Using Web Data and Services: Technology, Theory and Evidence

Xitong Li

Working Paper CISL# 2014-10

May 2014

Composite Information Systems Laboratory (CISL)
Sloan School of Management, Room E62-422
Massachusetts Institute of Technology
Cambridge, MA 02142

Using Web Data and Services: Technology, Theory and Evidence

By

Xitong Li

M.S., Management Research, MIT, USA, 2013
Ph.D., Control Science and Engineering, Tsinghua University, P.R. China, 2010
B.S., Control Science and Engineering, Tsinghua University, P.R. China, 2004

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2014

Signature of Author: ____

MIT Sloan School of Management
April 28, 2014

Certified By: ____

Stuart E. Madnick
John Norris Maguire Professor of Information Technologies & Professor of Engineering Systems
Thesis Supervisor

Certified By: ____

John Hauser
Kirin Professor of Marketing
Thesis Supervisor

Accepted By: ____

Ezra Zuckerman
Nanyang Technological University Professor
Chair of Ph.D. Program, MIT Sloan School of Management

Dissertation Committee Members

Erik Brynjolfsson

Schussel Family Professor of Management Science & Professor of Information Technology

MIT Sloan School of Management

John Hauser

Kirin Professor of Marketing

MIT Sloan School of Management

Stuart E. Madnick

John Norris Maguire Professor of Information Technologies & Professor of Engineering Systems

MIT Sloan School of Management

Using Web Data and Services:
Technology, Theory and Evidence

By

Xitong Li

Submitted to the MIT Sloan School of Management on May 9, 2014 in
partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Management

ABSTRACT

Many firms and individuals have been publishing data and services on the Web. It is necessary to develop advanced technology facilitating the use of Web data and services and to understand what impacts on firms and individuals would be. This thesis, composed of three essays, aims to explore (1) what technology could be developed to facilitate using Web data and services, and (2) what theoretical mechanisms are driving the impact of using Web data and services. The first essay describes an advanced technology for using Web services and the other two essays present some theoretical mechanisms and empirical evidences about how consumers are influenced by the data published on commercial webpages.

The first essay presents a classification of the data misinterpretation problems that may occur when composing Web services. After the problem scope is identified, it proposes an approach to automatic detection and reconciliation of data interpretation conflicts in Web services composition. To validate and evaluate the approach, the first essay describes a prototype and demonstrates the approach can significantly alleviate the reconciliation efforts for Web services composition.

The second essay explores how herding and social media Word of Mouth (WOM) drive product sales when commercial websites disclose the sales data in real-time on the product pages and integrate with social-networking platforms (e.g., Facebook, Twitter). Using a panel data set consisting of about 500 deals from Groupon.com, the second essay shows both herding and Facebook-mediated WOM lead to additional product sales, whereas Twitter-mediated WOM has no significant impact on sales. More importantly, it documents that herding and Facebook-mediated WOM are complements in driving sales.

Given the fact that many commercial websites integrate with social-networking platforms and the importance of social media endorsements, the third essay investigates if online review ratings would affect consumers' decisions of endorsing via Facebook and purchasing products. It builds a stylized Bayesian learning model and derives three hypotheses. The empirical findings largely support the hypotheses. In particular, the results show that a favorable valence of online reviews causes to increase consumers' social media endorsements and the estimated effect is greater when the variance in the review ratings is larger. Moreover, the findings reveal that consumers exhibit different behaviors when they consider endorsing *versus* purchasing products.

Thesis Supervisor: Stuart E. Madnick
Title: John Norris Maguire Professor of Information Technologies; Professor of Engineering Systems

Thesis Supervisor: John Hauser
Title: Kirin Professor of Marketing

Acknowledgements

This dissertation would not have been possible without the support of many, and I am immensely grateful to those who have given me intellectual, emotional and financial support during my PhD study at MIT Sloan School of Management.

To the members of my dissertation committee, Stuart Madnick, John Hauser, and Erik Brynjolfsson, I am honored and feel extremely proud and fortunate to call you my advisors. My deepest gratitude goes to Stuart Madnick. Stuart invited me to come to MIT as a visiting student in 2008 and has inspired me to dedicate my career in academia. Stuart took me as his doctoral student when I decided to apply to the PhD program of MIT Sloan School in 2009. During these years, Stuart has given me so much support from various aspects. It is difficult to imagine who I would have been without his support and mentorship. Thank Stuart for believing in me and giving me the opportunities to grow and explore. I also want to send my gratitude to John Hauser. I feel immensely fortunate to attend John's doctoral seminars in marketing and to get my research works exposed to him. I'm especially grateful for John's encouragements and inspiration. My job market research significantly benefited from his insights and guidance. John also gave me tremendous support and encouragement during the process of my job search. Thank John for believing in my work and caring me. My gratitude also goes to Erik Brynjolfsson. Every interaction with him has been an eye-opening experience and he always gives me insightful comments and suggestions. Thank Erik for supporting my work.

Hongwei (Harry) Zhu has been my coauthor and a great friend since I first came to MIT in 2008. Over the past few years, Harry has given me numerous intellectual and emotional supports. As a coauthor, Harry is a great colleague whom I have enjoyed working with. As a friend, Harry has always been generous with his time and provided me with encouragements whenever I need. I feel fortunate to have him around.

Several other faculty members at MIT also provided tremendous support and guidance. It is especially fortunate that I took Jens Hainmueller's course in political science. His course has greatly influenced my research interests and paradigm. The skills that I learned from his course are extremely useful in my current research works. I also want to extend my gratitude to Allen Moulton, Wamda Orlikowski, Nelson Repenning, Michael Siegel, Duncan Simester, Catherine Tucker, Marshall Van Alstyne, Juanjuan Zhang, Yanchong (Karen) Zheng. They have provided valuable feedback on my research works and about my job search.

I'm also extremely grateful that I have many great peers at MIT. As a senior peer, Lynn Wu has become a great scholar in our field and my role model. When we have been working together, I have learned so much from her insights, comments and suggestions. Thank Lynn for her providing me with tremendous support and encouragements. Besides Lynn, Yiwei Chen, Daria Dzyabura, You Hou, Shan Huang, Song Lin, Yiftach Nagar, Nell Putnam-Farr, Yanbo Wang, Sun Yang, Jiayin Zhang, Kexin Zheng are the ones make my PhD journey at MIT colorful and interesting. I will always remember the time we have been staying together. I feel fortunate that we have been great friends.

I would also like to thank my significant other Hailin Hu. Hailin has been giving me her love and support during the past years. She has cared my life so much and provided numerous encouragements for my living and study. She is also an excellent audience of my research works.

Last but not least, I would like to thank my parents Guosheng Li and Xiuying Li as well as my sister Huangtong Li. Thank you for always believing in me and standing behind all my endeavors.

Page intentionally left blank

Table of Contents

Page intentionally left blank

Introduction

Many firms and individuals have been publishing data and services on the Web. For example, Xignite.com has published a number of Web services (software components designed to support interoperable machine-to-machine interaction) that provide financial data. Facebook.com and Twitter.com offer a number of application programming interfaces (APIs) that can be used to retrieve data about the users' profiles and activities. Groupon.com and LivingSocial.com are constantly publishing the sales data on their deal pages. TripAdvisor.com and Yelp.com provide numerous user-generated reviews about hotels and restaurants. This thesis, composed of three essays, aims to explore (1) what technology could be developed to facilitate using Web data and services, and (2) what theoretical mechanisms drive the impact of using Web data and services. Specifically, the first essay describes an advanced technology for using Web services and the other two essays present some theoretical mechanisms (from the literature of economics, marketing and social psychology) and empirical evidences about how consumers are influenced by the data published on commercial webpages. Figure 1 illustrates the research framework and the interdependences of the three essays.

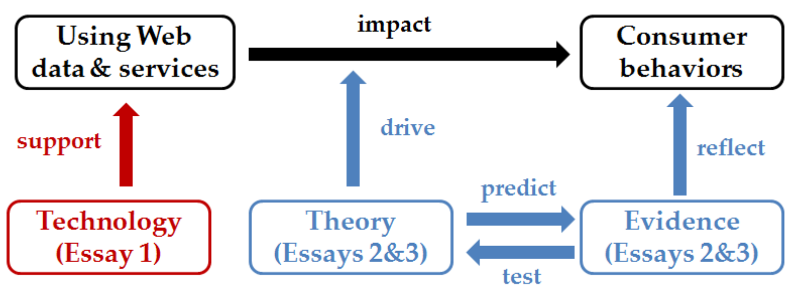

Figure 1: Research framework

Web services are accessible software components that can be invoked via open-standard Internet protocols (Yu et al. 2008). While a single Web service provides certain functionality, the real benefits of providing Web services come from reusing and composing them. Web services composition address the situation in which a business need cannot be accomplished by a single preexisting service, whereas a composite service consisting of multiple component services working together could satisfy the need. A successful Web service composition must ensure semantic interoperability so that data can be exchanged unambiguously among the involved services. Unfortunately, semantic interoperability is often hampered by data misinterpretation among independently-developed services. For example, a gallon in the U.S. (the so-called U.S. gallon) is approximately 3785 ml, while a gallon in the U.K. (the so-called Imperial gallon) is 4565 ml, almost a liter more. So when we learn that a particular car model has a fuel tank capacity of

15 gallons by querying a Web service (say from the U.K.), and learn about the gas mileage of 30 miles per gallon for the model by querying another Web service (say from the U.S.), we still need to know how to interpret the data exchanged (15 gallons) between the two services to compute the distance the car can go with a full tank of gas. The challenge of data misinterpretation grows when composing multiple services developed by independent provides which are distributed throughout the world and have disparate assumptions of data interpretation. Unfortunately, the current open standards for Web services provide limited technical support to address the various data misinterpretation problems.

Therefore, the first essay provides a comprehensive classification of the data misinterpretation problems that may occur when composing Web services. After identifying the problem scope, we develop an approach to automatic detection and reconciliation of data interpretation conflicts in Web services composition. The approach uses a lightweight ontology augmented with modifiers, contexts, and atomic conversions between the contexts. The open-standard descriptions of Web services are annotated to establish correspondences to the ontology. Given the naive Business Process Execution Language (BPEL) specification of the desired Web services composition with data interpretation conflicts, the approach can automatically detect the conflicts and produce the corresponding mediated BPEL. To validate and evaluate the proposed approach, we develop a prototype and show that the approach has desirable properties of software development methodology (e.g., adaptability, extensibility, and scalability) and can significantly alleviate the reconciliation efforts for Web services composition.

While many daily-deal sites (e.g., Groupon, LivingSocial) are constantly publishing the sales data in real-time on their deal pages, they have also integrated with major social-networking platforms (e.g., Facebook, Twitter) by placing the Facebook "Like" and Twitter buttons via the corresponding APIs. Highlighting the total number of vouchers sold in real-time allows potential buyers to observe prior others' purchasing decisions and may create an information cascade (e.g., herding) (Zhang and Liu 2012). Providing the Facebook "Like" and Twitter buttons allows shoppers to simultaneously endorse and share the deals to their social ties on Facebook/Twitter and may generate additional voucher sales. We call such mechanism as social media word-of-mouth (WOM). Given the web design of daily-deal sites, in the second essay we hypothesize that herding and social media WOM are two plausible mechanisms that affect consumers to purchase the deal vouchers. Moreover, we theorize that while herding helps update consumers' beliefs about the product quality through signaling, social media WOM can also have an advertising effect (Tucker 2012).

Since most daily-deal sites implement both herding and social media WOM in their web design, it is necessary to examine whether herding and social media WOM could interact with each other in driving sales. If they are complements, the current practice is optimal in that implementing both together would reinforce their positive effects on sales. But if they are substitutes, it is sub-optimal to implement both as one could cannibalize the effect of the other. We theorize that the advertising effect of social media WOM would complement with herding, but the signaling effect of social media WOM may substitute or complement with herding, depending on the nature of the informational signal provided by social media WOM. Overall, whether social media WOM and herding are complements or substitutes is an empirical question and will be answered in the second essay.

Specifically, the second essay explores how herding and social media WOM drive product sales when both are implemented together. Using a panel data set consisting of about 500 deals from Groupon.com, we find both herding and Facebook-mediated WOM lead to additional product sales, whereas Twitter-mediated WOM has no significant impact on sales. More importantly, we show herding and Facebook-mediated WOM are complements in driving product sales. The complementarity supports the current practice of daily-deal sites where both mechanisms are often implemented together. To further uncover the underlying mechanisms, we compare the estimated effects on two product categories: experience goods and search goods (Nelson 1974). While the values of experience goods (e.g., cleaning services, massage) are hard to ascertain before consumption, the values of search goods (e.g., shoes, clothing accessories) are relatively easier to ascertain before consumption. We find the herding effect is more salient for experience goods than for search goods, but the effect of Facebook-mediated WOM does not significantly differ between the two product categories. The comparison suggests that signaling product quality is the underlying mechanism of herding, while the effect of Facebook-mediated WOM is primarily through advertising, rather than signaling.

Given the fact that the extant literature documents the importance of consumers' social media activities (Aral et al. 2013, Malhotra et al. 2013) (for example, clicking Facebook "Like" button could generate additional voucher sales for daily-deal sites) and the current practice of many commercial websites integrating with social-networking platforms (e.g., Facebook, Twitter, Pinterest), it is deemed to explore what factors consumers would take into account in the decision-making of endorsing a product to their peers with established ties via social media.

We note that consumers' social media endorsing behaviors are distinct from their purchasing behaviors, because the motives and costs of endorsing a product to one's peers with established social ties are different from buying it for one's own consumption. The key distinction lies in that endorsing a product to one's peers, comparing to purchasing, is a social behavior and can associate with one's self-image (Berger and Schwartz 2011, Wojnicki and Godes 2008), but it incurs no financial cost. It is thus necessary to understand the similarity and difference between consumers' endorsing and purchasing behaviors.

User-generated reviews are another kind of data published on the Web (the third-party review websites). Prior research (Chevalier and Mayzlin 2006, Luca 2011) has focused on establishing the casual impact of online reviews on product sales, but it is unclear whether consumers' social media endorsing behaviors would be influenced by online reviews. Interestingly, we find that investigating the impact of online reviews on social media endorsements and product sales could allow us to uncover how similar and different consumers' social media endorsing behaviors are, compared to their purchasing behaviors. Therefore, the third essay explores if online reputation (restaurants' displayed Yelp ratings), which helps update consumers' perception of product value, is a causal factor that affects consumers' decisions of endorsing via Facebook and purchasing products (the restaurants' vouchers). We build a stylized Bayesian learning model and derive the hypotheses: (1) a higher online reputation leads to increased social media endorsements and voucher sales, but only when it is built upon a sufficient amount of review ratings; (2) these effects are greater for restaurants with more reviews; and (3) these effects are greater for restaurants with a larger variance in the review ratings. Interestingly, the third hypothesis contrasts to the predictions by some established theories (e.g., cue diagnosticity theory). We test the hypotheses using data of Groupon and LivingSocial deals and find supportive empirical evidence. In particular, We find the effect of displayed Yelp ratings on Facebook Likes is greater when the variance of ratings is larger, suggesting that perhaps consumers are risk averse in the decision-making of endorsing a product to their peers. But we don't find the evidence that the effect of displayed Yelp ratings on sales would change with the variance. The findings reveal that consumers exhibit different behaviors when they consider endorsing versus purchasing products.

This thesis opens the door to a variety of potential studies about social media and user-generated contents. Herein, we briefly discuss two of the potential studies. First, as discussed in the third essay, consumers' social media endorsing behaviors are distinct from their purchasing behaviors. More research is needed to further explore the distinction between the two kinds of consumer behaviors. Understanding the motives and costs of social media endorsements would help uncover why people are willing to endorse certain commercial products rather than the others. Moreover, studying the antecedences of one's social media endorsements would allow us to understand better the consequences on others' behaviors. Second, given the importance of user-generated reviews, it is necessary to explore how to encourage consumers to write about their experiences without biasing their opinions. Currently, a field experiment is being designed in collaboration with a local hotel and a third-party review site. We plan to investigate how online reviews solicited would differ from those written naturally by consumers without any solicitation. Potentially, solicitation would have two effects. One the one hand, solicitation may change consumers' opinions about their experiences. On the other hand, solicitation would encourage those consumers who otherwise would not have written the experiences go to write their opinions on the third-party review sites. The two effects of solicitation may change the distribution of the overall consumer-generated reviews.

References

Aral, S., C. Dellarocas, D. Godes. 2013. Social Media and Business Transformation: A Framework for Research. *Information Systems Research* **24**(1) 3-13.

Berger, J., E.M. Schwartz. 2011. What Drives Immediate and Ongoing Word of Mouth? *Journal of Marketing Research* **48**(5) 869-880.

Chevalier, J.A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345-354.

Luca, M. 2011. Reviews, reputation, and revenue: The case of Yelp. com. *Harvard Business School NOM Unit Working Paper* (12-016).

Malhotra, A., C. Kubowicz Malhotra, A. See. 2013. How to Create Brand Engagement on Facebook. *MIT Sloan Management Review* **54**(2) 18-20.

Nelson, P. 1974. Advertising as information. *The Journal of Political Economy* **82**(4) 729-754.

Tucker, C. 2012. Social advertising. *Available at SSRN 1975897*.

Wojnicki, A., D. Godes. 2008. Word-of-mouth as self-enhancement. *HBS Marketing Research Paper*.

Yu, Q., X. Liu, A. Bouguettaya, B. Medjahed. 2008. Deploying and managing Web services: issues, solutions, and directions. *The International Journal on Very Large Data Bases* **17**(3) 537-572.

Zhang, J., P. Liu. 2012. Rational Herding in Microloan Markets. *Management Science* **58**(5) 892-912.

Essay One

A Context-Based Approach to Reconciling Data Interpretation Conflicts in Web Services Composition

Abstract

We present a comprehensive classification of the data misinterpretation problems and develop an approach to automatic detection and reconciliation of data interpretation conflicts in Web services composition. The approach uses a lightweight ontology augmented with modifiers, contexts, and atomic conversions between the contexts. The WSDL descriptions of Web services are annotated to establish correspondences to the ontology. Given the naive Business Process Execution Language (BPEL) specification of the desired Web services composition with data interpretation conflicts, the approach can automatically detect the conflicts and produce the corresponding mediated BPEL. Finally, we develop a prototype to validate and evaluate the approach.

1. Introduction

Service-Oriented Computing (SOC) has become an increasingly important computing paradigm to develop and integrate distributed enterprise IT systems (Papazoglou et al. 2007). As a technology of choice for SOC, Web services, also simply called services, are accessible software components that can be invoked via open-standard Internet protocols (Yu et al. 2008). Web services composition addresses the situation in which a business need cannot be accomplished by a single pre-existing service, whereas a composite service consisting of multiple component services working together could satisfy the need. While the interface of a single (component or composite) service is described in Web Service Description Language (WSDL) (Christensen et al. 2001), the workflow logic of a composite service is usually defined in Business Process Execution Language (BPEL) (Alves et al. 2007), a standard from the Organization for the Advancement of Structured Information Standards (OASIS) for specifying the process of messages exchanged between Web services.

