratings serve as the measure of the leader's effectiveness. Second, consistent with role congruity theory, gender differences were moderated by the extent to which the examined firm was male or female-dominated. The perception tended to view the male leader as more effective if the organization was male dominant. Paustian-Underdahl et al. (2014) found the results were consistent with Eagly et al.'s (1995) results, although the effects seem to have waned over time (again suggesting time may have a moderating effect). Paustian-Underdahl et al.'s quantitative findings showed d = .12, 95% CI [-.09, .32], while Eagly et al.'s (1995) results were for d = .42, 95% CI [.32, .52].

Third, Paustian-Underdahl et al. (2014) found that specific leadership roles (i.e., business, education) may be seen as incongruent with a male gender role, thereby negatively affecting the perceived success of the male leader. Fourth, women are viewed to be more effective than men in middle management and senior leadership positions (Baker & Cangemi, 2016). Through the lens of Foschi's double standards of leadership model (Cited by Paustian-Underdahl, 2014) whereby women are perceived to be seen as more effective than men in top leadership positions because of the perceptions of the extra competence of the female leader, Paustian-Underdahl et al. (2014) contended that role congruity theory might be supplemented to illustrate the perceptions of additional competence may override perceptions of women's incongruity. Fifth, gender-related stereotypes regarding perceptions of female leadership effectiveness appear to have slightly shifted from Eagly et al.'s (1995) in that women leaders are being viewed more effectively by both men and women as the number of women has increased in the workplace.

Leadership Style and Gender. Extant research has indicated an interaction between leadership style and leader gender, and that interaction may moderate subordinate evaluation of the leader's effectiveness (Cenkci, & Özçelik, 2015). In a study of Taiwanese employees and how leadership style (authoritarian or benevolent—agentic or communal) interacts with the gender of the leader to impact employee performance, Wang et al. (2013) found evidence that indicated that the supervisors' gender affects the degree of perceived leadership effectiveness. Oakley (2000) showed that when women leaders engage more control-oriented and authoritarian leadership, they are more likely to experience negative feedback from subordinates rather than their male colleagues because of perceived incongruity between gender and the expected role. Research by Cheng & Lin (2012) found that as men engage in authoritarian and dominant behavior over their subordinates, their leadership is perceived as more effective, consistent with role congruity theory (Eagly & Karau, 2002).

Additionally, Cheng and Lin's (2012) research showed that women were expected to enact more benevolent leadership (consistent with perceived communal female traits), and subordinates reacted more positively when female leaders did so. In fact, subordinates performed at lower levels when female leaders showed less benevolent leadership than did subordinates led by men. However, an empirical study by Cenkci & Özçelik (2015) suggests that gender does not significantly affect the relationship between leadership style and employee work engagement. Nevertheless, the authors readily admit that culture could have moderated that relation as their study was conducted solely of Turkish firms, firms that are accustomed to male-dominance and authoritative leadership.

Scott and Brown (2006) found empirical evidence that subordinates have difficulty encoding leadership behaviors into underlying leadership traits when female leaders perform using agentic traits. Their findings suggest that female leaders are subjected to gender roles and leadership comparisons by their subordinates; however, Scott and Brown (2006) did not compare women's results against male leaders. Wang et al. (2013) sought to expand Scott and Brown's (2006) study to include that comparison. Findings from their initial study of fifteen Taiwanese electronic manufactures (mostly male) found that when male leaders engaged in benevolent leadership, their subordinates had stronger positive performances. Additionally, they discovered that authoritarian leadership and leader gender affect subordinate creativity, although there was no significant effect on task performance. Further findings indicated that female leaders engage in more authoritarian leadership than male leaders with no discernible differences in benevolent leadership between the genders. These findings led them to another study.

In their second study, Wang et al. (2013) collected data from Taiwanese customer service units (mostly female) within a large commercial bank. They found that female leader use of authoritarian leadership had a substantial adverse effect on subordinate performance. They also found that female leaders' benevolent leadership had a less favorable impact on subordinate task performance than men. The collaborating effect of benevolent leadership and leader gender on the follower's conscientiousness was not significant. In Study 2, their findings showed that female leaders' benevolent leadership was positively associated with follower behavior, suggesting that women leaders are more effective in predominantly female settings, consistent with Eagly et al.'s (1995) findings. However, in both studies, they found that female leaders' use of benevolence had weaker effects on subordinate performance than male leaders, especially in male-dominated settings.