A successful service composition must ensure semantic interoperability so that data can be exchanged unambiguously among the involved services. Unfortunately, semantic interoperability is often hampered by data misinterpretation among independently-developed services. For example, a gallon in the U.S. (the so-called U.S. gallon) is approximately 3785 ml, while the "same" gallon in the U.K. (the so-called Imperial gallon) is 4546 ml, almost a liter more. So when we learn that a particular car model has a fuel tank capacity of 15 gallons by querying a Web service (say from the U.K.), and learn about the gas mileage of 30 miles per gallon for the model by querying another Web service (say from the U.S.), we still need to know how to interpret the exchanged data (i.e., 15 gallons) between the two services to compute the distance the car can go with a full tank of gas. Apparently, additional information is still needed to correctly utilize the exchanged data. The challenge of data misinterpretation grows when composing multiple services developed by independent providers that are distributed throughout the world and have disparate assumptions of data interpretation. The basic Web services standards (e.g., WSDL, BPEL) generally ignore data semantics, rendering semantic interoperability far from reality. Several initiatives, e.g., OWL-S (Martin et al. 2007), WSMF/WSMO (Lausen et al. 2005) and METEOR-S (Patil et al. 2004), have proposed languages and frameworks to explicitly add semantics into service descriptions. Despite the foundations provided by these efforts, effective methods still need to be developed for reconciling data misinterpretation in Web services composition.

In this paper, we first present several real-world examples[1] of Web services and service composition with data misinterpretation problems. Those examples clearly demonstrate in reality how data misinterpretation affects the use of Web services and hampers their composition. Then, we develop a comprehensive classification of the various data misinterpretation problems that we have observed in the practice of Web services composition. The classification helps identify the scope of the problem domain. To address the challenging problems, we describe our approach to automatic detection and reconciliation of data interpretation conflicts in Web services composition. The approach is inspired by the Context Interchange (COIN) strategy for semantic interoperability among multiple data sources (Bressan et al. 2000; Goh et al. 1999) and the preliminary works of applying the strategy (Li et al. 2009a; Li et al. 2009b; Mrissa et al. 2007) to Web services composition. The approach uses a lightweight ontology to define a common vocabulary capturing only generic concepts shared by the involved services. The lightweight ontology also defines multiple contexts capturing different specializations (which are actually used by the involved services) of the generic concepts. Atomic conversions reconciling certain aspects of the differences need to be provided. Further, the WSDL descriptions of the involved services need to be annotated to establish correspondences between the data elements of WSDL descriptions and the concepts of the ontology. In this paper, we assume the service composition is specified using BPEL - in fact, our solution can be applied with any other composition specification languages. We call the BPEL composition ignoring data misinterpretation the naive BPEL. With the above descriptions in place, the reconciliation approach can automatically detect data interpretation conflicts in the *naive BPEL* and produce the corresponding *mediated BPEL* by incorporating appropriate conversions into the composition.

[1] Some of them are simplified from real-world Web services.

The mediated BPEL composition, now without any data interpretation conflict, is the output of the reconciliation approach and can be successfully deployed and executed.

We make three contributions that, to the best of our knowledge, have not appeared elsewhere:

First, we provide a set of new algorithms to automatically analyze data flows of service composition processes and reconcile data misinterpretation problems in the composition processes. The approach can significantly alleviate the reconciliation efforts and accelerate the development of Web services composition. Although the approach is demonstrated with BPEL composition only, it is a generalizable approach and can be easily adapted to analyze the data flow of a process specified in many other process modeling languages, such as process algebra, UML Activity Diagram and the Business Process Modeling Notation (BPMN). Thus, the approach can address semantic reconciliation in a broad context of Business Process Integration (BPI) (Becker et al. 2003) and workflow management (van der Aalst and Kumar 2003).

Second, we extend the W3C standard SAWSDL so that the extended SAWSDL can be used to annotate context information in WSDL descriptions. Specifically, we design two methods for context annotation to alleviate the complexity of handling the evolving data semantics of Web services. The extension for context annotation complies with SAWSDL so that the annotation task can be performed using any existing SAWSDL-aware tools, e.g., Radiant (Verma and Sheth 2007). Thus, this mechanism facilitates the annotation task and makes our approach practical, accessible and flexible.

Third, as part of this work, we develop and describe a working prototype – the Context Mediation Tool (CMT). By using the working prototype in a number of examples, we demonstrate the feasibility and applicability of our approach.

The reconciliation approach, as qualitatively and quantitatively evaluated in this paper, has the desirable properties of software development methodology (e.g., adaptability, extensibility and scalability) and can significantly alleviate the reconciliation efforts for Web services composition. Thus, the approach facilitates the application of SOC to develop Web-based information systems. This paper contributes to the literature on Service-Oriented Computing (Papazoglou et al. 2007), Business Process Integration (BPI) (Becker et al. 2003) and workflow management (van der Aalst and Kumar 2003). The rest of the paper is organized as follows. Section 2 describes the challenges of data misinterpretation problems when using and composing Web services. Section 3 and Section 4 present the reconciliation approach and the prototype. Section 5 presents the results of the validation and evaluation. Section 6 discusses the related work. Finally, Section 7 concludes the paper.

2. Challenges of Data Misinterpretation Problems

2.1 Motivating Examples of Web Services

2.1.1. Example 1: A Problematic Web Service. *Xignite*, Inc., an established U.S. Web services provider, has published a service named *XigniteEdgar* which consumes the stock ticker symbol of a company and returns its total assets. When requested using “ITWO” for i2 Technology, *XigniteEdgar* returns the data as shown in Figure 1. The returned total assets of i2 Technology is associated with the date “05/07/2009”. But should the users interpret the date as May 7th, 2009 or July 5th, 2009? How should the total assets of “313776” be interpreted? When invoked with “MSFT” for Microsoft, *XigniteEdgar* returns “68853” as Microsoft’s total assets. Is it possible that i2 Technology’s total assets are more than four times of Microsoft? Manual investigation shows the numeric figure for i2 Technology is in thousands, whereas that for Microsoft is in millions. If these assumptions of data interpretation were not explicitly clarified, users may incorrectly use *XigniteEdgar*, perhaps causing financial losses.

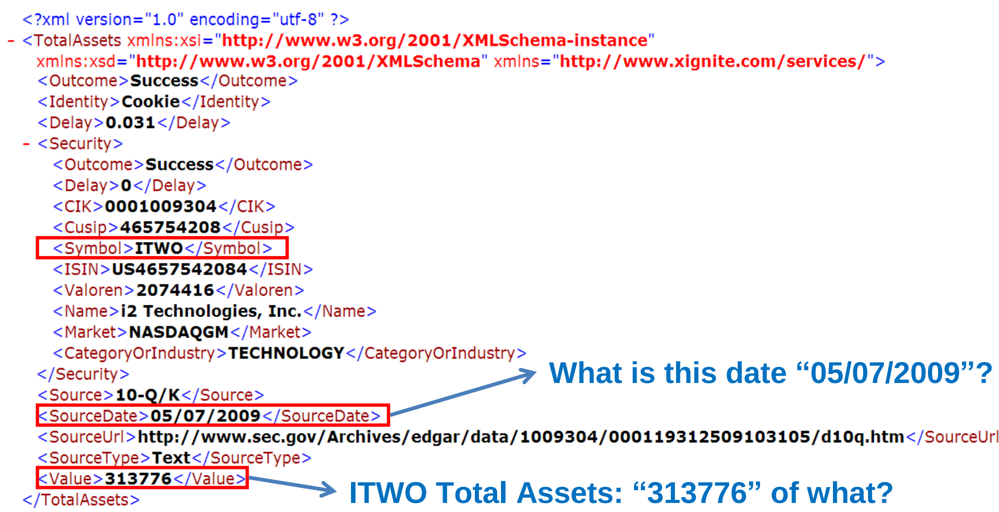

Figure 1: A problematic Web service with ambiguous data interpretation

2.1.2. Example 2: A Simple Composition of Two Component Services. Let’s consider a simple composition scenario with only two services in which a Chinese developer wants to develop a composite service *ConfHotelDeals*. Its function is to consume an international conference code and return the hotel expenses in the city where the conference is held. With the purpose of exploiting reuse, the developer decides to implement *ConfHotelDeals* by composing two existing services: *ConfInfo* and *HotwireDeals*[2]. Given a conference code, the operation *queryConfInfo* of *ConfInfo* provides basic information of the conference, including start and end dates and the city where the conference is held. The operation

[2] *HotwireDeals* originates from Hotwire.com, available at http://developer.hotwire.com/docs/Hotel_Deals_API.

queryDeals of *HotwireDeals* returns the room charges of the deals based on the city name and start/end dates. The composition process is illustrated in Figure 2. Unfortunately, these services have different assumptions about data interpretation. *ConfHotelDeals* is intended to return the monetary expenses in Chinese yuan ("RMB") and the hotel expense includes the value-added taxes. *ConfInfo* provides the dates in "dd-mm-yyyy". *HotwireDeals* assumes dates are in "mm/dd/yyyy" and returns the hotel deals in US dollars ("USD") without value-added taxes. If the data misinterpretation problems were not properly resolved, conflicts would happen in the composition process (as noted in Figure 2 by little "explosions") and the composite service *ConfHotelDeals* would not work correctly.

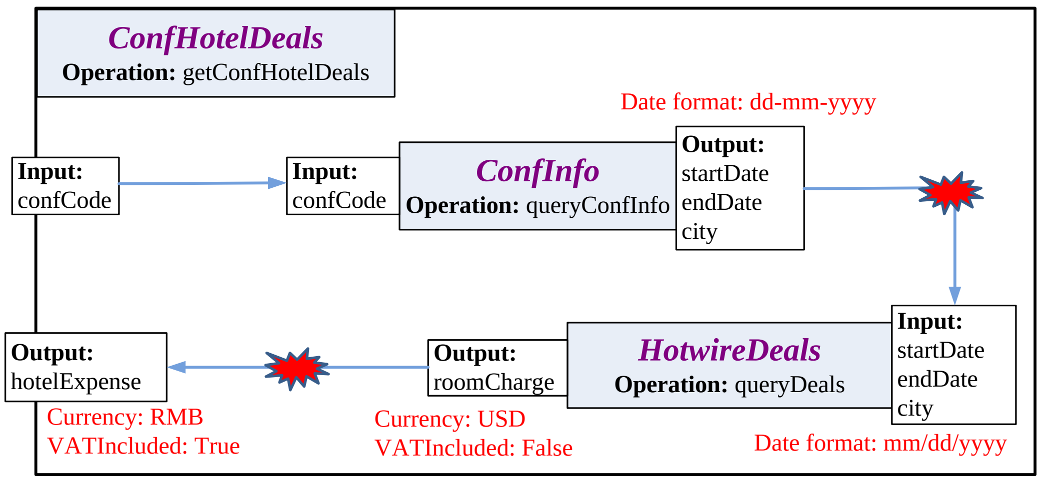

Figure 2: Example 2: Simple composition of two component services

2.1.3. Example 3: Composition Example of Multiple Services. Now let's consider a somewhat complicated scenario that a U.K. developer wants to develop a new Web service, *OpeningPriceMarketCap* (denoted as *CS* for *C*omposite *S*ervice), to obtain the opening stock price and market capitalization of a U.S. company on its <u>first</u> trading day. *CS* is intended for a U.K. analyst to monitor the U.S. stock market. The developer decides to implement the service by composing three existing services: *StockIPOWS*, *OpeningPriceWS* and *DailyMarketCap*, denoted as *S1*, *S2* and *S3* respectively. *S1* has the operation *getDateofIPO* that provides the IPO date of a company traded in the U.S. by using the company's ticker symbol. The operation *getOpeningPrice* of *S2* provides the opening stock price of a company on its first trading day. The operation *getDailyMarketCap* of *S3* provides the daily market capitalization of a company on a given date.

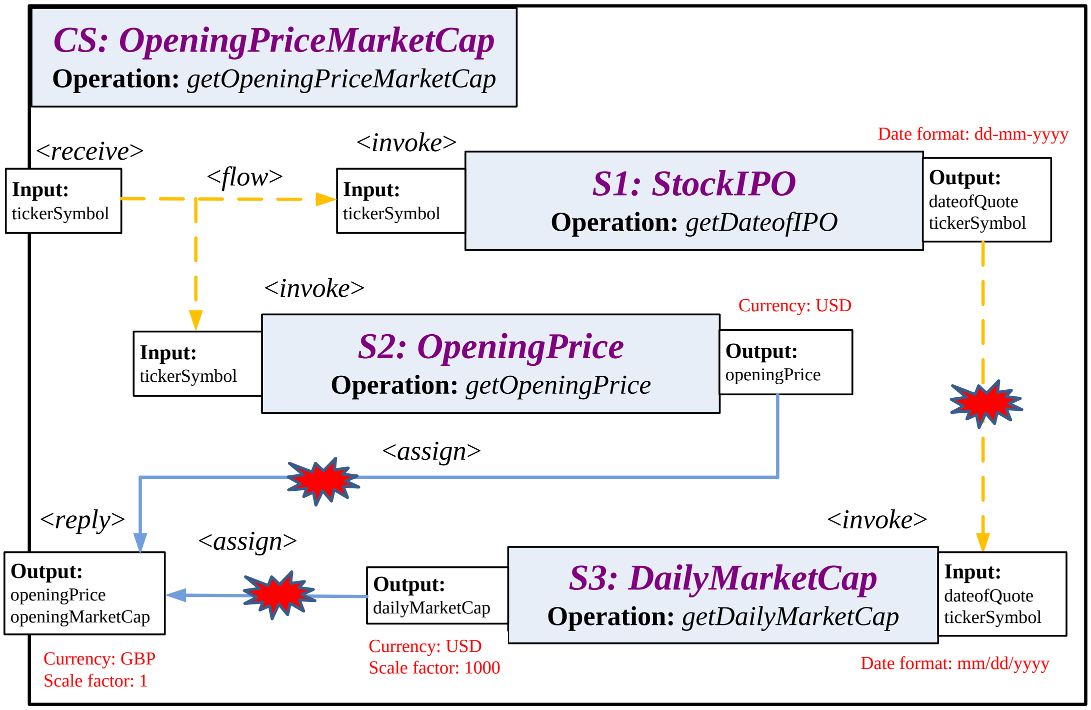

Figure 3: Example 3: Composition of multiple services

In principle, *CS* can be accomplished by a composition of *S1*, *S2* and *S3*. Specifically, the input *tickerSymbol* of *CS* needs to be transferred to both *S1* and *S2*. The output *openingPrice* of *CS* is obtained from the output *openingPrice* of *S2*. The output *openingMarketCap* of *CS* can be achieved by feeding the output of *S1* to the input of *S3* and delivering the output of *S3* to *CS*. According to this plan, the developer defines the workflow logic of the composition using a typical BPEL tool, such as ActiveVOS BPEL Designer[3]. The BPEL composition is graphically illustrated in Figure 3, where BPEL activities (e.g., <receive>, <invoke>) are enclosed in angle brackets. Since these four services are developed by independent providers, they have different assumptions about data interpretation in terms of data format, currency, and scale factors, as summarized in Table 1.

Table 1. Different Assumptions of Data Interpretation

Service	Date format	Currency	Scale factor
CS	-	GBP	1
S1	dd-mm-yyyy	-	-
S2	-	USD	1
S3	mm/dd/yyyy	USD	1000

[3] http://www.activevos.com/

Note that usually these assumptions are not explicitly represented in WSDL descriptions. As a result, existing BPEL tools (e.g., ActiveVOS BPEL Designer) cannot detect these conflicting assumptions and fail to alert data misinterpretation problems in the composition because the interpretation conflicts exist at the data instance level. If not reconciled, the data interpretation conflicts would result in severe errors and failures during the execution of the composition. This composition example (i.e., Example 3) will be used as the "walk-through" example in the rest of the paper.

2.2 Classification of Data Misinterpretation Problems

We classify data misinterpretation problems into representational, conceptual and temporal categories, as summarized in Table 2. The purpose of the classification is to help readers understand the problem scope of our solution and meanwhile draw the boundary of our study. Note that there exist a number of classification frameworks in the literature (Nagarajan et al., 2006; Sheth et al., 2005; Halevy, 2005). Those existing classifications tend to cover a broader range of semantic heterogeneity issues, some of which can be addressed by our approach (e.g., scale factors, currency), while others are not the focus of this paper, such as structural/schematic differences. The classification presented here exclusively focuses on data interpretation conflicts that may occur in Web services.

2.2.1. Representational. Different organizations may use different representations for a certain concept, which can result in representational misinterpretation problems. Five subcategories can be further identified at this level: format, encoding, unit of measure, scale factor, and precision. Format differences occur because there often exist multiple format standards, such as for representing date, time, geographic coordinates, and even numbers (e.g., "1,234.56" in USA would be represented as "1.234,56" in Europe). Encoding differences may be the most frequent cause of representational misinterpretation, because there are often multiple coding standards. For example, the frequently used coding standards for countries include the FIPS 2-character alpha codes, the ISO3166 2-character alpha codes, 3-character alpha codes, and 3-digit numeric codes. Also, IATA and ICAO are two standards for airport codes. Data misinterpretation problem can occur in the presence of different encoding standards (e.g., country code "BG" can stand for Bulgaria or Bangladesh, depending on whether the standard is ISO or FIPS). Besides the format and encoding differences, numeric figures are usually represented using different units of measure, scale factors, and precisions. For example, financial services use different currencies to report the data to consumers who prefer to use their local currencies. Scientific services may use different units of measure to record the data (e.g., meter or feet).

Table 2. Classification of Data Misinterpretation Problems

Categories		Explanations / Examples
Representational	Format	Different format standards for date, time, geographic coordinate, etc. Example: “05/07/2009” vs. “2009-05-07”
	Encoding	Different codes for country, airport, ticker symbol, etc. Example: Male/Female vs. M/F vs. H/D[4] vs. 0/1
	Unit of measure	Different units of currency, length, weight, etc. Example: 10 “USD” vs. 10 “EUR”
	Scale factor	Different scale factors of numeric figures Example: 10 “Billion”[5] vs. 10 “Million”
	Precision	Different precisions of numeric figures Example: “5.8126” vs. “5.81”
Conceptual	Subtle differences in conceptual extension	Different interpretations about whether or not a specific entity should be included Example: does the reported retail “price” include value-added taxes or not?
Temporal	Representational and conceptual data interpretation may change over time	Prices listed in Turkey are implicitly in Turkish liras (TRL) before 2005 but in Turkish New Lira (TRY) after January 1, 2005.