Johnson, Murphy, Zewdie, and Reichard (2008) conducted a series of four studies to provide an in-depth study of role congruity theory; to understand whether current leadership prototypes are complementary to sex roles; and to explain the extent to which gender role violations influence leadership effectiveness ratings. As a whole, these empirical studies demonstrate that leadership prototype dimensions do differ in salience between men and women, supporting role congruity theory. Additionally, an individual's sex-type (feminine or masculine) affects their opinions of a leader's gender consistent actions. Their studies suggest that male leaders tend to be perceived by subordinates as more effective than female leaders, *ceteris paribus*. They found that sex disparities appeared more pronounced for leadership effectiveness ratings than other factors such as likeability. Further findings indicated that female followers expect that their female leaders should also be sensitive. Those women who did not display sensitivity were rated as less effective. However, contrary to role congruity theory, Johnson et al. (2008) found that sensitive female leaders were liked as well as influential female leaders, but both were liked more than sensitive male leaders.

Moderating Effects and Evolving Attitudes. However, not all researchers agree. Bosak and Sczesny (2011) sought to show that, as more women continue to enter the workforce, especially those in male-dominated areas, the perceived incongruity associated with traits ascribed to each gender (and, thereby, stereotypes) would decrease over time. In their study, their second hypothesis posited that the perceived variation in women's roles would mediate the effects on masculine traits attributed to women, and perceived adjustment(s) in leadership abilities will mediate the effects on feminine characteristics associated with leaders. Their empirical study of 160 management students recruited from the University of Bern, Switzerland, found that participants projected that women would significantly increase in masculine traits over time, reducing the perceived role incongruity between gender and leader. The traits ascribed to men and leadership remained stable over time. Further, they found evidence to suggest that perceived changes in women's roles did mediate the impact on beliefs about masculine traits in female leadership.

They conducted a second study with 196 participants (106 male, 90 female) at the University of Bern. In this study, Bosak and Sczesny (2011) posited that the incongruity in beliefs about leaders and women should diminish as role distributions shift due to a projected (manipulated for the experiment) increase in women's masculine traits. Their second study found evidence that individuals perceive leadership as a masculine behavior, especially in male-dominated work settings. As women became leaders, their findings showed that followers assigned masculine traits to the female leaders and evaluated them equal to men in the same leadership position. Additionally, they found that male leaders increased in feminine traits as the role distributions became equal. Because that was a surprise and could not be easily explained, Bosak and Sczesny postulated that that shift occurred potentially due to the manipulation of the experiment. Further research in this area is needed.

In an empirical study of 2267 people (1546 male and 721 female) using a customized 360-degree feedback evaluation, Pfaff, Boatwright, Potthoff, Finan, Ulrey, & Huber (2013) found that subordinates' perceptions of the female leaders use of nine of ten task-oriented behaviors were much higher than for men. Women utilized relationally oriented actions more so than their male counterparts did. However, subordinates and other leaders/supervisors perceived that women and men leaders use task-oriented behavior almost equally, while female middle managers view

themselves as engaging in task-oriented behaviors more than men recognize themselves in the same regard. Since task-oriented behaviors are historically associated with agentic leadership, these findings do not support role congruity. Rather, these results seem to confound more than clarify. Future research into understanding this disparity is warranted.

Exposure to Female Leaders. Dasgupta & Asgari (2004) identify research that demonstrates that some automatic attitudes and beliefs (stereotypes, for example) may be malleable when specific motivational factors spur people or when individuals practice specific strategies to avoid stereotypical or prejudicial responses. Their empirical research showed that seeing women in high profile *and* counter-stereotypical leadership positions (business leaders, judges, scientists, etc., those considered more masculine) has a robust impact on women's unconscious beliefs about their ingroup (Dasgupta & Asgari, 2004). As the female participants in their study were exposed to more famous female leaders, they were more likely to associate leadership traits to women more than those who were exposed to controlled exemplars (Dasgupta & Asgari, 2004).