2.2.2. Conceptual. The same term and representation is often used to refer to similar but slightly different data concepts. This category of misinterpretation usually occurs when the extension of the concept has different assumptions of the interpretation, such as whether or not a specific entity is included by the concept. For example, a retail price reported by European services usually includes the value-added taxes, while retail prices reported by US services, especially for purchases to be done in a store, usually do not include the value-added taxes.[6] An even more challenging problem in this category is referred to as “Corporate Householding” (Madnick et al. 2003) which refers to misinterpretation of corporate household data. For example, the answer to “What were the total sales of IBM” varies depending on whether the sales of majority owned subsidiaries of IBM should be included or not. The answers can be very different due to different reporting rules adopted in different countries or for different purposes. Besides the entity aggregation issue, the conceptual extension of the inter-entity relationship may also have different interpretations. For instance, in order to answer the question “How much did MIT purchase from IBM in the last fiscal year?”, we need to clarify whether the purchasing relationship between MIT and IBM should be interpreted as direct purchasing (i.e., purchased directly from IBM) or indirect purchasing through other channels (e.g., third-party brokers, distributors, retailers). In some cases,

[4] In France.
[5] Of course, these categories can be nested – for example, there can be different meanings of scale factor, such as “Billion” means one thousand million in USA but it used to mean one million million in the UK.
[6] Usually called “sales taxes” in the USA

only the direct purchasing from IBM to MIT are considered, whereas in other cases indirect purchasing through other channels also needs to be included (Madnick and Zhu 2006).

2.2.3. Temporal. Most of the above-mentioned possibilities of data interpretation may change over time (Zhu and Madnick 2009). For example, a Turkish auction service may have listed prices in millions of Turkish liras (TRL),[7] but after the Turkish New Lira (TRY) was introduced on January 1, 2005, it may start to list prices in unit of Turkish New Lira. Also, an accounting service may or may not aggregate the earnings of Merrill Lynch into that of Bank of America which acquired the former in September 2008. Considering the highly dynamic and distributed environment of Web services, these data misinterpretation problems resulting from the temporal evolvement would become very challenging. Due to length limit, we will not address the temporal issues in this paper, but our approach can be extended to resolve them.

2.3 Deficiency of Existing Approaches

To address the abovementioned problems, we must identify the data interpretation conflicts that may occur in naive BPEL composition and rewrite it to reconcile the identified conflicts. The existing approaches usually perform the identification and reconciliation of interpretation conflicts in a manual way. As depicted in the upper half of Figure 4, after the naive BPEL is produced, a manual inspection of potential conflicts is conducted. Once an interpretation conflict is detected, the naive BPEL is modified by inserting an ad-hoc conversion to transform the output of the upstream service to the needed input of the downstream one. These steps (as indicated as "Identify conflicts" and "Rewrite") are continued iteratively until a valid BPEL is produced. The ad-hoc, "brute-force" approaches tend to produce "spaghetti" code that is difficult to debug and maintain. In summary, the brute-force approaches suffer from the following deficiencies: 1) It is error-prone to manually inspect the naive BPEL, especially when the composition involves a large number of data elements as well as Web services and has complicated workflow logic. Also, it is error-prone to manually define customized conversion code and insert it to the composition; 2) It is difficult to reuse the conversion code, as it usually defined and inserted in the composition in an ad-hoc way; and 3) Every time an involved service is changed (or removed) or a new service is added, the *Identifying conflicts* and *Rewrite* steps need to be manually performed again and new custom conversions may need to be inserted in the composition. As a result, the brute-force approaches potentially make the number of custom conversions very large and difficult to maintain over time.

[7] About one million TRL equaled one US dollar.

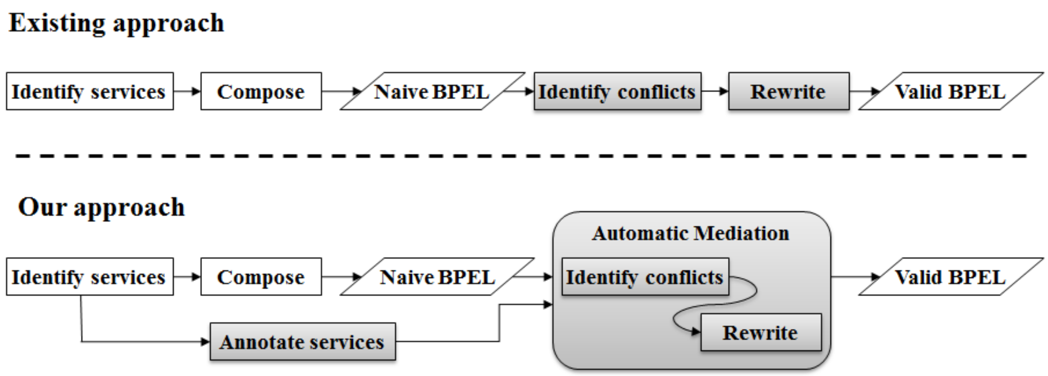

Figure 4: Comparison of existing approach and our proposed approach.

The situation could become even worse when the number of services involved in the composition is large and the involved services are highly dynamic. For example, the recent SOA implementation of a Texas health and human resource system consists of over a hundred Web services and more than 20 composite services.[8] According to a recent Application Integration Survey, data integration accounts for about 40% of software development costs.[9] Another survey conducted in 2002 reveals that approximately 70% of the integration costs were spent on identifying interpretation differences and developing custom code to reconcile these differences (Seligman et al. 2002). Therefore, it is important to develop a systematic and disciplined approach to addressing the various data misinterpretation problems for Web services composition.

We have developed an improved approach to rectify these deficiencies. Our approach automates the "*Identify conflicts*" and "*Rewrite*" steps as an intelligent mediation step (see the lower half of Figure 4). By using the proposed approach, developers do not need to read the naive BPEL to identify the conflicts or to decide where the conversions need to be inserted. We provide a tool that fully automates the mediation step and produces the valid BPEL.

Note that our approach requires the services in the composition be annotated to explicitly capture the assumptions that affect the interpretations of data. Although semantic annotation is a new step, it allows for the separation of declarative semantic descriptions from the programming code. It also enables automatic identification and reconciliation of semantic conflicts. As we will show in Section 5.2.2, this separation offers tremendous benefits to our approach.

---

[8] Source from the email communication between the authors and SourcePulse.com, a software services firm.
[9] http://www.slideshare.net/mlbrodie/powerlimits-of-relational-technology

3. Context-Based Approach

In this section, we describe our context-based approach to reconciling data interpretation conflicts in Web services composition. The approach consists of methods for representing semantic assumptions and mediation algorithms for identifying conflicts and rewriting the BPEL to reconcile the identified conflicts. The lightweight ontology (Zhu and Madnick 2007) is used to facilitate semantic annotation.

3.1 Representation of Ontology and Contexts

3.1.1. Lightweight Ontology. Ontology is a collection of concepts and the relationships between these concepts. Ontologies are often used for Web query processing (Storey et al. 2008), Web services composition (Mrissa et al. 2007), and data reliability assessment (Krishnan et al. 2005). In practice, there are various types of ontologies ranging from lightweight, rather informal, to heavyweight, more formal ones (Wache et al. 2001). Lightweight ontologies are simple and easy to create and maintain since they only include the high-level concepts. On the other hand, they do not directly provide all the depth and details of a typical formal ontology. In contrast, formal ontologies are often relatively complex and difficult to create (Zhu and Madnick 2007).

To combine the strengths and avoid weaknesses of these ontology approaches, we adopt an augmented lightweight ontology approach that allows us to automatically derive a fully specified ontology from concisely described high-level concepts and contexts. By “lightweight”, we mean the ontology only requires generic concepts used by the involved services and the hierarchical relationships between the concepts. The different assumptions of the services for interpreting the generic concepts are represented as contexts using the vocabulary and structure offered by the ontology.

Figure 5 presents a graphical representation of the lightweight ontology for Example 3 (see Section 2.1.3). Concepts are depicted by round rectangles and *basic* is the special concept from which all other concepts inherit. Like traditional ontologies, the lightweight ontology has two relationships: *is_a* and *attribute*. For instance, concept *openingPrice is a* type of *stockMoneyValue*. An attribute is a binary relationship between a pair of concepts. For example, attribute *dateOf* indicates that the *date* concept is the “date of” attribute of concept *stockMoneyValue*. In practice, it is frequently straightforward to identify generic concepts among multiple independent services. For example, *S3* has an output *dailyMarketCap* and *CS* has an output *openingMarketCap*. Both of them correspond to a generic concept *marketCapital*. However, *S3* provides the data instances of *dailyMarketCap* using currency “USD” and scale factor “1000”, while *CS* interprets and furnishes the data instances of *openingMarketCap* using currency “GBP” and scale factor “1”. To accommodate the different data interpretations, the construct modifier is introduced to allow multiple variations (i.e., specializations) to be associated with different services. In other words, modifier is used to capture additional information that affects the interpretations of the generic concepts. A generic concept can have multiple modifiers, each of which indicates an orthogonal dimension of the variations. Also, a modifier can be inherited by a sub-concept from its ancestor concepts.

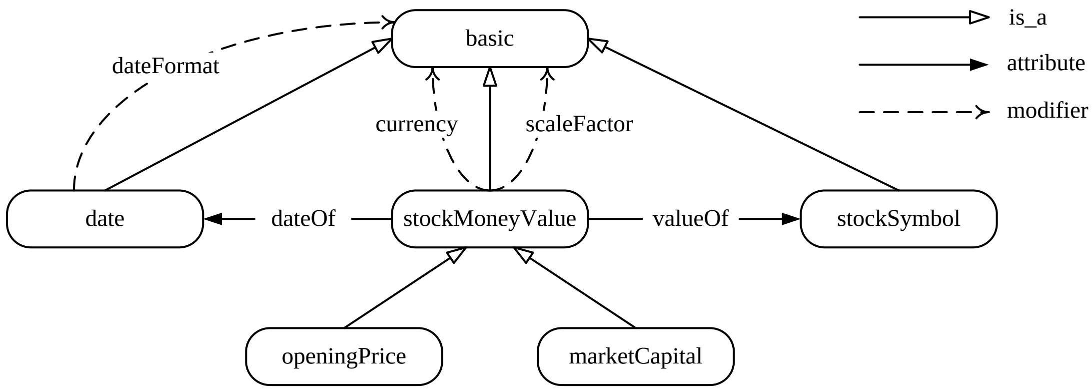

Figure 5: Lightweight ontology shared by involved services of the composition.

Modifiers are depicted by dashed arrows in Figure 5. For example, concept *stockMoneyValue* has two modifiers, *currency* and *scaleFactor*, which indicates that its data instances need to be interpreted according to two dimensions: money currency and scale factor, respectively. Also, concept *date* has modifier *dateFormat* that indicates its data instances can be interpreted by different date formats. The actual interpretation of a generic concept depends on modifier values. For instance, *CS* interprets concept *openingMarketCap* using currency “GBP”. Thus, the value of modifier *currency* is “GBP” in case of CS. According to Table 1, the modifier value of *currency* is “USD” in case of *S2* and *S3*. That means that different services may be associated with different values assigned to the modifiers. In our work, the different value assignments to a collection of modifiers are referred to as different *contexts*, and in a certain context each modifier is assigned by a specific modifier value. Specifically, a context is conceptually a set of assignments of all the modifiers of the ontology and can be described by a set of <modifier, value> pairs. Further, each service involved in the composition may be associated with a context which corresponds to its assumption of data interpretation. For example, the different assumptions in Table 1 are described using four contexts associated with the four services involved in the composition, as shown in Table 3. As a result, interpretation differences among these services can be treated as context differences.

Table 3. Context Definition of Involved Services in the Composition

Service	Context
CS	*ctxt0* = [<*dateFormat*, NULL>, <*currency*, GBP>, <*scaleFactor*, 1>]
S1	*ctxt1* = [<*dateFormat*, dd-mm-yyyy>, <*currency*, NULL>, <*scaleFactor*, NULL>]
S2	*ctxt2* = [<*dateFormat*, NULL>, <*currency*, USD>, <*scaleFactor*, 1>]
S3	*ctxt3* = [<*dateFormat*, mm/dd/yyyy>, <*currency*, USD>, <*scaleFactor*, 1000>]

3.1.2. Semantic and Context Annotation. Web services are usually described using the WSDL specification at a syntactic level, rather than a semantic level. To facilitate semantic interoperability, semantic annotation is widely used to establish correspondences between the data elements of WSDL descriptions and the concepts of an ontological model (Patil et al. 2004; Sivashanmugam et al. 2003). The annotations are recommended to be done using the W3C standard, Semantic Annotation for WSDL and XML Schema (SAWSDL) (Farrell and Lausen 2007). SAWSDL allows any language for expressing an ontological model and enables developers to annotate the syntactic WSDL descriptions with pointers to the concepts (identified via URIs) of the ontological model (Kopecký et al. 2007; Verma and Sheth 2007). Thus, SAWSDL is an appropriate industrial standard for us to establish the correspondence between the syntactic WSDL descriptions and the lightweight ontology.

SAWSDL provides an attribute *modelReference* for specifying the correspondence between WSDL components (e.g., data/element types, input and output messages) and the concepts of an ontology. However, SAWSDL *per se* does not provide any mechanism for context annotation that is required for resolving data misinterpretation problems in service composition. Thus, we extend SAWSDL with two annotation methods that use the modelReference attribute: (1) Global context annotation: we allow the <wsdl:definitions> element of the WSDL specification to have the modelReference attribute and use its value to indicate that all data elements of a WSDL description subscribe to a certain context identified via the URI value; (2) Local context annotation: for any data element, in addition to the URI value indicating the corresponding ontological concept, we allow the modelReference attribute to have an additional URI value to indicate the context of the data element. Global context annotation affects the entire WSDL description and allows the developers to succinctly declare the context for all elements of the WSDL description. Local context annotation provides a mechanism for certain elements to have their contexts different from the globally declared context. In case a small number of elements in a WSDL description have contexts different from that of the other elements, this *overriding* capability can be useful to simplify the annotation task.

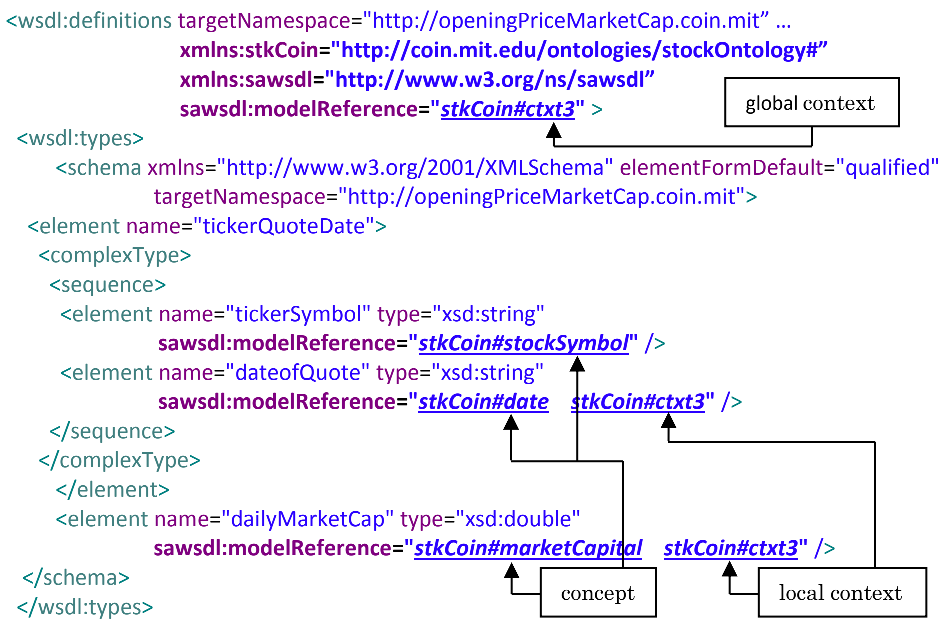

Figure 6: Excerpt of annotated WSDL description of S3
using global and local context annotations

Figure 6 shows the annotated part of *S3*'s WSDL description in which the annotations are highlighted in bold. Each leaf data element of *S3* has the modelReference attribute to point to its corresponding concept in the ontology. For example, the elements *tickerSymbol* and *dateofQuote* correspond to the concepts *stockSymbol* and *date*, respectively. Since *S3* use context *ctxt3* (see Table 3), the modelReference attribute of the element <wsdl:definitions> has the value "stkCoin#ctxt3" which is the URI of context *ctxt3* defined in the ontology. The modelReference attribute of a data element can have one value, or two values separated by a whitespace.[10] In case of only one value, it is the URI of the concept to which the data element corresponds. In case of two values, the former value is the URI of the concept and the latter is the URI of the context in which the data element is interpreted. It is worth noting that both global and local context annotations comply with the SAWSDL standard. Both the global and local context annotations are used in Figure 6. Although the local annotation does not actually override the global context, we include it for illustration purposes.

If business needs were to change over time and we later needed to shift the date format of *S3* from "mm/dd/yyyy" to "dd-mm-yyyy", the only thing we need to do is to update the context of the *dateofQuote* element of *S3* to context *ctxt1* (see Table 3) by means of the local context annotation. Then, our approach can automatically determine and reconcile possible interpretation differences resulting from

[10] SAWSDL allows the modelReference attribute to have multiple values separated by whitespaces.

the date format change. As a result, the global and local context annotations promote the flexibility of our solution to handle the evolving semantics of services.

3.1.3. Conversions between Different Contexts. Context differences, once detected, need to be reconciled using conversion programs to convert the exchanged data from the source value *vs* to the target value *vt*. In our work, a conversion is defined for each modifier between two different modifier values. Below is a general representation of the conversions, where *C* is the generic concept having a modifier *m*, *mvs* and *mvt* are two different values of *m* in the source context *ctxt_s* and the target context *ctxt_t*, respectively. In fact, *mvs*, *mvt* can be derived by querying the context definition according to *ctxt_s*, *ctxt_t* (see Table 3).

cvt(*C*, *m*, *ctxt_s*, *ctxt_t*, *mvs*, *mvt*, *vs*, *vt*)

The conversions defined for individual modifiers are called *atomic conversions*. At least one atomic conversion is specified for each modifier to reconcile the difference indicated by different modifier values. Since there exist three modifiers in the example ontology (see Figure 5 and Table 3), we specify three atomic conversions: $cvt_{\text{dateFormat}}$, cvt_{currency} and $cvt_{\text{scaleFactor}}$.

Our solution is agnostic about the actual implementation of the atomic conversions. In practice, depending on its complexity, an atomic conversion can be implemented using an XPath function[11] or an external (e.g., third-party) service. For example, the atomic conversion $cvt_{\text{dateFormat}}$ for converting the date format from "dd-mm-yyyy" to "mm/dd/yyyy" can be implemented using the following XPath function:

$cvt_{\text{dateFormat}}$: *Vt* = concat(substring-before(substring-after(*Vs*,"-"),"-"),"/",

substring-before(*Vs*,"-"),"/",substring-after(substring-after(*Vs*,"-"),"-"))

Also, the atomic conversion $cvt_{\text{scaleFactor}}$, which converts a number value from the scale factor *mvs* to *mvt*, can be implemented using the following XPath function:[12]

$cvt_{\text{scaleFactor}}$: *Vt* = *Vs* * *mvs* div *mvt*

In complex cases, the conversions may have to be implemented by invoking external (e.g., third-party) services, such as by using Web wrapper services (Madnick et al. 2000). For example, it is needed to invoke an external currency exchange service *CurrencyConverter*[13] (denoted as *S4* for short) which consumes the source and target currencies *mvs*, *mvt* and a money value *vs* and converts to another money value *vt*. Thus, *S4* can be used to implement the atomic conversion cvt_{currency}.