In their second study that involved students in a university setting, Dasgupta and Asgari (2004) found a mediating factor: how often people are exposed to women in leadership mediates their gender-specific attitudes, stereotypes, and beliefs. In other words, results showed that as more of the students encountered female instructors, the less they expressed gendered stereotypes automatically. Additionally, female students showed a marked increase in women's automatic stereotypical responses to other women if they encountered mostly male faculty members. Finally, their studies support social role theory by showing that women are much more likely to be involved in communal care-taking roles while men are disproportionally located in more authoritative, agentic roles, both of which perpetuate gender stereotypes. Following these studies, Dasgupta and Asgari (2004) suggested that changes in automatic gender-stereotypical responses demonstrated in their studies could potentially decrease the perceived incongruity between gender roles and leadership roles. They maintain that differences in perceived discrepancies and automatic gender stereotypes may offer women access to more agentic leadership positions, provide perceivers more opportunities to evaluate women who already occupy agentic leadership roles as more effective, and reduce the power of gender stereotypes through a conscious belief change over time (Lemoine et al., 2016).

Empirical findings from Yang and Aldrich's (2014) research of entrepreneurial teams and start-ups suggested that women would generally face fewer disadvantages in business leadership, except for male/female perceived spousal characteristics (male as the breadwinner, the female as a homemaker) and related accountability. Yang & Aldrich (2014) found that ascribed attributes, despite growing egalitarian value systems, remain a fundamental basis for assigning rewards and distributing leadership. Further findings indicated that gender acts as an underlying and often unacknowledged "cultural and cognitive principle lurking beneath the surface of all social interactions" (Yang & Aldrich, 2014, p. 322).

Among the most significant sociocultural threat to leadership gender equality, especially in the highest levels of leadership, maybe the lingering stereotypes and implicit biases regarding the gendered idea of leadership roles (Pfaff et al., 2013). Pfaff et al.'s research shows that women in middle management are perceived as being prepared both cognitively and behaviorally to ascend to the most elite levels of firm leadership. Empirical data collected by Festing et al. (2015)

indicated that gender-related biases on HRM practices might perpetuate male-oriented leadership practices and prevent women from attaining higher levels of leadership. Beaman, Chattopadhyay, Duflo, Pande, and Topalova (2009) found that prior exposure to female leaders improves perceptions of leader effectiveness for women in general and reduces the influences of gender stereotypes.

LIMITATIONS

This article is limited to the degree of the scope of the surveyed articles, data, and research. Additionally, this research reviews the extant literature and does not conduct any empirical or quantitative study. Instead, it relies on the data collected from other researchers. This paper did not evaluate the degree of leader power and its potential influence(s) on a subordinate's reaction. As far back as 1991, Ragins suggested that gender differences regarding power may significantly impact the leader-subordinate exchange. She argues that female leaders lacking in power will lose subordinate support. This area may be fruitful for further research but is not examined in this setting. This article also did not profoundly consider the distinctive leadership stylings or choices and how those may impact the subordinate's perception of leadership effectiveness beyond what was presented.

In this research, gender was viewed through biological sex rather than as a spectrum or construct. For the purpose of this study and following research by psychologists Ellemers (2014, 2018) and Ramos, Barreto, Ellemers, Moya, Ferreira, & Calanchini (2016), gender was regarded as the observations of how men and women behave as explained by inherent biological differences between them. Other than a cursory review, research into potential moderating effects of disparate cultures was not thoroughly explored.

IMPLICATIONS

Firms must do more than merely placate women or encourage their ascension in the ranks of leadership. Women must be empowered, given proper training, and afforded the opportunities to lead. "As long as companies do not go beyond the usual measures…they will only address the symptoms" of the problem (Festing et al., 2015, p. 74). For managers, identifying the requisite leadership style to fit the followers' needs is paramount; for female leaders, it is critical (Eagly & Karau, 2002). Understanding that gender stereotypes and role congruities influence style will allow both men and women to identify the appropriate situational leadership style. Further, followers may be trained to better understand their roles in the leadership-followership dyad to alleviate or prevent a stereotypical reaction to the leader's gender.

Granted, over time, the stereotypical masculinization of leadership has waned (Koenig et al., 2011; Paustian-Underdahl et al., 2014); however, women still must be higher qualified than men, outperform to receive a promotion, and receive less pay than similarly positioned men with commiserate expertise, training, and tenure (Lammers & Gast, 2017). Additional workforce training is costly, especially if the additional training does not transfer to the entire organization (Gloor et al., 2020). Gloor et al. (2020) suggest that women face a potential backlash if women are trained to use more agentic characteristics, and organization-wide interventions and training may

solidify stereotypes. Their suggestion of "team-based interventions" (p. 507) echoes research that more diverse teams tend to diminish gender bias (Kalysh et al., 2016). Targeted training using diverse teams may help mitigate training costs while potentially reducing gender inequity (Gloor et al., 2020; Kalysh et al., 2016; Onesto, 2017).

Additionally, firms should explore ways that diverse teams may be utilized to increase the organizational culture's attitudes towards diversity and inclusion. Socialization activities where men and women collaborate in small team settings have been shown to help break down stereotypical beliefs (Madsen & Scribner, 2017). Reviewing or adopting espoused corporate values that incorporate diversity in the workforce, including leadership positions, may promote more positive attitudes toward all minorities, not just women.

Companies and managers should create a mentoring program that will help women in the job promotion process and the acclimation to its organizational culture. Mentoring should not be considered merely as a part of the onboarding process. On the contrary, continued mentorship over time aids both women and minorities with increased job satisfaction and retention (Fuller, Cliffe, & Moorosi, 2015; Roebuck & Smith, 2011).

FUTURE RESEARCH

Some areas for future research have already been noted. A model created by evidence gathered by Vial, Napier, and Brescoll (2016) provides an opportunity to explore the moderating effect of power within the self-reinforcing cycle of illegitimacy faced by women leaders. Since power and its results were beyond this article's scope, an analysis of their research and testing their substantial model may explain the dichotomous perceptions between male and female leaders. Following the example of Bosak and Sczesny (2011), researchers should continue long-term, longitudinal empirics to determine if time has had a moderating effect on follower attitudes toward female leaders. In other words, as the baby boomers leave the workforce, what impact has generation X or millennials had and did each generation view female leaders differently? These two questions should generate much research and debate.

Additionally, climate and cultural impacts cannot be neglected, and continued study in these areas will be beneficial. Wang et al.'s (2013) and Bosak and Sczesny's (2011) research suggest cultural differences may influence the subordinate's response to and evaluation of a leader based on gender. The degree to which stereotypes are pervasive may result from cultural norms and may vary between large cultural groups (Hispanic as compared to Asian, for example) and within subgroups (Cuban, Mexican, Japanese, and Chinese). Hofstede's research could be compared with other scholarship to examine the cultural impact of gender, and, in particular, how team training may challenge stereotypes and socialize newer norms in smaller settings by culturally and gender diverse teams. Finally, research examining the question "do women face continued obstacles to leadership in senior levels?" may provide useful information for promoting and retaining high-quality female leaders.

CONCLUSION

Women have faced obstacles and difficulties in attaining and possessing leadership positions at all levels. The article analyzed gender through the lens of role congruity theory, as postulated by Eagly and Karau (2002), to determine to what extent followers perceived gendered differences in leadership effectiveness. Much research, both empirical and theoretical, shows that women leaders are perceived differently by followers as compared to their male counterparts. Several studies point to the moderating effects of leadership style and culture. Additionally, some studies (Bosak and Sczesny, 2011; Pfaff et al., 2013) present findings that show the potential that time moderates attitudes and behaviors toward women leaders, thereby leveling perceptions toward genders in leadership. While much study has occurred, future areas, especially those involving power as moderators, may provide a better understanding of a leader's perceived effectiveness between genders.

REFERENCES

Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, *60*(1), 421–449.

Ayman, R. (1993). Leadership theory and research: Perspectives and directions. In R. Chemers, M. M, Ayman (Ed.). Academic, New York.

Baker, J., & Cangemi, J. (2016). Why are there so few women CEOs and senior leaders in corporate America? *Organization Development Journal, 34*(2), 31–43.

Baxter, J. (2015). Who wants to be the leader? The linguistic construction of emerging leadership in differently gendered teams. *International Journal of Business Communication*, *52*(4), 427–451.

Becker, J., Ayman, R., & Korabik, K. (2002). Discrepancies in self/subordinates' perceptions of leadership behavior. *Group & Organization Management*, *27*(2), 226–244.

Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M. (1977). *Status characteristics and social interaction*. New York: Elsevier Scientific.

Billing, Y. D. Alvesson, M. (1994). *Gender, managers, and organizations*. Berlin: Walter du Gruyter.