---

[11] The BPEL specification and most BPEL engines (e.g., ActiveBPEL) support XPath 1.0.

[12] Note that this is a general purpose conversion function that works for any values of *mvs* and *mvt*.

[13] *CurrencyConverter* originates from http://www.ac-markets.com/forex-resources/currency-converter.aspx External services for conversions may also need to be annotated with concepts and contexts.

It is worth noting that $cvt_{\text{scaleFactor}}$ and cvt_{currency} are defined as parameterized conversions: the source and target modifier values *mvs*, *mvt* are used as parameters of the conversions. A parameterized conversion can be applied to handle any pair of different modifier values *mvs* and *mvt* (i.e., a dimension of the context differences) and thus is not limited to a specific one. For example, cvt_{currency} can be used to convert money value between any pair of currencies. Using parameterized conversions can largely reduce the number of predefined atomic conversions and significantly enhance the scalability of our reconciliation solution.

3.2 Reconciliation Algorithms

In Web services composition, context conflicts can occur when a piece of data from the source service in one context is transferred to, and consumed by, the target service in another context. Figure 7 shows the typical scenario where a context conflict occurs in the composition. In Figure 7, there exists a data transfer where the data *data_s* from service *WS_s* is transferred to service *WS_t* and consumed as data *data_t*. Using context annotation, both *data_s* and *data_t* are instances of concept *C* which has a modifier *m*. Also, *WS_s* and *WS_t* are annotated with two different contexts *ctxt_s*, *ctxt_t*, respectively. As a result, according to the context definition of the ontology, *data_s* and *data_t* are interpreted differently by *WS_s* and *WS_t* if the modifier value of *m* in *ctxt_s* (i.e., *mvs*) is different from the value *mvt* of *m* in *ctxt_t*. In such a case, a context conflict occurs within the data transfer. In the following sections, we present three successive algorithms that automate the identification and reconciliation of context conflicts in the composition process. Example 3 will be used to demonstrate the algorithms.

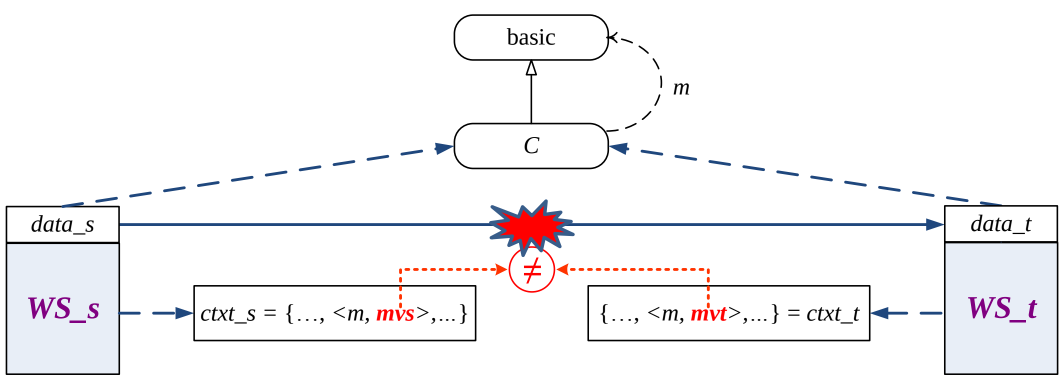

Figure 7: Scenario of context conflict in Web services composition.

3.2.1. Identifying Data Transfers. Recall that the BPEL composition that ignores context conflicts is called the naive BPEL. Since context conflicts occur within data transfers, it is needed to analyze the data flow of the naive BPEL and identify all the data transfers. Each data transfer can be represented using the following form, where *ws_s* and *ws_t* are the source and target services, *data_s* and *data_t* are the data elements involved in the data transfer, and *type* indicates if the data transfer is explicit or implicit.

dataTrans(*type*, *data_s*, *ws_s*, *data_t*, *ws_t*)

Each explicit data transfer involves two variables and can be easily identified according to the <assign> activity which is used to copy the data from the source variable to the target variable. As shown in Figure 3, there are two <assign> activities in the composition process of Example 3: they are to transfer the data *dailyMarketCap* and *openingPrice*, respectively. Thus, two explicit data transfers are identified.

Each implicit data transfer involves one variable shared by two activities interacting with participant services having potentially different contexts. The BPEL specification provides four types of interaction activities: <receive>, <reply>, <invoke>, and <onMessage> contained in <pick>. For an output variable, its source interaction activity may be <receive>, <onMessage> or <invoke>. For an input variable, its target interaction may be <reply> or <invoke>. By examining each variable in the composition, all implicit data transfers in the BPEL composition can be identified.

Algorithm 1. Identifying Explicit and Implicit Data Transfers

```
Input: BPEL process proc.
Output: The set of explicit data transfers EDT = {edt},
        the set of implicit data transfers IDT = {idt}.
set EDT = ∅, IDT = ∅;
for each <assign> activity asn in proc
  var_s ← getSourceVariable(asn), var_t ← getTargetVariable(asn)
  act_s ← getSourceInteractionActivity(proc, asn),
  act_t ← getTargetInteractionActivity(proc, asn)
  edt ← getDataTransfer(var_s, var_t, act_s, act_t)
  EDT ← EDT ∪ {edt}
for each variable var in proc
  L_var ← getInteractionActivitySeries(proc, var)
    for each source activity act_s1 in L_var
      act_s2 ← getNextSourceActivity(L_var, act_s1),
      T_var ← getTargetActivitySeries(L_var, act_s1, act_s2)
      for each target activity act_t in T_var
        idt ← getDataTransfer(var, act_s1, act_t)
        IDT ← IDT ∪ {idt}
return EDT, IDT;
```

Algorithm 1 is developed to identify explicit and implicit data transfers. Using Algorithm 1, three implicit and two explicit data transfers are identified in the composition process of Example 3, as shown in Table 4. Instead of explicitly using the <assign> activity, the output of *S1* is directly transferred and consumed as the input of *S3* through variable *tickerQuoteDate*. An implicit data transfer is thus identified, where the source and target interaction activities are the invocation of *S1*, *S3*, respectively. In Figure 1-3, the composition process involves <receive>, <reply> and <invoke>; it does not involve <onMessage>.

Table 4. Data Transfers in the Composition Process of Example 3

dt1	*dataTrans* (*implicit*, *tickerSymbol*, *CS*, *tickerSymbol*, *S1*)
dt2	*dataTrans* (*implicit*, *tickerSymbol*, *CS*, *tickerSymbol*, *S2*)
dt3	*dataTrans* (*implicit*, *tickerQuoteDate*, *S1*, *tickerQuoteDate*, *S3*)
dt4	*dataTrans* (*explicit*, *openingPrice*, *S2*, *openingPrice*, *CS*)
dt5	*dataTrans* (*explicit*, *marketCap*, *S3*, *openingMarketCap*, *CS*)

3.2.2. Detecting Context Conflicts. When a data transfer is identified, the annotated WSDL descriptions of its source and target services (denoted as *ws_s* and *ws_t*, respectively) can be derived through <partnerLinkType> of the BPEL composition. According to the context annotation, the concept *C* corresponding to the transferred data is obtained. Also, if the source data *data_s* and the target data *data_t* are annotated with contexts, their contexts are denoted as *ctxt_s*, *ctxt_t*, respectively. In order to determine possible context conflicts, all modifiers of concept *C* need to be examined. When a certain modifier *m* has different values *mvs*, *mvt* in *ctxt_s* and *ctxt_t*, respectively, a context conflict is thus determined. The scenario of determining context conflicts is illustrated earlier in Figure 7. For example, *dt3* (see Table 4) is an implicit data transfer involving variable *tickerQuoteDate* which contains two data elements *dateofQuote* and *tickerSymbol*. In the WSDL descriptions of *S1* and *S3*, *dateofQuote* is annotated to concept *date* of the ontology. Concept *date* has a modifier *dateFormat* with different values in the contexts of *S1* and *S3*: “dd-mm-yyyy” for *S1* and “mm/dd/yyyy” for *S3* (see Table 3). As a result, a context conflict occurs when *dateofQuote* is transferred through data transfer *dt3* from *S1* to *S3*. There is no conflict for *tickerSymbol* because it has no modifier.

Each context conflict can be represented using the following form:

$$ctxtConflict(dt, C, ctxt\_s, ctxt\_t, [(m_i, mvs_i, mvt_i)]_{i=\{1,\ldots,n\}})$$

where *dt* is the data transfer in which the context conflict occurs. $[(m_i, mvs_i, mvt_i)]_{i=\{1,\ldots,n\}}$ depicts the array of n modifiers with different values in *ctxt_s* and *ctxt_t*. Algorithm 2 is developed to automate the procedure of conflict determination. As shown in Table 5, three context conflicts in the naive BPEL composition are determined.

Algorithm 2. Detecting Context Conflicts

```
Input: BPEL process proc, the set of data transfers DT = {dt},
       the set of annotated WSDL description WS = {ws}, Ontology onto;
Output: The set of context conflicts CC = {cc};
set CC = ∅
for each data transfer dt in DT
  ws_s ← getSourceService(dt, proc, WS), ws_t ← getTargetService(dt, proc, WS)
  data_s ← getSourceDataElement(ws_s, dt), data_t ← getTargetDataElement(ws_t, dt)
  c ← getConcept(ws_s, data_s)
  ctxt_s ← getContext(ws_s, data_s), ctxt_t ← getContext(ws_t, data_t)
  for each modifier m of c in onto
    mvs ← getModifierValue(c, m, ctxt_s), mvt ← getModifierValue(c, m, ctxt_t)
      if mvs ≠ mvt
       then cc ← getContextConflict(C, m, ctxt_s, ctxt_t, mvs, mvt)
            CC ← CC ∪ {cc}
return CC;
```

Table 5. Context Conflicts in the Composition Process of Example 3

cc1	*ctxtConflict* (*dt3*, *date*, *ctxt1*, *ctxt3*, [(*dateFormat*, “dd-mm-yyyy”, “mm/dd/yyyy”)])
cc2	*ctxtConflict* (*dt4*, *openingPrice*, *ctxt2*, *ctxt0*, [(*currency*, “USD”, “GBP”)])
cc3	*ctxtConflict* (*dt5*, *marketCap*, *ctxt3*, *ctxt0*, [(scaleFactor, “1000”, “1”); (currency, “USD”, “GBP”)])

3.2.3. Incorporating Conversions. Once a context conflict is determined within a data transfer, it is needed to assemble an appropriate conversion to reconcile the conflict. The appropriate conversion is either a predefined atomic conversion or a composite one assembled using several atomic conversions. For reconciliation, the identified conversion is incorporated into the data transfer to convert the data in the source context to the target context.

When the determined context conflict occurs in an implicit data transfer, the data transfer needs to be made explicit in order to incorporate the conversion. Suppose *var* is the variable involved in the implicit data transfer. To make the data transfer explicit, it is needed to create a new variable named *var_t* which has the same element type as *var*, and to insert an <assign> activity into the data transfer for copying *var* to *var_t*. As shown in Table 5, data transfer *dt3* is an implicit data transfer where a context conflict of date format occurs. To make *dt3* explicit, a new variable *tickerQuoteDate_t* is declared using the same element type as variable *tickerQuoteDate*. Since *tickerQuoteDate* has two data elements *dateofQuote* and *tickerSymbol*, the <assign> activity inserted into *dt3* has two <copy> activities for copying *dateofQuote* and *tickerSymbol* of *tickerQuoteDate* to that of *tickerQuoteDate_t*. Then, the input variable of the invocation of *S3* is changed from variable *tickerQuoteDate* to variable *tickerQuoteDate_t*. After this step, all data transfers with context conflicts are made explicit.

When a context conflict involves only one modifier, it can be reconciled using a predefined atomic conversion. For example, the context conflict *cc1*, as shown in Table 5, involves modifier *dateFormat* of concept *date*. It is thus easy to identify the atomic conversion $cvt_{\text{dateFormat}}$ that can reconcile *cc1*. The conversion $cvt_{\text{dateFormat}}$ is applied through substituting the input *vs* of the XPath function as data element *dateofQuote*. Also, the context conflict *cc2* involves modifier *currency* of concept *openingPrice*, which can be reconciled using the atomic conversion cvt_{currency}. As discussed in Section 3.1.3, cvt_{currency} is implemented by the external currency converter service *S4* rather than using XPath function. Thus, an <invoke> activity is inserted in the data transfer *dt4* of *cc2* in order to convert *openingPrice* in “USD” from *S2* to the equivalent price in “GBP”, an output data of *CS*. Necessary <assign> activities are also inserted to explicitly transfer the exchanged data.

Algorithm 3. Incorporating Conversions

```
Input: BPEL process proc, the set of annotated WSDL description WS = {ws},
       the set of context conflicts CC = {cc},
       the set of predefined atomic conversions CVT = {cvt};
Output: Mediated BPEL process mediatedProc;
mediatedProc = proc
for each context conflict cc in CC
  dt ← getDataTransfer(cc)
  if isImplicit(dt) == 'TRUE'
   then var ← getVariable(dt), var_t ← declareNewVariable(var),
        insertAssign(mediatedProc, dt, var, var_t)
  AMV = [(m_i, mvs_i, mvt_i)] ← getArrayOfModifierValues(cc)
  if |AMV| == "1"
    then cvt ← getAtomicConversion(cc, m, CVT)
         insertConversion(mediatedProc, cvt)
  else
    for each (m_i, mvs_i, mvt_i) in AMV
      cvt_i ← getAtomicConversion(cc, m_i, CVT), insertConversion(mediatedProc, cvt_i)
return mediatedProc;
```

When a certain context conflict involves two or more modifiers, no predefined atomic conversion can reconcile the context conflict, as each atomic conversion is defined with only one modifier. In this case, the context conflict can still be reconciled using the composition of multiple atomic conversions, each of which is defined with one of the modifiers involved in the context conflict. For example, the context conflict *cc3* involves two modifiers *scaleFactor* and *currency* of concept *marketCapital*. Among the predefined atomic conversions, modifier *scaleFactor* and *currency* correspond to $cvt_{\text{scaleFactor}}$, cvt_{currency}, respectively. Therefore, *cc3* can be reconciled using the composition of the two atomic conversions, successively applying $cvt_{\text{scaleFactor}}$ and cvt_{currency}. Specifically, the output data *dailyMarketCap* from *S3* is first converted by $cvt_{\text{scaleFactor}}$ from the scale factor “1000” to “1”, and then converted by cvt_{currency} from the currency “USD” to the equivalent amount in “GBP”. After the two-step composite conversion consisting of $cvt_{\text{scaleFactor}}$ and cvt_{currency}, the exchanged data is converted and transferred to the output data *openingMarketCap* of *CS*. Algorithm 3 is developed to automate the procedure of assembling conversions and generating the mediated BPEL to reconcile the determined context conflicts.

4. Prototype Implementation

We implemented a proof-of-concept prototype, named Context Mediation Tool (CMT), as a JAVA application, to demonstrate the reconciliation approach. The lightweight ontology with structured contexts is defined using the COIN Model Application Editor[14] which is a Web-based tool for creating and editing COIN-style ontology and contexts in RDF/OWL. Atomic conversions between the contexts are defined in a specification file. The WSDL descriptions of the composite and component services (e.g., *CS* and *S1* ~ *S3* of Example 3) are annotated using our context annotation method. To facilitate the annotation task, we extended an open-source Eclipse plug-in for semantic annotation (i.e., Radiant[15]) and developed the context annotation tool *Radiant4Context*. We assume naive BPEL composition processes with possible data misinterpretation problems are defined using any typical BPEL tool.

CMT is used to create a mediation project and consume all the above documents. The reasoning engine implemented within CMT can automatically perform the reconciliation algorithms described in Section 3.2. Take Example 3 for instance. CMT first performs Algorithm 1 to identify the three implicit and two explicit data transfers in the naive BPEL composition process. Then, CMT continues to use Algorithm 2 to determine the three context conflicts. Finally, CMT uses Algorithm 3 to select three atomic conversions $cvt_{\mathrm{dateFormat}}$, $cvt_{\mathrm{scaleFactor}}$ and cvt_{currency} from predefined conversion library[16] and incorporates them into corresponding data transfers to reconcile the conflicts.

CMT has three working areas for the mediation tasks, as shown in Figure 8. The first working area requires the user to import the involved documents of the composition into the mediation project. To monitor the results of different mediation steps, the second working area, Mediation Stage, allows the user to choose one of the four consecutive stages, including Naive BPEL Process, Data Transfers, Context Conflicts, and Mediated BPEL Process. These stages provide the intermediate and final results that the approach produces while addressing context differences among services involved in the composition. Eventually, CMT produces the mediated BPEL composition process. Note that CMT can perform all the mediation steps in an automatic and consecutive way.

[14] http://interchange.mit.edu/appEditor/TextInterface.aspx?location=MIT

[15] http://lsdis.cs.uga.edu/projects/meteor-s/downloads/index.php?page=1

[16] We recommend that libraries of such atom conversions be established that can be reused for future compositions.

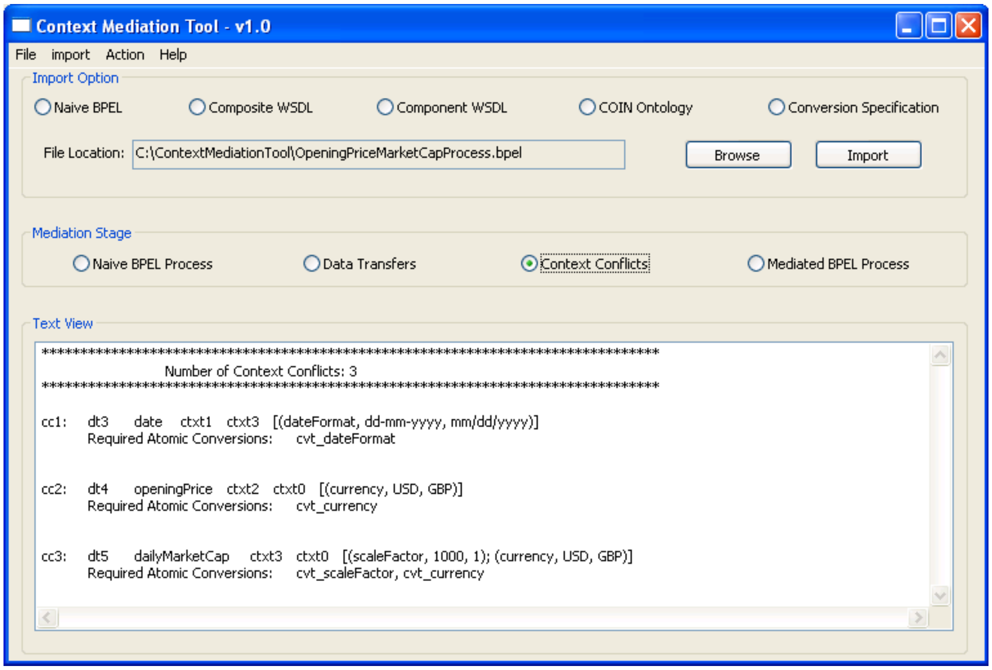

Figure 8: Snapshot of CMT at stage context conflicts.