Billing, Y. D. (2011). Are women in management victims of the phantom of the male norm? *Gender, Work and Organization*, *18*(3), 298–317.

Bongiorno, R., Bain, P. G., & David, B. (2014). If you're going to be a leader, at least act like it! Prejudice towards women who are tentative in leader roles. *British Journal of Social Psychology*, *53*(2), 217–234.

Bosak, J., & Sczesny, S. (2011). Exploring the dynamics of incongruent beliefs about women and leaders. *British Journal of Management*, *22*(2), 254–269.

Bosley, E. (2018). *Developing their voices: The experiences of women senior executives in federal government as they develop voice*. [The George Washington University].

Brands, R. A., Menges, J. I., & Kilduff, M. (2015). The leader-in-social-network schema: Perceptions of network structure affect gendered attributions of charisma. *Organization Science*, *26*(4), 1210–1225.

Brenner, O. C., Tomkiewicz, J., & Schein, V. E. (1989). The relationship between sex role stereotypes and requisite management characteristics revisited. *Academy of Management Journal*, *32*(3), 662–669.

Brescoll, V. L. (2011). Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, *56*(4), 622–641.

Brescoll, V. L. (2016). Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders. *Leadership Quarterly*, *27*(3), 415–428.

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues*, *57*(4), 725–741.

Carli, L. L., & Eagly, A. H. (2016). Women face a labyrinth: An examination of metaphors for women leaders. *Gender in Management*, *31*(8).

Cenkci, A. T., Ozcelik, G. (2015). Leadership styles and subordinate work engagement: The moderating impact of leader gender. *Global Business and Management Research: An International Journal*, *7*(4), 8–20.

Cheng, M., & Lin, Y. (2012). The effect of gender differences in supervisors' emotional expression and leadership style on leadership effectiveness. *African Journal of Business Management*, *6*(9), 61–72.

Collins, B. J., Burrus, C. J., & Meyer, R. D. (2014). Gender differences in the impact of leadership styles on subordinate embeddedness and job satisfaction. *Leadership Quarterly*, *25*(4), 660–671.

Cuadrado, I., Navas, M., Molero, F., Ferrer, E., & Morales, J. F. (2012). Gender differences in leadership styles as a function of leader and subordinates' sex and type of organization. *Journal of Applied Social Psychology*, *42*(12), 3083–3113.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 642–658.

de Klerk, S., & Verreynne, M. L. (2017). The networking practices of women managers in an emerging economy setting: negotiating institutional and social barriers. *Human Resource Management Journal*, *27*(3), 477–501.

Díaz-García, M. C., & J.-M. (2010). Entrepreneurial intention: The role of gender. *International Entrepreneurship and Management Journal*, *6*, 261–283.

Eagly, A. H. (2007). Female leadership advantages and disadvantages: Resolving the contradictions. *Psychology of Women Quarterly*, *31*(1), 1–12.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth*. Harvard University Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598.

Eagly, A. H., Karau, S. J., & Johnson, B. T. (1992). Gender and leadership style among school principals. *Educational Administrative Quarterly*, *28*(1), 76–102.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, *117*(1), 125–145.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*, 3–22.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes, M. T., Trautner (Ed.), *The developmental social psychology of gender* (Mahwah, NJ, pp. 123–174). Erlbaum.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315.

Elkins, K. (2020). Here's how much men and women earn at every age. *CNBC.* Retrieved from https://www.cnbc.com/2019/04/02/heres-how-much-men-and-women-earn-at-every-age.html.

Ellemers, N. (2014). Women at work: How organizational features impact career development. *Policy Insights Behavioral Brain Science*, *1*, 46–54.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298.

Ellemers, N., Rink, F., Derks, B., & Ryan, M. K. (2012). Women in high places: When and why promoting women into top positions can harm them individually or as a group (and how to prevent this). *Research in Organizational Behavior*, *32*, 163–187.

Elprana, G., Felfe, J., Stiehl, S., & Gatzka, M. (2015). Exploring the sex difference in affective motivation to lead: Furthering the understanding of women's underrepresentation in leadership positions. *Journal of Personnel Psychology*, *14*(3), 142-152.

Esser, A., Kahrens, M., Mouzughi, Y., & Eomois, E. (2018). A female leadership competency framework from the perspective of male leaders. *Gender in Management*, *33*(2), 138–166.