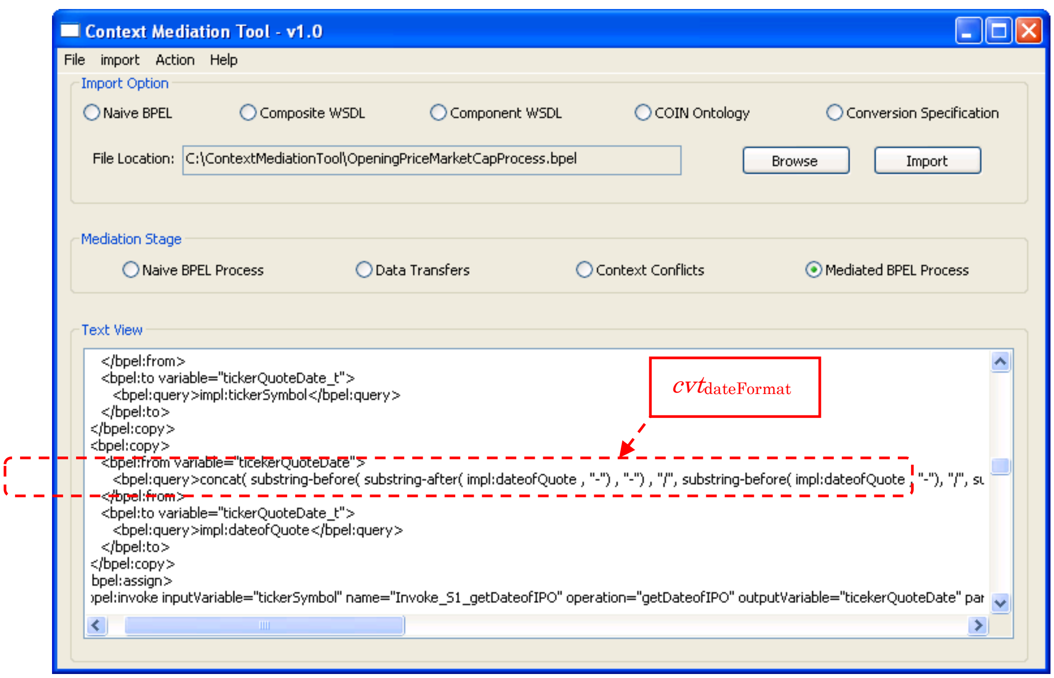

Figure 9: Snapshot of CMT at stage mediated BPEL process.

Figure 8 shows the snapshot of CMT at the stage Context Conflicts where the three context conflicts in the composition process of Example 3 and corresponding atomic conversions required for the reconciliation are identified. At the stage Mediated BPEL Process, CMT produces the mediated BPEL composition process with incorporated conversions. Figure 9 shows the snapshot of CMT in which the XPath function for the conversion $cvt_{\mathrm{dateFormat}}$ is embedded in the mediated BPEL composition process.

5. Validation and Evaluation

5.1 Validation

We validated the solution approach by applying it to several composition processes that involve various interpretation conflicts. Here we show the results of applying the approach to Example 3 (see Section 2.1.3) and Example 2 (see Section 2.1.2). For Example 3, Figure 10 shows the snapshot of the naive BPEL composition process defined using ActiveVOS BPEL Designer. Note that we have used a schematic notation in Figure 3 to illustrate the naïve BPEL composition process. Since the interpretation conflicts exist at the data instance level, ActiveVOS BPEL Designer cannot detect the conflicts of data interpretation and fails to alert any error. But severe errors and failures will occur when one attempts to executes the naive BPEL composition.

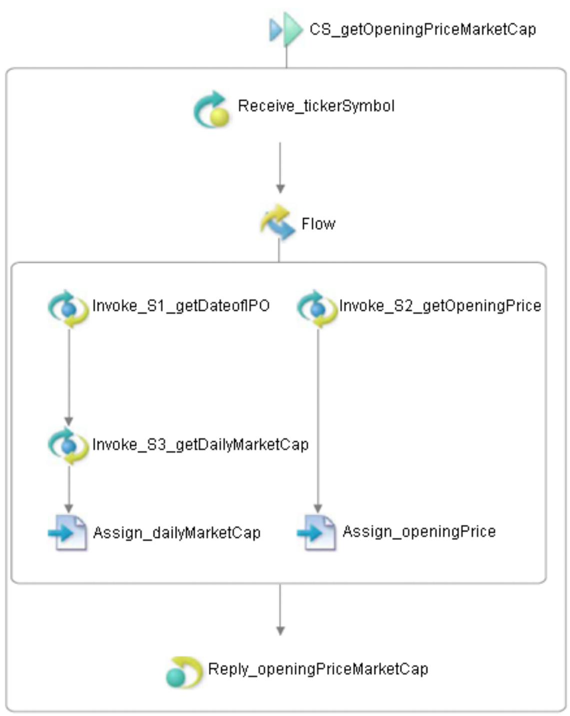

Figure 10: Naive BPEL composition process with context conflicts.

The prototype CMT can automatically produce the mediated BPEL composition consecutively. After the mediated BPEL composition is produced, we import it into ActiveVOS BPEL Designer for validation purpose. Figure 11 shows the snapshot of the mediated BPEL process with the incorporated conversions. As we can see, CMT inserts a <assign> activity into the composition process between the invocations of *S1* and *S3* in order to reconcile the conflict of date format (i.e., *cc1* in Table 5). In fact, CMT embeds the XPath conversion function $cvt_{\text{dateFormat}}$ in the <copy> element of the <assign> activity and uses it to convert the date format from “dd-mm-yyyy” to “mm/dd/yyyy”. To reconcile the conflict of currency (i.e., *cc2* in Table 5), CMT inserts the invocation of the external currency converter service *S4*. By invoking *S4*, the output *openingPrice* in “USD” from *S2* is converted to the equivalent price in “GBP” as the output of *CS*. Finally, CMT inserts a <assign> activity and a <invoke> activity consecutively in the composition process to reconcile the conflicts of scale factor and currency (i.e., *cc3* in Table 5). The XPath conversion function $cvt_{\text{scaleFactor}}$ is embedded by CMT in the <copy> element of the <assign> activity and used to reconcile the conflict of scale factor. *S4* is used to reconcile the conflict of currency (see *cc2* and *cc3* in Figure 11).

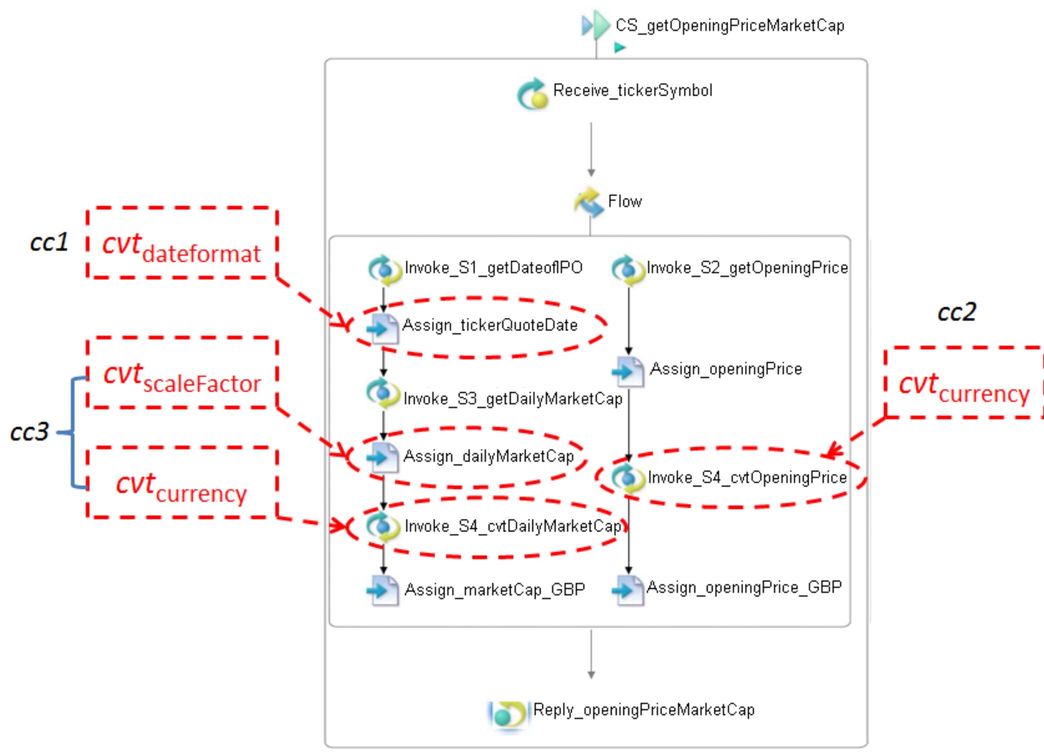

Figure 11: Mediated BPEL composition process with incorporated conversions.

In order to validate the correctness of the mediated BPEL composition process, we provide a number of testing data values for the input of *CS* and the output of the services (i.e., *S1* ~ *S3* and *S4*). We utilize the simulation feature of ActiveVOS BPEL Designer to simulate the execution of the mediated BPEL process. The execution results indicated that: a) the mediated BPEL process properly completed without any deadlocks or errors; b) all the context conflicts were successfully reconciled – different date formats, scale factors and currencies were correctly converted between the involved services; and c) *CS* produced the expected output: *openingPrice* and *openingMarketCap*.

For Example 2 (see Figure 2), three context conflicts are determined using CMT: the date format difference, the currency difference, and the VAT difference – *HotwireDeals* provides the room charge not including value-added taxes, while *ConfHotelDeals* is expected to provide the hotel expense including the taxes. Similar to Example 3, the date format difference and the currency difference can be resolved by $cvt_{\text{dateFormat}}$ and cvt_{currency}, respectively. Differently, the VAT difference needs to be resolved by using a new conversion cvt_{VAT} which is implemented as an external service *TaxesCalculator*. *TaxesCalculator*'s operation *getVATAdded* consumes a money value without value-added taxes and returns the money value with value-added taxes. In a similar way, CMT produces the mediated composition for *ConfHotelDeals* with all determined context conflicts reconciled. Figure 12 illustrates the mediated composition process with all necessary conversions (indicated in bolded red boxes) inserted.

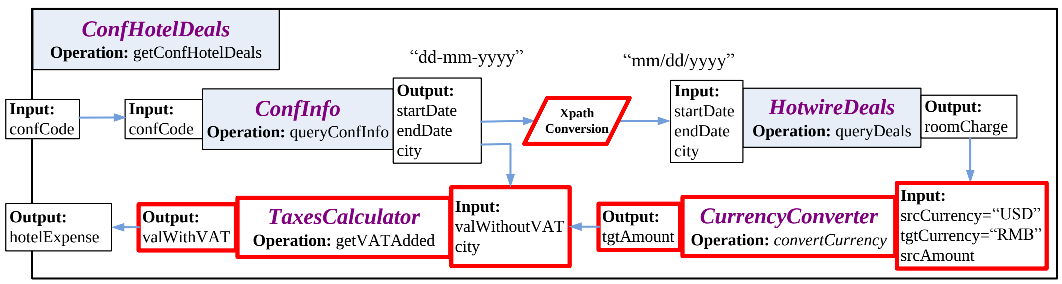

Figure 12: Mediated composition with conceptual VAT difference reconciled.

5.2 Evaluation

The reconciliation approach is evaluated both qualitatively and quantitatively. The evaluation results are presented in the following two subsections.

5.2.1. Qualitative Evaluation. The qualitative evaluation of the reconciliation approach is conducted by checking whether it can handle more general and complicated composition situations. Specifically, we try to answer the following two questions: (1) What types of data misinterpretation problems can the approach address? and (2) What types of Web services composition can the approach support? The method of qualitative evaluation used in this paper is similar to the method of key feature comparison, which is a credible method for evaluating software engineering-based approaches (VIDE 2009) and recently used by (Abeywickrama and Ramakrishnan 2012) as well.[17]

For the first question, we find that the use of modifiers in a lightweight ontology is a quite versatile modeling technique. It allows for the representation of each type of interpretation conflicts discussed in Section 2.2. For example, to address the difference of date format or currency (a kind of unit

[17] Note that the key feature comparison of our work with the prior approaches is presented in Section 6.

of measure) at the representational level, we use the modifier of date format or currency and corresponding conversions (i.e., $cvt_{\mathrm{dateFormat}}$, cvt_{currency}) and demonstrate the feasibility through Example 3. In Example 2 we use the modifier of value-added taxes and the conversion cvt_{VAT} to deal with the difference of value-added taxes, a kind of conceptual-level data misinterpretation problems. Other conceptual-level problems like those of "Corporate Householding" (Madnick et al. 2003) and temporal-level problems (Zhu and Madnick 2009) can also be modeled using appropriate modifiers and addressed in a similar way. With the ontology/context modeling and semantic annotation in place, all the possible data misinterpretation problems in Table 2 that may occur in Web services composition can be addressed by the approach.

Since BPEL becomes the OASIS standard for defining Web services composition in practice, the approach presented in this paper focuses on addressing BPEL-based composition processes. BPEL specification provides four types of interaction activities (i.e., <receive>, <reply>, <invoke> and <onMessage> within <pick>) to define interaction patterns between the composition process and participant services. Also, BPEL provides several basic workflow constructs (e.g., *sequence*, *parallel*, *choice* and *iteration*) to define the composition processes. In our work all these interaction activities and workflow constructs have been taken into consideration when we developed Algorithm 1. In other words, Algorithm 1 can be used to automatically inspect any composition process defined using BPEL and identify data transfers within the process. For example, we demonstrate the capability of the approach to address Example 3 which involves three of the four types of interaction activities (i.e., <receive>, <reply>, <invoke>) and the *sequential* and *parallel* workflow constructs. <onMessage> is similar to <receive>, as both handle the message arrival. Thus, Algorithm 1 analyzes <onMessage> in a similar way as it does for <receive>. Since control-flow conditions of choice and iteration are irrelevant to the identification of data transfers, Algorithm 1 will examine each workflow branch defined by the construct of choice or iteration in a similar way as it does for the sequence or parallel workflows. After the data transfers in the composition process are identified using Algorithm 1, Algorithm 2 and Algorithm 3 are used to determine and reconcile possible data interpretation conflicts. Therefore, the approach can support any Web services composition defined using the BPEL and WSDL standards.

5.2.2. Quantitative Evaluation. A quantitative evaluation of the proposed approach is carried out with the focus on assessing human efforts needed for reconciling data interpretation conflicts. Although a direct measurement of human efforts can be obtained through empirical experiments, it is often difficult to set up such appropriate experiments to reliably and objectively measure the evaluation metrics. Instead, we will consider the complexity of how mediation is accomplished in the brute-force approach compared with our approach.

Let us suppose an extreme case where there are *N* services (including the composite service) that have different data interpretations and interact with each other in the composition. In such a case, there are *N*(N-1)/2* service-to-service interactions in the composition. Thus, the brute-force approach (see the discussion in Section 2.3) has to examine each of the service-to-service interactions to ensure the interoperability between every two interacting services. Each service-to-service interaction involves an XML message probably with multiple data elements. Suppose on average there are *K* data elements in the XML message between any two services and *D* dimensions of data interpretation conflicts (e.g., currency and scale factor) associated with each data element, then in total the brute-force approach has to examine *K*D*N*(N-1)/2* possible places where data interpretation conflicts might occur. Wherever a data interpretation conflict is detected, the brute-force approach has to construct a conversion and insert it to the appropriate place in the composition. As the number of services *N* and the number of data elements in XML messages *K* increase, the amount of manual work of inspecting and rewriting BPEL increases quickly. Maintaining manually created BPEL over time is also labor-intensive and error-prone.

In contrast, our reconciliation approach requires manual creation of a lightweight ontology, annotation of each service, and provision of atomic conversions, each of which concerns only one data interpretation dimension. Although this may appear to be undesirable beforehand, it actually reduces the amount of pairwise manual inspection and conversion construction using annotation for individual services. More importantly, our approach can automatically examine the XML message between each service-to-service interaction, identify context conflicts, and build and insert appropriate conversions in the composition. Thus, the key advantage of our reconciliation approach lies in the automatic generation of mediated BPEL which otherwise would require significant amount of manual work as in the brute-force approach.

Let us use a specific example to demonstrate the advantage of our approach. Assume that the developer of Example 3 later wanted to serve diverse users that require any combination of 10 different currencies and 4 scale factors (i.e., 1, 1K, 1M, 1B). The component services, e.g., *S3*, may also change their currencies and scale factors. In such a case, both the output *dailyMarketCap* of *S3* and the output *openingMarketCap* of *CS* may use 40 (=10×4) different data interpretations. To convert the output *dailyMarketCap* of *S3* to the output *openingMarketCap* of *CS*, it would be most likely for the developers to manually specify 1560 (=39×40) custom conversions if they used the brute-force approach. An even worse case would arise if currencies and scale factors of *CS*, *S2* and *S3* changed over time independently. Comparatively, our approach only requires two parameterized conversions (i.e., $cvt_{\text{scaleFactor}}$ and cvt_{currency}). More importantly, as long as no additional dimension of data interpretation difference is introduced, there is no need to define new conversions even if the involved services were to be added (or removed) in the composition, or the workflow logic of the composition process were to be changed. In practice such situations frequently happen because the implementations of Web services and service composition often evolve in the fast-changing global business environment.

There are two points to note regarding the examples in this paper. First, for reasons of brevity and simplicity, the examples in the paper only include a few web services. There are large complex applications built using hundreds of web services, they would not be so easy for a human to examine the naive BPEL and resolve all the conflicts – and do that error-free. Second, the scalability issue not only exists at initial development of the composite application but over its entire life cycle. If a change is needed to the application or happens to the specifications of one or more of the web services, then the entire resolution process must be reviewed and appropriate changes made by the human. With our approach, most of this is automated, only the context specifications (and occasionally the ontology) have to be updated.

6. Related Work and comparison

The basic Web services standards (e.g., WSDL, BPEL) generally ignore data semantics, rendering semantic composition and interoperability far from reality. A research area, referred to as Semantic Web Services (SWSs), has emerged to apply Semantic Web technologies to Web services (Burstein et al. 2005; McIlraith et al. 2001; Sycara et al. 2003). OWL-S (Martin et al. 2007), WSMF/WSMO (Fensel and Bussler 2002; Lausen et al. 2005) and METEOR-S (Patil et al. 2004; Sivashanmugam et al. 2003) are three major initiatives that have developed languages and frameworks to explicitly add semantics into the Web services descriptions. Despite the ontological foundations provided by these efforts, it is still necessary to develop effective approaches to semantic composition.

Data misinterpretation among Web services can be considered as a semantic heterogeneity problem. However, the literature provides only a few approaches to handle the challenging problem in Web services composition. The initial work in (Spencer and Liu 2004) proposes to use data transformation rules to convert the data exchanged between services. This work requires a common ontology described in OWL (particularly in description logic) and the correspondences between the ontology and WSDL descriptions defined using OWL-S. Rather than using OWL-S, the approach in (Nagarajan et al. 2006; Nagarajan et al. 2007) proposes to perform semantic annotation by using WSDL-S which is the ancestor of SAWSDL and more consistent with existing industrial standards and practices. The approach focuses on addressing schematic differences of the exchanged messages by using schematic conversations (e.g., XSLT). The work in (Gagne et al. 2006; Sabbouh et al. 2008) proposes a set of mapping relations to establish direct correspondences between the messages of two WSDL-based services.

Then, the common ontology can be constructed based on these correspondences and data-level differences are resolved by predefined conversions. Generally, those approaches require each participant services to be annotated and mapped to a common ontology serving as the global schema. However, it is more costly to construct and maintain this type of global schema than the lightweight ontology used in our approach, which only needs a small set of generic concepts. More importantly, the mappings or transformation rules required by those approaches are created manually to perform direct conversions between the exchanged messages. In contrast, the actual conversions in our approach can be automatically composed using a small number of atomic, parameterized conversions. Furthermore, those approaches only focus on dealing with a pair of participant services, rather than a composition consisting of multiple services.

To the best of our knowledge, the work in (Mrissa et al. 2006a; b; Mrissa et al. 2007), which also draws on the original COIN strategy, is most related to this paper. However, our solution is significantly distinct from their work in multiple aspects. (1) Their work ignores considering the composite service whose context may be different from any component service, while our solution can address both composite and component services. (2) They embed context definition in WSDL descriptions using a non-standard extension. As a result, their approach suffers from the proliferation of redundant context descriptions when multiple services share the same context. In contrast, we avoid this problem by separating ontology and context definitions from the annotated WSDL descriptions. (3) Only context conflicts between the <invoke> activities in the BPEL composition are considered in their work, while context conflicts between all interaction activities (e.g., <receive>, <reply>, <invoke> and <onMessage>) can be handled using our solution. (4) Since in their work each context conflict needs to be reconciled using the a priori specification of an external service, they miss the opportunity to reuse predefined atomic conversions and the capability of conversion composition. In our work we define a parameterized atomic conversion for each modifier and use reasoning algorithms to automatically generate composite conversions consisting of atomic conversions to handle complex context differences. Thus, the number of predefined conversions is largely reduced.

In addition to the literature on Web services, it is worth noting some interesting works (Sun et al. 2006; Tan et al. 2009; Hamid et al. 2010) from the domain of process/workflow management. Sun et al. (2006) develop a data-flow specification for detecting data-flow anomalies within a process/workflow, including missing data, redundant data and potential data conflicts. With a different focus from our work, their work provides no automatic approach that can be used to produce the data-flow specification. Also, semantic heterogeneity of the data exchanged is not considered in their work. We believe that Algorithm 1 can be adapted to construct data-flow specification, so that potential data-flow anomalies can be also addressed. Both Tan et al. (2009) and Hamid et al. (2010) focus on developing mediator services that could address the workflow inconsistencies between services involved in the composition. Our work complements those studies in that we focus on resolving data misinterpretation conflicts in the composition.

7. Conclusion

Differences of data interpretation widely exist among Web services and severely hamper their composition and interoperability. To this end, we adopt the context perspective to deal with the data misinterpretation problems. We describe the lightweight ontology with structured contexts to define a small set of generic concepts among the services involved in the composition. The multiple specializations of the generic concepts, which are actually used by different services, are structured into different contexts so that the differences can be treated as context differences. We introduce a flexible, standard-compliant mechanism of semantic annotation to relate the syntactic WSDL descriptions to the ontology. Given the naive BPEL composition ignoring semantic differences, the reconciliation approach can automatically determine context conflicts and produce the mediated BPEL that incorporates necessary conversions. The incorporated conversions can be predefined atomic conversions or composite conversions that are dynamically constructed using the atomic ones. The context-based reconciliation approach has desirable properties of adaptability, extensibility and scalability. In the long run, it can significantly alleviate the reconciliation efforts for Web services composition.

Our approach has two limitations. First, the lightweight ontology enriched with modifiers and contexts needs to be defined manually. Although the ontology has a small number of generic concepts compared to other heavyweight ontologies, efforts are required to define the ontology. Second, our approach requires the participant services be annotated with respect to the ontology. Although it is a nontrivial task, the semantic annotation allows for separation of declarative semantic descriptions from the programming code (e.g., JAVA and ASP.NET) and provides the prerequisite through which our approach can automatically detect and reconcile the data misinterpretation conflicts. To alleviate the cost of the annotation task, we have extended an open-source Eclipse plug-in (i.e., Radiant) and developed a context annotation tool. Thus, developers can easily use our context annotation tool to add context information.

Fortunately, there has been a growing trend (Savas et al. 2009) that authors of data services are encouraged to provide certain metadata definition and semantic annotation. Also, researchers have begun to develop various solutions (Uren et al. 2006; Mrissa et al. 2007; Di Lorenzo et al. 2009), albeit with limited scope, to produce context information for interpreting the data provided by Web services.

Therefore, we expect over time such context information will become increasingly available in the published Web services so that our proposed approach can be used more easily and smoothly.

Future work is needed to address the limitations of our approach. Specifically, we plan to develop techniques to automate the construction of the lightweight ontology for Web services. Also, we intend to integrate existing annotation methods (Uren et al. 2006) with our approach to facilitate semantic annotation. Additionally, we plan to adapt several existing service discovery techniques and integrate them with our approach so that the necessary external mediation services could be more easily discovered and used by the tool CMT. Despite the identified future work, our approach, even in its current form, can substantially reduce the effort and possible errors of manual Web services composition. We expect our approach and the prototype can be applied in the practice of SOC and the development of Web-based information systems.

References

Alves, A., A. Arkin, S. Askary, C. Barreto, B. Bloch, F. Curbera, M. Ford, Y. Goland, A. Guizar, N. Kartha. 2007. Web services business process execution language version 2.0. *OASIS Standard* **11**.

Becker, J., A. Dreiling, R. Holten, M. Ribbert. 2003. Specifying information systems for business process integration–A management perspective*. *Information Systems and E-Business Management* **1**(3) 231-263.

Bressan, S., C. Goh, N. Levina, S. Madnick, A. Shah, M. Siegel. 2000. Context Knowledge Representation and Reasoning in the Context Interchange System. *Applied Intelligence* **13**(2) 165-180.

Burstein, M., C. Bussler, T. Finin, M.N. Huhns, M. Paolucci, A.P. Sheth, S. Williams, M. Zaremba. 2005. A semantic Web services architecture. *Internet Computing, IEEE* **9**(5) 72-81.

Christensen, E., F. Curbera, G. Meredith, S. Weerawarana. 2001. *Web services description language (WSDL) 1.1*. W3C Recommendation.

Farrell, J., H. Lausen. 2007. Semantic Annotations for WSDL and XML Schema. *W3C Recommendation, Available at http://www.w3.org/TR/2007/REC-sawsdl-20070828/*.

Fensel, D., C. Bussler. 2002. The Web Service Modeling Framework WSMF. *Electronic Commerce Research and Applications* **1**(2) 113-137.

Gagne, D., M. Sabbouh, S. Bennett, S. Powers. 2006. *Using Data Semantics to Enable Automatic Composition of Web Services*.

Goh, C.H., S. Bressan, S. Madnick, M. Siegel. 1999. Context interchange: new features and formalisms for the intelligent integration of information. *ACM Transactions on Information Systems (TOIS)* **17**(3) 270-293.

Kopecký, J., T. Vitvar, C. Bournez, J. Farrell. 2007. SAWSDL: Semantic Annotations for WSDL and XML Schema. *IEEE INTERNET COMPUTING* **11**(6) 60-67.

Krishnan, R., J. Peters, R. Padman, D. Kaplan. 2005. On data reliability assessment in accounting information systems. *Information Systems Research* **16**(3) 307.

Lausen, H., A. Polleres, D. Roman. 2005. Web Service Modeling Ontology (WSMO). *W3C Member Submission* **3**.

Li, X., S. Madnick, H. Zhu, Y. Fan. 2009. *An Approach to Composing Web Services with Context Heterogeneity*. Los Angeles, CA, USA.

Li, X., S. Madnick, H. Zhu, Y.S. Fan. 2009. *Reconciling semantic heterogeneity in Web services composition*. Phoenix, AZ, USA.

Madnick, S., A. Firat, M. Siegel. 2000. *The Caméléon Web Wrapper Engine*. Cairo, Egypt.

Madnick, S., R. Wang, X. Xian. 2003. The design and implementation of a corporate householding knowledge processor to improve data quality. *Journal of management information systems* **20**(3) 41-70.

Madnick, S., H. Zhu. 2006. Improving data quality through effective use of data semantics. *Data & Knowledge Engineering* **59**(2) 460-475.

Martin, D., M. Burstein, D. McDermott, S. McIlraith, M. Paolucci, K. Sycara, D. McGuinness, E. Sirin, N. Srinivasan. 2007. Bringing Semantics to Web Services with OWL-S. *World Wide Web* **10**(3) 243-277.

McIlraith, S.A., T.C. Son, H. Zeng. 2001. Semantic Web Services. *IEEE INTELLIGENT SYSTEMS* **16**(2) 46-53.

Mrissa, M., C. Ghedira, D. Benslimane, Z. Maamar. 2006. *Context and Semantic Composition of Web Services*. Krakow, Poland.

Mrissa, M., C. Ghedira, D. Benslimane, Z. Maamar. 2006. *A Context Model for Semantic Mediation in Web Services Composition*. Tucson, Arizona, USA.

Mrissa, M., C. Ghedira, D. Benslimane, Z. Maamar, F. Rosenberg, S. Dustdar. 2007. A context-based mediation approach to compose semantic Web services. *ACM Transactions On Internet Technology* **8**(1) 4.

Nagarajan, M., K. Verma, A.P. Sheth, J. Miller, J. Lathem. 2006. *Semantic Interoperability of Web Services - Challenges and Experiences*. Chicago, USA.

Nagarajan, M., K. Verma, A.P. Sheth, J.A. Miller. 2007. Ontology driven data mediation in web services. *International Journal of Web Services Research* **4**(4) 104-126.

Papazoglou, M.P., P. Traverso, S. Dustdar, F. Leymann. 2007. Service-Oriented Computing: State of the Art and Research Challenges. *IEEE Computer* **40**(11) 38-45.

Patil, A.A., S.A. Oundhakar, A.P. Sheth, K. Verma. 2004. *Meteor-s web service annotation framework*.

Sabbouh, M., J.L. Higginson, C. Wan, S.R. Bennett. 2008. *Using Mapping Relations to Semi Automatically Compose Web Services*.

Seligman, L.J., A. Rosenthal, P.E. Lehner, A. Smith. 2002. Data Integration: Where Does the Time Go? *IEEE Data Engineering Bulletin* **25**(3) 3-10.

Sivashanmugam, K., K. Verma, A. Sheth, J. Miller. 2003. *Adding Semantics to Web Services Standards*. IEEE Computer Society, Las Vegas, Nevada, USA.

Spencer, B., S. Liu. 2004. *Inferring Data Transformation Rules to Integrate Semantic Web Services*. Springer Verlag, Hiroshima, Japan.

Storey, V., A. Burton-Jones, V. Sugumaran, S. Purao. 2008. CONQUER: A Methodology for Context-Aware Query Processing on the World Wide Web. *Information Systems Research* **19**(1) 3-25.

Sun, S.X., J.L. Zhao, J.F. Nunamaker, O.R.L. Sheng. 2006. Formulating the data-flow perspective for business process management. *Information Systems Research* **17**(4) 374-391.

Sycara, K., M. Paolucci, A. Ankolekar, N. Srinivasan. 2003. Automated discovery, interaction and composition of Semantic Web services. *Web Semantics: Science, Services and Agents on the World Wide Web* **1**(1) 27-46.

Uren, V., P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, F. Ciravegna. 2006. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web* **4**(1) 14-28.

van der Aalst, W., A. Kumar. 2003. XML-based schema definition for support of interorganizational workflow. *Information Systems Research* **14**(1) 23-46.

Verma, K., A. Sheth. 2007. Semantically Annotating a Web Service. *IEEE INTERNET COMPUTING* **11**(2) 83-85.

Wache, H., T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, S. Hübner. 2001. *Ontology-based integration of information-a survey of existing approaches*. Seattle, WA, USA.

Yu, Q., X. Liu, A. Bouguettaya, B. Medjahed. 2008. Deploying and managing Web services: issues, solutions, and directions. *The International Journal on Very Large Data Bases* **17**(3) 537-572.

Zhu, H., S. Madnick. 2009. Reconciliation of Temporal Semantic Heterogeneity in Evolving Information Systems. *Ingénierie des Systèmes d'Information (Networking and Information Systems)* **14**(6) 59-74.

Zhu, H., S.E. Madnick. 2007. *Scalable Interoperability Through the Use of COIN Lightweight Ontology*. SPRINGER-VERLAG, Seoul, Korea.

Essay Two

Herding and Social Media Word-of-Mouth: Evidence from Groupon

Abstract

Understanding the various social influence mechanisms that affect consumers' online shopping behaviors has become more important with the widespread adoption of social media. This study explores how herding and social media word-of-mouth (WOM) drive product sales. While herding helps updating consumers' beliefs about the product quality, social media WOM can have an advertising effect in addition to providing quality signals. Using a panel data set consisting of about 500 deals from Groupon.com, we find both herding and Facebook-mediated WOM lead to additional product sales, whereas Twitter-mediated WOM has no significant impact on sales. More importantly, we theorize the interaction effect between herding and social media WOM and show herding and Facebook-mediated WOM are complements in driving product sales. The complementarity supports the current practice of daily-deal sites where both mechanisms are often implemented together. To uncover the underlying mechanisms, we find the herding effect is more salient for experience goods than for search goods, but the effect of Facebook-mediated WOM does not significantly differ between the two product categories. The comparison suggests that signaling product quality is the underlying mechanism of herding, while the effect of Facebook-mediated WOM is primarily through advertising, rather than signaling. Our findings are robust to a number of different estimation specifications and identification strategies.

1. Introduction

Reaching out hundreds of millions of customers with deep-discounted vouchers, daily-deal sites (e.g., Groupon.com, LivingSocial.com) have become a popular platform for advertising and encouraging consumers to try new products (Dholakia 2011). Since Groupon has launched in 2008, daily-deal businesses have been growing exponentially. As of April 2012, consumers in North America have spent approximately $7 million a day[18] (more than $2.5 billion a year) on daily deals and it is projected to reach $4 billion a year by 2015.

Although daily deals have become a popular marketing vehicle, their underlying economic mechanisms have rarely been examined. The few studies on daily deals use existing marketing frameworks, such as couponing (Kumar and Rajan 2012) and price discrimination (Edelman et al. 2010),

[18] See http://savvr.com/2012/04/top-10-highest-grossing-daily-deals-of-all-time/

to understand the economics of daily deals. However, daily deals differ from traditional marketing vehicles in at least two important aspects and deserve to be investigated separately. First, daily-deal sites explicitly highlight the total number of vouchers sold in real-time. By allowing potential buyers to observe prior others' purchasing decisions, daily-deal sites can create an information cascade (i.e., herding), driving even more sales for the popular deals (Zhang and Liu 2012). Second, daily-deal sites can generate a word-of-mouth (WOM) effect via social media platforms, such as Facebook and Twitter. By clicking on the Facebook "Like" or Twitter button on a deal page, shoppers can simultaneously endorse and share the deal to their social ties. Thus, WOM via social media can positively affect sales as well. Lastly, these two aspects of daily deals—herding and social media WOM—are often implemented together. These two mechanisms could potentially interact with each other to affect sales.

The extant literature documents the economic implications of herding and online WOM. Information cascade as a result of herding may affect product adoption and sales, because observing past actions can help updating consumers' beliefs about product quality, especially when prior knowledge is imperfect (Cai et al. 2009, Duan et al. 2009). Similarly, online WOM may help updating consumers' beliefs about product quality and consequently spur sales, especially when WOM comes from social ties (Wojnicki and Godes 2008), as opposed to anonymous strangers such as online reviewers. Simultaneously, WOM via social media may also serve as an advertising vehicle (Chen et al. 2011). Advertising through social contacts is, on average, more reliable than public announcements (Tucker 2012), because friends are more likely to have similar tastes (homophily) or know about a person's idiosyncratic preferences (tie strength). Thus, the advertising effect could further enhance the effect of social media WOM in spurring sales.[19]

While exploring the economic implications of both mechanisms is beneficial, it is important to examine whether herding and social media WOM could interact with each other in driving sales. If herding and social media WOM are substitutes, it is sub-optimal to implement both as one could cannibalize the effect of the other. In such case, firms should choose the mechanism with the highest return on investment and design the appropriate marketing strategy. For example, if action-based herding is more salient in generating future sales, early promotions, such as through celebrity endorsement, could be beneficial, because prior sales can have a "multiple effect" on later sales (Moretti 2011). But if social media WOM is more effective, it would be better to generate buzz via social media platforms, such as encouraging users to "like" a product on Facebook or "tweet" about it on Twitter. On the other hand, however, if herding and social media WOM are complements, it is more beneficial to implement both

[19] In the literature, advertising could have two different effects: informative and persuasive. The informative effect of advertising is to increase the product awareness, while the persuasive effect of advertising is to improve consumers' attitude about the product. In this paper, we refer advertising primarily to its informative effect.

together than separately even when one mechanism dominates the other. As complements, the two mechanisms reinforce each other to generate additional sales such that the sum of both is greater than the sum of the separate parts. Because the nature of their interaction dictates different marketing strategies, it is important to explore whether herding and social media WOM are complements or substitutes. While little prior research explores the interaction between herding and social media WOM, many daily-deal sites (e.g. Groupon.com) implement the two mechanisms in the web design, providing an ideal setting to investigate the interaction effect between the two.