Ferguson, T. W. (2018). Female leadership and role congruity within the clergy: Communal leaders experience no gender differences yet agentic women continue to suffer backlash. *Sex Roles*, *78*(5–6), 409–422.

Festing, M., Knappert, L., & Kornau, A. (2015). Gender-specific preferences in global performance management: An empirical study of male and female managers in a multinational context. *Human Resource Management*, *51*(1), 55–79.

Fitzsimmons, T. W., Callan, V. J., & Paulsen, N. (2012). Gender disparity in the c-suite: Do male and female CEOs differ in how they reached the top? *Leadership Quarterly*, *32*(2), 245–266.

Fritz, C., & van Knippenberg, D. (2017). Gender and leadership aspiration: The impact of organizational identification. *Leadership and Organization Development Journal*, *38*(8), 1018–1037.

Fuller, K., Cliffe, J., & Moorosi, P. (2015). Women's leadership preparation within the senior leadership team. *Planning and Changing*, *46*(3), 388–415.

Garcia-Retamero, R., & López-Zafra, E. (2006). Prejudice against women in male-congenial environments: Perceptions of gender role congruity in leadership. *Sex Roles*, *55*(1–2), 51–61.

Geys, B. (2014). Employees' preferences toward (un) likable managers. *The Leadership Quarterly*, *25*(5), 875–884.

Gloor, J. L., Morf, M., Paustian-Underdahl, S., & Backes-Gellner, U. (2020). Fix the game, not the dame: Restoring equity in leadership evaluations. *Journal of Business Ethics*, *161*(3), 497–511.

Groeneveld, S., Bakker, V., & Schmidt, E. (2020). Breaking the glass ceiling, but facing a glass cliff? The role of organizational decline in women's representation in leadership positions in Dutch civil service organizations. *Public Administration*, *98*, 441–464.

Haslam, S. A., & Renneboog, L. (2011). Who gets the carrot and who gets the stick? Evidence of gender disparities in executive remuneration. *Strategic Management Journal*, *321*(July 2010), 301–321.

Hearn, J. (2019). Gender, work and organization: A gender – work – organization analysis. *Gender, Work & Organization*, *26*(October 2018), 31–39.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674.

Heilman, M. E., & Caleo, S. (2018). Combatting gender discrimination: A lack of fit framework. *Group Processes and Intergroup Relations*, *21*(5).

Heilman, M. E., & Chen, J. J. (2005). Same behavior, different consequences: Reactions to men's and women's altruistic citizenship behavior. *Journal of Applied Psychology*, *90*(3), 431–441.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, *92*(1), 81–92.

Hernandez Bark, A. S., Escartín, J., Schuh, S. C., & van Dick, R. (2016). Who leads more and why? A mediation model from gender to leadership role occupancy. *Journal of Business Ethics*, *139*(3), 473–483.

Hoover, A. E., Hack, T., Garcia, A. L., Goodfriend, W., & Habashi, M. M. (2019). Powerless men and agentic women: Gender bias in hiring decisions. *Sex Roles*, *80*, 667–680.

Jogulu, U. D., & Wood, G. J. (2006). The role of leadership theory in raising the profile of women in management. *Equal Opportunities International*, *25*(4), 236–250.

Johnson, S. K., Murphy, S. E., Zewdie, S., & Reichard, R. J. (2008). The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders. *Organizational Behavior and Human Decision Processes*, *106*(1), 39–60.

Kalysh, K., Kulik, C. T., & Perera, S. (2016). Help or hindrance? Work-life practices and women in management. *Leadership Quarterly*, *27*(3), 504–518.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*(3), 280–290.

Kiser, A. I. T. (2015). Workplace and leadership perceptions between men and women. *Gender in Management*, *30*(8), 598–612.

Klatt, J., Eimler, S. C., & Krämer, N. C. (2016). Makeup your mind: The impact of styling on perceived competence and warmth of female leaders. *The Journal of Social Psychology*, *165*(5), 483–497.

Ko, I., Kotrba, L., & Roebuck, A. (2015). Leaders as males?: The role of industry gender composition. *Sex Roles*, *72*(7–8), 294–307.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371–392.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, *137*(4), 616–642.

Koland, D. F. (2015). *A new game: Shifting the leadership culture to close the gender gap in corporate America*. [University of St. Thomas].

Lammers, J., & Gast, A. (2017). Stressing the advantages of female leadership can place women at a disadvantage. *Social Psychology*, *48*(1), 28–39.

Lemoine, G. J., Aggarwal, I., & Steed, L. B. (2016). When women emerge as leaders: Effects of extraversion and gender composition in groups. *Leadership Quarterly*, *27*(3), 470–486.

Lord, R. G., & Emrich, C. G. (2001). Thinking outside the box by looking inside the box: Extending the cognitive revolution in leadership research. *Leadership Development: A Review in Context*, *11*(4), 551–579.

Lord, R. G., & Maher, K. J. (1993). *Leadership and information processing: Linking exceptions and performance*. New York: Routledge.

Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, *91*(4), 777–785.

Madsen, S. R., & Scribner, R. T. (2017). A perspective on gender in management The need for strategic cross-cultural scholarship on women in management and leadership. *Cross Cultural and Strategic Management*, *24*(2), 231–250.

Newman, A., & Butler, C. (2014). The influence of follower cultural orientation on attitudinal responses towards transformational leadership: evidence from the Chinese hospitality industry. *International Journal of Human Resource Management*, *25*(7), 1024–1045.

Oakley, J. G. (2000). Gender-based senior barriers to management positions: Understanding scarcity of female CEOs. *Journal of Business Ethics*, *27*(4), 321–334.

Offor, E. E. (2011). *Analysis of sex stereotyping on women's positive evaluation and promotion to executive leadership roles*. [Univeristy of Maryland].

Onesto, L. M. (2017). *Women's leadership effectiveness within the technology industry: How gender roles and emotional intelligence impact followers' evaluations*. [The Chicago School of Professional Psychology].

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology*, *99*(6), 1129–1145.

Peachey, J. W., & Burton, L. J. (2011). Male or female athletic director? Exploring perceptions of leader effectiveness and a (potential) female leadership advantage with intercollegiate athletic directors. *Sex Roles*, *64*(5–6), 416–425.

Pfaff, L. A., Boatwright, K. J., Potthoff, A. L., Finan, C., Ulrey, L. A., & Huber, D. M. (2013). Perceptions of women and men leaders following 360-degree feedback evaluations. *Performance Improvement Quarterly*, *26*(1), 35–56.

Phillips, J. S., & Lord, R. G. (1982). Schematic information processing and perceptions of leadership in problem-solving groups. *Journal of Applied Psychology*, *67*(4), 486–492.

Porter, D. D. S. (2009). *A study of the perceptions of female leaders' qualifications, leadership style, and effectiveness among elective and selective leaders*. [Clark Atlanta University].

Post, C. (2015). When is female leadership an advantage? Coordination requirements, team cohesion, and team interaction norms. *Journal of Organizational Behavior*.

Powell, G. N. (2011). The gender and leadership wars. *Organizational Dynamics*, *40*, 1–9.

Ragins, B. R. (1991). Gender effects in subordinate evaluations of leaders: Real or artifact? *Journal of Organizational Behavior*, *12*(3), 259–268.

Ramos, M. R., Barreto, M., Ellemers, N., Moya, M., Ferreira, L., & Calanchini, J. (2016). Exposure to sexism can decrease implicit gender stereotype bias. *European Journal of Social Psychology*, *46*(4), 455–466.

Rhee, K. S., & Sigler, T. H. (2015). Untangling the relationship between gender and leadership. *Gender in Management*, *30*(2), 109–134.

Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, *57*(4), 637–655.

Ridgeway, C. L. (2013). Framed by gender: How gender inequality persists in the modern world. *Social Forces*, *92*(1), 401–405.

Ridgeway, C. L., & Jacobson, C. K. (1977). Sources of status and influence in all female and mixed sex groups. *The Sociological Quarterly*, *18*(3), 413–425.

Robinson, J. L., & Lipman-Blumen, J. (2003). Leadership behavior of male and female managers, 1984-2002. *Journal of Education for Business [H.W. Wilson - EDUC]*, *79*(1), 28–33.

Roebuck, D. B., & Smith, D. N. (2011). Wisdom from executive female leaders: What can organizations, executive education programs, and graduate students learn? *Journal of Executive Education*, *10*(1), 43–74.