The nature of the information each mechanism provides is key to understand how herding and social media WOM interact to influence consumers' purchasing decisions. Because social media WOM has the dual role of updating consumers' beliefs about product quality and advertising products to potential buyers, its interaction with herding is nuanced. As social media WOM and herding can both signal product quality, they could produce redundant information to consumers, substituting each other's effect. On the other hand, if consumers interpret the information from the two mechanisms to be different, herding and social media WOM can complement each other (Kirmani and Rao 2000). In addition to signaling, social media WOM also serves as an advertising vehicle, broadcasting product information to one's social ties and thus increasing the product awareness. This advertising effect can be distinct from signaling and thus generate complementarities between herding and social media WOM. As potential buyers become aware of a product through the advertising of social media WOM, they are more likely to buy when existing sales are already high (herding) because these prior purchases provide a favorable signal about the product. Hence, high existing sales amplify the advertising effect of social media WOM. Thus, whether herding and social media WOM are complements or substitutes is an empirical question and deserves to be investigated carefully. If social media WOM primarily affects sales through advertising, we may expect a complementary relationship with herding. However, if social media WOM and herding both drive sales through signaling similar information about the product, the two mechanisms can act as substitutes.

Using a panel data set consisting of about 500 deals from Groupon.com, we first find that herding and Facebook-mediated WOM both positively affect voucher sales. Economically, all else equal, a 10% increase in the existing sales, on average, is associated with 0.98 additional voucher sales in the next hour. Interestingly, a 10% increase in the total number of Facebook Likes on average leads to 1.24 additional voucher sales, whereas Twitter-mediated WOM has no effect on sales (perhaps due to the transient nature of tweets). We also find evidence of complementarities between herding and Facebook-mediated WOM, as they positively interact with each other to generate additional sales. To uncover the underlying mechanisms behind these findings, we compare experience goods -- for which product quality is relatively difficult to ascertain before consumption (such as cleaning services, massage) and search goods -- for which product quality is relatively easier to ascertain before consumption (such as shoes, clothing accessories) (Nelson 1974). We find that herding is more salient for experience goods than for search goods, while the effect of Facebook-mediated WOM does not significantly differ between the two product categories. These results show that our measured herding effect is less likely to come from alternative explanations, because we should expect a stronger effect of herding for experience goods than for search goods if signaling is the main mechanism behind the herding effect. Similarly, we can attribute the effect of Facebook-mediated WOM to advertising as opposed to signaling because increasing product awareness should benefit sales regardless of whether it is a search and an experience good. Because Facebook-mediated WOM does not affect sales differently for search goods and experience goods, it is likely that the dominant mechanism for social media WOM to affect sales is advertising. This also explains the complementarities between herding and social media WOM because signaling product quality (herding) can amplify the advertising effect of social media WOM. Our results are robust to multiple model specifications and several identification strategies.

2. Related Literature

Since daily deals have become a popular marketing tool, it is important to understand their underlying mechanism. Kumar and Rajan (2012) is one of the first to study daily deals and use a framework from the couponing literature to analyze the profitability of social coupons. There are also a growing number of working papers on daily deals (Byers et al. 2012, Dholakia 2011, Edelman, et al. 2010, Wu et al. 2013). Our work differs from these studies by examining the implications of two distinct mechanisms – herding and social media WOM – that differentiate daily deals from traditional marketing vehicles, such as coupons and rebates. Thus, our work contributes to the literature on both herding and online WOM.

The seminal theoretical work on herding shows that agents make decisions using their private but imperfect information as well as observing decisions that prior others have made (Banerjee 1992, Bikhchandani et al. 1992). When prior decisions are converged to a single choice, subsequent agents would simply follow the converged choice regardless of their own private information. A growing number of empirical papers have documented the herding effect. For example, in the context of micro-financing both Herzenstein et al. (2011) and Zhang and Liu (2012) find the evidence of herding among lenders on Prosper.com.

WOM, a well-established construct in the marketing literature, is shown to increase product awareness (Liu 2006). Trusov et al. (2009) examine the effect of WOM marketing on the growth of memberships at a social-networking site and compare its effect with traditional marketing vehicles. Aral and Walker (2011) use a randomized field experiment to empirically test the effectiveness of social media WOM on the adoption of free applications hosted on Facebook.com. Nevertheless, there is still scant empirical evidence for the effectiveness of social media WOM on actual sales. Using a natural experiment from an information policy change at Amazon, Chen et al. (2011) document the effects of observational learning and online WOM on the sales rank of music CD sold on Amazon. Our work differs from their study in two ways. First, our work focuses on the effect of social media WOM in which agents have some form of established social relationships, such as friends on Facebook or followers on Twitter. Because of the prior established social relationships, WOM messages can be pushed to consumers via the social media platforms. By contrast, Chen, et al. (2011) examine the effect of online WOM using anonymous reviewers on Amazon who do not have any prior social relationship with potential buyers and thus these WOM messages can only be pulled on the product page as opposed to having them pushed to buyers via social media. Therefore, social media WOM in our setting serves as an advertising vehicle in addition to helping consumers update their beliefs about product quality. The potential to increase the pool of potential buyers through advertising also distinguishes social media WOM from traditional online WOM such as Amazon's product reviews, where users must first become aware of the product before finding the reviews. This advertising effect is also a primary driver for the complementarity between herding and social media WOM. This result may be absent in traditional WOM studies. Second, instead of using Amazon sales rank as an approximation to actual sales by Chen, et al. (2011), we use the accurate sales data that allow us to more precisely quantify the effects of herding and social media WOM.

3. Theory

Since daily-deal sites provide an ideal setting for studying the implications of herding and social media WOM, we choose Groupon.com, perhaps the largest daily-deal site, as our research setting. Groupon features a single deal everyday on the main page of each local market. Figure 1 shows a screenshot of a typical feature deal. Shoppers can see the characteristics of the deal, such as a brief description, vendor, discounted voucher price and percentage. The total number of vouchers sold is prominently displayed in real-time, allowing shoppers to observe prior others' purchasing decisions. The Facebook "Like" and Twitter buttons are displayed below the sales information, allowing shoppers to share the deal with their Facebook friends and/or Twitter followers. While the number of tweets is unobservable on the deal page, the number of Facebook Likes is either unobservable or shown inconspicuously. When it is observable in some cases,[20] the number of Facebook Likes is always at the

[20] The number of Facebook Likes may be observable depending on shoppers' Web browsers, albeit in the smallest font possible at the bottom of the deal page. When it is displayed, it could potentially bias the estimation if one also

bottom of the deal page in a much smaller font than other information. Because people often neglect information that is not visually prominent (Nisbett and Ross 1980), we assume that the primary channel for Facebook Likes to affect future sales is through sharing the deal with Facebook friends as opposed to observing the actual number of Facebook Likes.

Figure 1: Screenshot of a typical deal featured by Groupon.com, in which the total number of voucher sales and the buttons for Facebook Likes and Twitter are circled

3.1 Herding

The economics literature on herding (Banerjee 1992, Bikhchandani, et al. 1992) suggests that people make decisions after internalizing others' observable behaviors. The herding effect is particularly salient under the condition of imperfect information when potential adopters are uncertain about the product and infer their own utility by observing others' prior decisions (Duan, et al. 2009). When previous decisions converge, consumers will follow the converged decision and disregard their own private information. Daily-deal sites facilitate the herding mechanism because they prominently display the cumulative sales of each deal in real-time (see Figure 1). Herding can provide a useful signal for

---

interprets the effect of Facebook Likes as a type of herding. Based on our trials, Internet Explorers does not display the number of Facebook Likes, and it has the largest market share in the browser market, accounting slightly over 40% (based on the statistics by September 2011, see http://www.tomshardware.com/news/browsers-ie-chrome-firefox-mozilla,14410.html). Thus, we could expect about 40% Groupon shoppers would not be able to observe the number of Facebook Likes. In such case, the effect of Facebook Likes is merely through social media WOM.

updating consumers' beliefs about the deal value. Intuitively, suppose there are two restaurant deals with identical characteristics, uninformed customers would expect the deal with higher existing sales to be more valuable than the other. Recent survey by Dholakia (2011) shows that more than 80% of the deal shoppers are new customers who are likely to be uninformed about the product and therefore more likely to herd when existing sales are high.

Interestingly, most of the deals on Groupon.com are for experience goods whose values are hard to infer before consumption (Nelson 1974). For experience goods (e.g., massage, spa, and cleaning services), observing existing sales is one of the few available signals for consumers to infer the product quality, making it important for influencing purchase decisions. By contrast, the value of search goods (tangible products, such as shoes, glasses, and clothing accessories) is relatively easier to ascertain before consumption. Since observing existing sales is just one of the many information signals that consumers can use to infer their utilities, its effect on future sales for search goods is less salient than it is for experienced goods. Given the majority of deals on daily deal sites are experienced goods, we expect herding to significantly affect sales.

Hypothesis (H1): *All else equal, a deal with more existing sales is likely to receive more additional sales in the next period.*

3.2 Social Media WOM

Word-of-mouth (WOM) refers to the dissemination of information from one person to another. It can be measured using volume, such as the total number of messages transmitted, or valence, such as the sentiment of the disseminated information. While the valence of WOM can affect sales through conveying positive or negative sentiments, the volume of WOM affects sales through increasing product awareness (Chen, et al. 2011). Although both volume and valence can influence product sales (Chevalier and Mayzlin 2006, Chintagunta et al. 2010), there is some evidence that WOM volume is more effective than valence (Liu 2006). In this study, we focus on the volume of WOM via social media platforms such as Facebook and Twitter for each featured deal.

Groupon provides the Facebook "Like" and Twitter buttons on the deal page (see Figure 1), allowing shoppers to share the deal with their Facebook friends and/or Twitter followers. For example, after clicking on the Facebook "Like" button, a user's Facebook wall will record this activity and is displayed immediately on the news feeds of his or her Facebook friends (see Figure 2). Upon seeing the deal information from the news feeds, these Facebook friends can view the Groupon deal by following the enclosed link. Twitter uses a similar mechanism to distribute deal information. We theorize that there are potentially two underlying mechanisms for social media WOM to generate voucher sales.

Figure 2: A Groupon deal is shared on Facebook news feed

Figure 3: A Groupon deal is shared on Twitter

The first mechanism is advertising, specifically, disseminating information about a product to a wider pool of potential buyers (informative role of advertising). Clicking on the Facebook "Like" button on a deal page can spread information about the deal to one's Facebook friends, increasing the awareness for the product. According to a recent statistics, an average Facebook user has 229 friends on Facebook,[21]

[21] See http://embracedisruption.com/2013/01/08/an-average-facebook-user-has-229-friends-100-social-stats-from-2012/

which suggests that on average a Facebook Like can potentially push a deal to 229 Facebook users. Upon seeing the deal, the informed users can choose to purchase the advertised deal. Extant research shows that traditional online WOM from anonymous strangers in the public (Chen, et al. 2011, Liu 2006) can affect various economic outcomes through advertising. Because the same piece of information gets more attention when it comes from friends than when it comes from strangers (Granovetter 1973), we believe that social media WOM can have a stronger effect in advertising than traditional online WOM that comes from strangers. Accordingly, we expect social media WOM can effectively spread information about a product and attract more potential buyers (Tucker 2012).

The second mechanism for social media WOM to generate sales is signaling. Extant research has shown that people enjoy an enhanced self-image or identity (Akerlof and Kranton 2000). To improve their image, consumers often choose to associate themselves with superior products and brands in front of others (Berger and Heath 2007). Using WOM as a self-enhancement tool (Wojnicki and Godes 2008), consumers tend to endorse superior products to their peers. Because consumers' propensity to click the Facebook "Like" button depends on their perception about the deal value (Li 2013), choosing to "Like" a product on Facebook can be viewed as a public social endorsement, signaling a favorable product quality to one's peers and thus increasing their propensity to buy. Because the endorsement comes from social connections on Facebook, the signaling effect through "liking" a deal on Facebook can be stronger than through traditional online endorsements such as voting up an online news article (Muchnik et al. 2013) or writing positive online reviews. Thus, we expect that Facebook Likes can affect voucher sales by signaling a favorable product quality to peers.

Twitter has similar mechanisms for simultaneously advertising and endorsing a deal to social contacts. Clicking on the Twitter button automatically pushes a tweet about the deal to one's followers (see Figure 3). Similar to Facebook Like, this allows Twitter users to advertise and endorse the deal to a broader base of followers, increasing the potential pool of buyers and their propensity to buy. Thus, we expect Twitter to also positively affect voucher sales.

Hypothesis (H2): *All else equal, a deal shared by more Facebook Likes and/or Twitter messages (tweets) receives more additional sales in the next period.*

3.3 Interaction between Herding and Social Media WOM

As discussed in Section 3.2, there are two underlying mechanisms—advertising and signaling—that can contribute to the overall effect of social media WOM on sales. In this section, we theorize the interaction effect between herding and social media WOM according to the two underlying mechanisms.

First, social media WOM can serve as an advertising vehicle to disseminate product information to users on social media platforms (Facebook and Twitter) and thus increase the awareness of the product. Upon seeing a deal through social media WOM, potential buyers can decide to make a purchase based on their expected valuation for the deal. Because daily deal shoppers are likely to be uninformed buyers (Dholakia 2011), external signals, such as existing sales, can significantly influence their perception of the deal value. When existing sales are low, it can negatively affect potential buyers' perception about the product and their propensity to buy. Thus, even if social media WOM is effective in enticing a large number of potential buyers, the return to advertising is diminished when low existing sales fail to convert a potential buyer to make the purchase. On the other hand, when many people have already bought the deal, consumers are more likely to be converted because high existing sales can signal high product quality. Essentially, acting as complements, the high existing sales amplify the advertising effectiveness of social media WOM in generating additional sales.

Second, social media WOM, such as Facebook Likes and tweets, can serve as public endorsements for one's social contacts (Li 2013) and provide an information signal to help improve consumers' perception about the product value (Wojnicki and Godes 2008). While herding gleans information by aggregating preferences from anonymous strangers, the information signal embedded in the social media WOM come from established social ties. Because social contacts are more likely to have similar tastes (homophily) or know about a person's idiosyncratic preferences, the information signal derived from social media WOM can be different from that derived from herding. However, quality signals from social contacts on social media platforms do not necessarily supersede signals provided through herding. Herding provides action-based information about the product, while social endorsements may come from individuals who never bought the product themselves. Thus, herding and social media WOM could complement each other as the former provides action-based information albeit from strangers while the latter provides endorsements from social ties. When potential buyers see endorsements for the product from friends via social media and also observe many prior others have bought the product, their beliefs in the endorsement would be reinforced by the high existing sales. Accordingly, the quality signal from the high existing sales amplifies the signaling effectiveness of social media WOM. In this case, we would expect herding and social media WOM to complement each other in affecting sales.

On the other hand, the tie strength on social media platforms, such as Facebook or Twitter, is often at best weak (Bapna et al. 2011). When they are sufficiently weak that these social contacts do not differ from online strangers in the public, the quality signals from herding and social media WOM about a product can actually be similar enough and thus redundant in that the two mechanisms may cannibalize each other. In this case, social media WOM and herding can actually substitute each other in affecting product sales.

Therefore, depending on the nature of the information social media WOM provides, whether herding and social media WOM are complements or substitutes in affecting product sales is an empirical question. Acting as an advertising vehicle, social media WOM complements herding to drive sales. As a way to provide an information signal about product quality, social media WOM and herding can also be complements if consumers interpret the signals they provide to be different and reinforce each other. However, if the strength of ties on social media platforms is sufficiently weak that they do not differ from anonymous strangers, social media WOM and herding are likely to provide redundant signals. In this case, they can also be substitutes.

Hypothesis (H3a): *All else equal, herding and social media WOM are complements in affecting sales.*

Hypothesis (H3b): *All else equal, herding and social media WOM are substitutes in affecting sales.*

4. Data and Empirical Methodology

4.1 Data Collection

All data in our study are collected from public sources by using Cameleon Web Wrapper (Firat et al. 2000). Specifically, data about deal characteristics and sales are extracted directly from Gropon.com. We use public APIs provided by Facebook and Twitter to extract the number of Facebook Likes and tweets that are associated with each deal. When available, we also gather product ratings from Yelp.com and Citysearch.com.

We sample 6 metropolitan areas in the US, including East Coast (Boston, New York City), Central (Chicago, Houston) and West Coast (Los Angeles, San Francisco) from July 1st to September 27th, 2011. Our data collection is discontinued on September 27, 2011, because Gropon.com stops displaying the accurate number of sales for each deal.22 Accordingly, the data set includes 526 featured deals in the 6 metropolitan areas. For each deal, we collect the number of voucher sales, the number of Facebook Likes and tweets hourly (from 1:00am to 11:59pm) during the first day when the deal is featured. Also, we collect the discounted voucher price, the original/face value, the product category, and the average product rating from Yelp/Citysearch. 26 (about 4.9%) out of the 526 deals have errors.

[22] See http://allthingsd.com/20111010/groupon-makes-it-less-possible-to-track-how-well-it-is-doing/

Results from the t-test and Chi-square test show the deals with errors are not systematically different from the rest of the sample. Therefore, we can safely remove the erroneous deals, resulting in an unbalanced panel data set consisting of 500 deals with at most 24 hourly periods.

4.2 Descriptive Statistics

Table 1 presents the descriptive statistics about various deal characteristics. In our data, 113 deals are related to restaurants and pubs, and 348 are other experience goods, such as spas, massage, and cleaning services. Overall, experience goods (including restaurants and pubs) account for 92.2% and the remaining 39 deals are for search goods (tangible products), such as shoes, glasses, and clothing accessories. On average, the discounted voucher price is $103.92 with an average discount rate of 57%. The average number of vouchers sold in the first day is 937, generating $97,338 in revenue for a typical featured Groupon deal. The average number of Facebook Likes and tweets in the first day are 106.08 and 10.46, respectively, and they are statistically different (t=9.43, p<0.001). This suggests that Groupon shoppers are more likely to share deals via Facebook than via Twitter. More than 80% of the deals have ratings from Yelp and/or Citysearch, with an average rating of 3.88 (sd=0.75) out of a maximal rating of 5.

Table 1. Descriptive Statistics

Variable	N	Mean	s.d.	Min	Max
Voucher price	500	103.92	349.39	2	2999
Original value	500	267.21	827.66	5	7900
Discount rate	500	57.05	10.89	33.33	95.00
Rating	410	3.88	0.75	1	5
Total voucher sales	468	936.66	1829.77	0	28569
Total Facebook Likes	468	106.08	217.27	0	2612
Total tweets	468	10.46	30.32	0	460

Notes: The descriptive statistics of total sales, total Facebook Likes and tweets are based on 468 deals for which the observations at the end of the day (11:59pm) are collected in the data set.

4.3 Estimation Specification

Given the panel structure of the data, we use a fixed-effect specification as the main model in the analysis. Because it can eliminate any time-invariant unobserved heterogeneity, fixed-effect specifications have been used to identify the effect of herding in various empirical studies (Duan, et al. 2009, Zhang and Liu 2012).