Rosette, A. S., & Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology*, *95*(2), 221–235.

Rosser, V. J. (2003). Faculty and staff members' perceptions of effective leadership: Are there differences between women and men leaders? *Equity and Excellence in Education*, *36*(1), 71–81.

Rosser, S. V. (2010). *The gender gap in patenting: A feminist issue preventing equality*. (V. Nardi, Ed.). New York: Nova Sciences Publishers, Inc.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*, 165–179.

Saint-Michel, S. E. (2018). Leader gender stereotypes and transformational leadership: Does leader sex make the difference? *Management (France)*, *21*(3), 944–966.

Schaubroeck, J. M., & Shao, P. (2012). The role of attribution in how followers respond to the emotional expression of male and female leaders. *Leadership Quarterly*, *23*(1), 27–42.

Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, *57*(4), 675–688.

Schein, V. E. (2007). Women in management: Reflections and projections. *Women in Management Review*, *22*(1), 6–18.

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite. *Journal of Applied Psychology*, *57*(2), 95–100.

Schnarr, K. (2012). Are female executives finally worth more than men? *Ivey Business Journal*, *1*(1), 1–3.

Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational Behavior and Human Decision Processes2*, *101*(2), 230–242.

Sheth, S., Gal, S., Hoff, M., & Ward, M. (2020). 7 charts that show the glaring gap between men's and women's salaries in the US. Retrieved from https://markets.businessinsider.com/news/stocks/gender-wage-pay-gap-charts-2017-3-1029049751

Sojo, V. E., Wood, R. E., Wood, S. A., & Wheeler, M. A. (2016). Reporting requirements, Targets, And quotas for women in leadership. *Leadership Quarterly*, *27*(3), 519–536.

The Conference Board of Canada, & Canada, T. C. B. of. (2013). He says, she says: gender gap persists in attitudes toward women's advancement in the workplace. Retrieved from www.conferenceboard.ca/press/newsrelease/13-05-15/he_says_she_ says_ gender_gap_persists_in_attitudes_toward_women_s_advancement_in_the_workplace.aspx

Triana, M. D. C., Richard, O. C., & Yücel, İ. (2017). Status incongruence and supervisor gender as moderators of the transformational leadership to subordinate affective organizational commitment relationship. *Personnel Psychology*, *70*(2), 429–467.

Van Engen, M, L., van der Leeden, R., & Willemsen, T. M. (2001). Gender, context, and leadership styles: A field study. *Journal of Occupational and Organisational Psychology*, *74*, 581–598.

Vial, A. C., Napier, J. L., & Brescoll, V. L. (2016). A bed of thorns: Female leaders and the self-reinforcing cycle of illegitimacy. *Leadership Quarterly*, *27*(3).

Vinkenburg, C. J., van Engen, M. L., Eagly, A. H., & Johannesen-Schmidt, M. C. (2011). An exploration of stereotypical beliefs about leadership styles: Is transformational leadership a route to women's promotion? *Leadership Quarterly*, *22*(1), 10–21.

Vu, H. T., Duong, H. T., Barnett, B., & Lee, T. (2017). A role (in)congruity study on Vietnamese journalists' perception of female and male leadership. *Asian Journal of Communication*, *27*(6), 648–664.

Wang, A. C., Chiang, J. T. J., Tsai, C. Y., Lin, T. T., & Cheng, B. S. (2013). Gender makes the difference: The moderating role of leader gender on the relationship between leadership styles and subordinate performance. *Organizational Behavior and Human Decision Processes*, *122*, 101–113.

Wood, W., & Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles*, *73*(11–12), 461–473.

Yang, T., & Aldrich, H. E. (2014). Who's the boss? Explaining gender inequality in entrepreneurial teams. *American Sociological Review*, *79*(2), 303–327.

Zheng, W., Kark, R., & Meister, A. L. (2018). Paradox versus dilemma mindset: A theory of how women leaders navigate the tensions between agency and communion. *Leadership Quarterly*, *29*(5), 584–596.

Blank inside back cover.

QUARTERLY REVIEW OF BUSINESS DISCIPLINES

VOLUME 7 NUMBER 3 NOVEMBER 2020

This issue is now available online at www.iabd.org.

A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES
SPONSORED BY UNIVERSITY OF NORTH FLORIDA
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)