We denote the natural log of cumulative sales of a deal i up to the t^{th} hour by $Y_{i,t}$, t=1,2,…,24. As the cumulative sales are explicitly highlighted in real-time on Groupon.com (see Figure 1), the one-hour lagged cumulative sales $Y_{i,t-1}$ reflects the aggregate purchases before the t^{th} hour and thus can be used to operationalize and measure the herding effect. $y_{i,t}$ is the incremental sales occurring during the t^{th} hour. According to the estimation specification suggested by Zhang and Liu (2012), the herding effect can be identified by the coefficient of $Y_{i,t-1}$ on $y_{i,t}$ after controlling for deal-specific heterogeneity and other time-varying variables. We also control for nonlinear time trends in a later specification. To control for other time-varying confounding factors, we also included several robustness checks in later sections.

The natural log of cumulative number of Facebook Likes or tweets associated with deal i up to the t^{th} hour is denoted by $FB_{i,t}$, $TW_{i,t}$, respectively. We use one-hour lag of the cumulative Facebook Likes and tweets ($FB_{i,t-1}$, $TW_{i,t-1}$) in the estimation to avoid the potential endogeneity, because there may be some confounding factors that simultaneously drive Facebook Likes (or tweets) and sales in the same period. We use their natural log transformations for the three key independent variables in all the estimation specifications,[23] because they are heavily skewed (see Table 1).

$$y_{i,t} = \alpha + \beta_1 Y_{i,t-1} + \beta_2 FB_{i,t-1} + \beta_3 TW_{i,t-1} + \mu_i + v_t + \epsilon_{i,t} \tag{1}$$

In Equation (1), μ_i controls for deal-specific time-invariant heterogeneity, including any observable and unobservable time-invariant deal characteristics, such as voucher price, quality of the good, and the location for the voucher redemption. In particular, the unobservable quality about the deal is likely to be time-invariant during the 24-hour period and thus can be captured by μ_i. We use v_t to control for common shocks at different hours over a day. For example, suppose consumers' online shopping behaviors, on average, are more active in the later afternoon than in the early morning, the hour dummies v_t control for the common time shocks throughout the whole day. $\epsilon_{i,t}$ is the unobserved disturbance term which is assumed to be orthogonal to other independent variables. Based on this assumption, Equation (1) estimates the effects of herding (β_1) and WOM mediated via Facebook and Twitter (β_2, β_3) using the within-deal variance. Table 2 presents the overall mean and standard deviation of the time-varying variables in the data set.

[23] In the data set, some deals have zero cumulative sales, Facebook Likes or tweets at certain time points. To include those observations, we add 0.5 to the cumulative amounts before taking the natural log transformations for the three key independent variables.

Table 2. Pearson Correlation among Time-Varying Variables

Variable	Mean	s.d.	1	2	3
1. Incremental sales: $y_{ij,t}$	41.46	117.84	1		
2. Log of past cumulative sales: $Y_{ij,t-1}$	4.29	2.29	0.345	1	
3. Log of past cumulative Facebook Likes: $FB_{ij,t-1}$	2.30	1.93	0.274	0.811	1
4. Log of past cumulative tweets: $TW_{ij,t-1}$	1.41	0.87	0.175	0.472	0.514

Notes: The means, standard deviations (s.d.) and Pearson correlations are based on the pooled including 10550 observations in the data.

The literature on product diffusion and social contagion suggests that people adopt when they come in contact with others who have already adopted, spreading like epidemics (Young 2009). The adoption rate increases as the user base grows but may decrease when the product starts saturating the market. This non-linear shape for sales over time may cause a spurious relationship between cumulative past sales and the current sales. Therefore, it is important to control for product diffusion when measuring the effect of herding. The extant literature (Carare 2012, Duan, et al. 2009) suggests that adding the linear and quadratic terms of product age into the estimation specification can address the issue of non-linear shape of sales over time, such as from product diffusion. In our context, the age of the deal is operationalized as the number of hours that have passed since the deal is featured. To maintain enough degrees of freedom for the estimation, we adopt an approach, suggested by Duan, et al. (2009), that allows the coefficients of the deal age and its quadratic term to vary across different cities but remain constant within the same city. Accordingly, in our study we use an enhanced estimation specification, Equation (2), in which *j* is the index for the 6 metropolitan areas, *t* is the number of hours as the deal age, and γ_{j1}, γ_{j2} are the coefficients of linear and quadratic deal ages, controlling for city-specific product diffusion pattern and common time trends.

$$y_{ij,t} = \alpha + \beta_1 Y_{ij,t-1} + \beta_2 FB_{ij,t-1} + \beta_3 TW_{ij,t-1} + \gamma_{j1} t + \gamma_{j2} t^2 + \mu_i + v_t + \epsilon_{ij,t} \quad (2)$$

5. Results

We first use the fixed-effect estimation with robust standard errors clustered at the deal level to estimate the main effects of herding and social media WOM.[24] We then compare these estimates between search goods and experience goods. If the herding effect is real, we would expect it to be stronger for experience goods than for search goods. We would also expect a similar finding for social media WOM if the primary mechanism behind it is signaling. Next, we explore the complementarities between herding and social media WOM. In Section 5.4, we will use dynamic Generalized Methods of Moments (GMM) in the robustness check.

5.1 Effects of Herding and Social Media WOM

Table 3 reports the results from the fixed-effect estimation. While Columns (1)-(4) are estimated using Equation (1), Column (5) is estimated using Equation (2) which controls the non-linear shape of sales over time by including city-specific linear and quadratic time trends. The coefficients across the columns are fairly stable and the variance inflation factors (VIF's) are reasonably small, suggesting that multicollinearity is not an issue. According to Column (5) of Table 3, the cumulative sales are positively associated with future sales in the next hour, suggesting a positive herding effect on sales. Similarly, the cumulative number of Facebook Likes is positively associated with future sales in the next hour, suggesting a positive effect of Facebook-mediated WOM on sales. Economically, all else equal, a 10% increase in the existing sales, on average, is associated with 0.98 additional voucher sales in the next hour, and a 10% increase in the total number of Facebook Likes on average leads to 1.24 additional voucher sales in the next hour. Column (6) reports the estimated standardized coefficients and reveals that the magnitudes of the effects of herding and Facebook-mediated WOM are fairly comparable.

Interestingly, while the effect of Facebook-mediated WOM on sales is significantly positive, the estimated coefficient of cumulative tweets is far from being statistically significant. This suggests that Twitter-mediated WOM has minimal impact on sales, perhaps due to the transient nature of tweets. The fact that only Facebook Likes but not tweets matters for future sales indicates that the effect of Facebook-mediated WOM on sales is likely to be causal, rather than being confounded by some omitted factors. Suppose some confounding factors could simultaneously affect voucher sales and encourage consumers to share Groupon deals on social media, we expect them to bias our estimations for Facebook Likes and tweets similarly. Since only Facebook Likes affects sales while tweets do not, we can attribute the sales effect to come from Facebook-mediated WOM as opposed to the unobserved confounding factors.

[24] Allowing for any arbitrage form of serial correlation, robust standard errors clustered at the panel level consistently converge to the true standard errors, as the number of clusters approaches infinity (Wooldridge, 2010).

Table 3. Fixed-Effect Estimation with Clustered Standard Errors

	(1)	(2)	(3)	(4)	(5)	(6)
Past cumulative sales: $Y_{ij,t-1}$	12.92***			8.90***	9.80***	22.42***
	(2.85)			(2.42)	(2.50)	(5.72)
Past cumulative Facebook Likes: $FB_{ij,t-1}$		15.28***		12.71***	12.41***	23.96***
		(2.93)		(2.47)	(2.50)	(4.83)
Past cumulative tweets: $TW_{ij,t-1}$			11.17	8.74	8.92	7.78
			(7.58)	(7.45)	(7.98)	(6.96)
Deal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	10550	10550	10550	10550	10550	10550
Number of clusters	500	500	500	500	500	500

Notes: Dependent variable is hourly voucher sales $y_{ij,t}$. All standard errors are clustered at the deal level and reported in parentheses. Columns (1)-(4) are estimated using Equation (1). Column (5) is estimated using Equation (2) which controls the product diffusion process by including city-specific linear and quadratic time trends. Column (6) reports the standardized coefficients estimated using Equation (2). *p < 0.10, **p < 0.05, ***p < 0.01

5.2 Differential Effects for Search Goods and Experience Goods

One necessary condition for herding to create information cascade is that consumers have imperfect information about the product value when they make a decision (Banerjee 1992, Bikhchandani, et al. 1992). After all, if consumers have perfect information about the product value, observing prior others' decisions would not increase their information sets. At best, herding provides a rough information signal for product quality as it only shows prior others' decisions but not the private information they use to make the decisions. Thus, herding is only a useful mechanism when quality information about the product is difficult to obtain. Since Groupon shoppers are largely new customers (Dholakia 2011), they are likely to have imperfect information about the deals, especially for experience goods (Nelson 1974) whose true values are difficult to ascertain before consumption. For experience goods such as restaurants, spa, massage, and cleaning services, herding is an important mechanism for consumers to infer the product value and make purchase decisions. By contrast, for search goods (Nelson 1974) whose values are relatively easier to ascertain before consumption, potential buyers have more alternative cues to infer the product value and would not need to rely on observing prior other's purchases. Information derived from herding would then play a less important role in helping consumers update their beliefs about the product. Thus, we expect the herding effect to be more salient for experience goods than for search goods. If we find that the effects between search goods and experience goods are similar, it is likely that our estimates on herding are biased. Therefore, comparing the effect of herding between search goods and experience goods can allow us to ascertain whether herding is truly driving sales.

As discussed in Section 3.2, social media WOM can affect sales through both advertising and signaling. Social media WOM can increase product awareness through advertising, and in the meanwhile, social media WOM can signal product quality through social endorsement and thus improve consumers' beliefs about the product. We expect the advertising effect to be similar between search goods and experience goods since improving product awareness should not depend on the product's inherent characteristic. However, we expect the signaling effect of social media WOM to be more salient for experience goods than for search goods. Because it is difficult to infer the product quality of experience goods, any additional signal, such as ones derived from social media WOM, can be useful for consumers to make a purchase decision. Therefore, if we observe a stronger effect of social media WOM for experience goods than for search goods, it is likely that signaling is the primary mechanism behind social media WOM to affect sales. By contrast, if we do not observe a difference between the two product categories, the positive effect of social media WOM is likely to come from advertising as opposed to signaling.

To examine the above propositions, we classify all the deals in our data set into the two categories: search goods and experience goods[25], and estimate the effects using the main specification Equation (2) that includes nonlinear time trends. Table 4 reports the results. Comparing the estimates in Columns (1) and (2), we find that the estimated coefficient of past cumulative sales for experience goods is positive and statistically significant, whereas the corresponding estimate for search goods is much smaller. One possible alternative explanation for the difference in the estimates between experience goods and search goods is that experience goods could be more popular. Since the inherent popularity of a product is unlikely to change over a 24-hour period, our fixed-effect model has already controlled for this variation. Furthermore, we compare the average final sales of experience goods and search goods and find the difference in the average final sales of the two categories is not statistically significant (955.9 vs. 718.7, p=0.44). The difference in the median of the final sales is even closer (488 *vs*. 435). This comparison rules out the possibility that inherent product characteristics could explain the difference in the estimated coefficients in Columns (1) and (2).

We notice that there are only 39 search goods in our data set as shown in Column (1) of Table 4, while 461 experience goods are analyzed in Column (2). To address the difference in the sample size of the two categories, we randomly select 39 experience goods from all the 461 experience goods and estimate the effects using the random subsample.[26] As shown in Column (3), the estimated coefficient of past cumulative sales for experience goods is again much larger and more statistically significant than the corresponding estimates for search goods. To test that the estimate for experience goods is indeed greater than for search goods, we add interaction effects between the product category and herding and between product category and social media WOM. Results in Columns (4)-(6) shows that the difference in the estimated associations between $Y_{ij,t-1}$ and $y_{ij,t}$ for experience goods and search goods is statistically significant while we do not find a significant difference for Facebook Likes and tweets. These comparison results together enhance our confidence that the estimated positive association between the cumulative sales and future sales results from herding. If some other mechanism were the primary driver for the results, it is unlikely that they affect experience goods and search goods differently.

[25] Although the dichotomous classification of search vs. experience goods is well recognized in the literature since the pioneering work by Nelson (1974), we note that most products can be viewed as a collection of search attributes and experience attributes. To measure the differential effects, we only classify tangible products as search goods (including shoes, glasses, and clothing accessories) and we assume the values of tangible products are relatively easier to ascertain before consumption than the rest in the other category (experience goods). We also use some other methods to classify search and experience goods and get similar results.

[26] We draw a number of different random subsamples of 39 experience goods and the results are qualitatively similar.

Table 4. Differential Effects for Search Goods vs. Experience Goods

	(1) Search Goods	(2) Experience Goods	(3) Random Sample of Experience Goods	(4)	(5)	(6)
Past cumulative sales: $Y_{ij,t-1}$	6.73*	10.38***	47.35**	19.27**	21.53**	21.04**
	(3.65)	(2.79)	(22.22)	(9.12)	(9.80)	(9.24)
Past cumulative Facebook Likes: $FB_{ij,t-1}$	7.14	12.80***	18.33	14.11*	13.54*	13.68*
	(5.10)	(2.67)	(14.22)	(7.30)	(7.31)	(7.50)
Past cumulative tweets: $TW_{ij,t-1}$	0.13	9.67	-14.87	-8.80	-8.18	-7.03
	(4.41)	(8.63)	(12.03)	(9.29)	(9.04)	(7.51)
$Y_{ij,t-1} \times$ Prod_dum				-5.86	-17.10*	-18.05*
				(3.68)	(9.10)	(9.92)
$FB_{ij,t-1} \times$ Prod_dum					13.53	12.43
					(8.32)	(8.13)
$TW_{ij,t-1} \times$ Prod_dum						7.10
						(13.08)
Deal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	811	9739	727	1538	1538	1538
Number of clusters	39	461	39	78	78	78

Notes: Dependent variable is hourly voucher sales $y_{ij,t}$. All standard errors are clustered at the deal level and reported in parentheses. Columns (1) and (2) are estimated based on all search goods and experience goods in the data set, respectively. Column (3) is estimated based on a random sample of 39 experience goods. Columns (4)-(6) compare the effects of herding and social media WOM for the 39 search goods and the random sample of 39 experience goods. All columns are estimated using Equation (2) which controls the product diffusion process by including city-specific linear and quadratic time trends. *p < 0.10, **p < 0.05, ***p < 0.01

We do not find any evidence that social media WOM to affect search goods and experience goods differently, as shown in Columns (5) and (6). This suggests that the primary driver for Facebook-mediated WOM to affect sales is advertising, rather than signaling. If signaling product quality were the primary mechanism for Facebook-mediated WOM to affect sales, we would have detected a significant difference in its effect between experience goods and search goods, just as what we have found for herding. On the other hand, there is no reason to expect that increasing product awareness through advertising should affect sales differently between search goods and experience goods.

5.3 Results of Complementarity between Herding and Facebook-mediated WOM

To explore the interaction effect between herding and social media WOM, we include the interaction terms between the past cumulative sales and Facebook Likes and between past cumulative sales and tweets, as shown in Equation (3).

$$y_{ij,t} = \alpha + \beta_1 Y_{ij,t-1} + \beta_2 FB_{ij,t-1} + \beta_3 TW_{ij,t-1} + \gamma_{j1} t + \gamma_{j2} t^2$$
$$+ \eta_1 Y_{ij,t-1} \times FB_{ij,t-1} + \eta_2 Y_{ij,t-1} \times TW_{ij,t-1} + \mu_i + v_t + \epsilon_{ij,t} \quad (3)$$

To reduce multicollinearity, we de-meaned the variables in the interaction terms. The uncentered variance inflation factors (VIF's) of all the key independent variables are below the critical values, indicating multicollinearity is not an issue. The estimates are reported in Table 5 and Column (1) is reproduced as in Column (5) of Table 3 for readers' convenience. As shown in Columns (2)-(4), the interaction terms between the past cumulative sales and Facebook Likes are positive and statistically significant, after controlling for deal-specific heterogeneity, common time shocks, and linear and nonlinear time trends at the city level. This result supports that Facebook-mediated WOM can positively interact with herding in driving future sales, while all the estimated coefficients related to Twitter-mediated WOM are statistically insignificant. The absence of any effect from Twitter is likely due to its transient nature that is less likely to have a lasting effect on sales. The lack of complementarities between Twitter and herding can in fact enhance the evidence of complementarities between Facebook-mediated WOM and herding. If some confounding factors were to drive both sales and the complementarities between Facebook-mediated WOM and herding, we would expect these factors to similarly affect the interaction between Twitter-mediated WOM and herding. Because it is difficult to envision a scenario where a confounding factor only affects one social media channel and its interaction effect with herding but not the other, we can attribute the positive interaction between Facebook-mediated WOM and herding to be a true complement in driving sales.

As we theorize in Section 3.3, if the primary mechanism for social media WOM to affect sales is advertising that increases product awareness, we would expect more people to arrive at Groupon's landing page because of the advertising effect of social media WOM. These consumers are more likely to buy if the product has high existing sales. Hence, quality signals derived from high existing sales can reinforce the advertising effect on social media to convert these consumers into buying. Therefore, the evidence of complementarities between Facebook-mediated WOM and herding is also consistent with our findings that the overall effect of Facebook-mediated WOM is primarily through advertising.

Table 5. Interaction Effects between Herding and Social Media WOM

	(1)	(2)	(3)	(4)
Past cumulative sales: $Y_{ij,t-1}$	9.80***	15.97***	10.52***	16.17***
	(2.50)	(2.95)	(2.37)	(3.05)
Past cumulative Facebook Likes: $FB_{ij,t-1}$	12.41***	8.48***	11.66***	8.65***
	(2.50)	(1.99)	(2.47)	(2.06)
Past cumulative tweets: $TW_{ij,t-1}$	8.92	-1.55	6.83	-1.01
	(7.98)	(6.25)	(6.88)	(5.84)
$Y_{ij,t-1} \times FB_{ij,t-1}$		4.46***		5.14***
		(0.89)		(1.03)
$Y_{ij,t-1} \times TW_{ij,t-1}$			2.46	-2.52
			(1.57)	(1.58)
Deal fixed effects	Yes	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes	Yes
Number of observations	10550	10550	10550	10550
Number of clusters	500	500	500	500

Notes: Dependent variable is hourly voucher sales $y_{ij,t}$. Variables in the interaction terms are centered (de-meaned). All standard errors are clustered at the deal level and reported in parentheses.
*p < 0.10, **p < 0.05, ***p < 0.01

5.4 Robustness Checks

In this section, we conduct a set of robustness checks to verify our findings. First, we use a dynamic GMM framework to further explore the causality of our findings. Then, we use revenue instead of number of products sold as an alternative dependent variable.

5.4.1 Dynamic GMM

While fixed-effect estimation is more efficient under the condition that the disturbance terms $\epsilon_{ij,t}$ are serially uncorrelated, the first-differencing estimation is more efficient when $\epsilon_{ij,t}$ follow a random walk (Wooldridge 2010). Wooldridge (2010, pp. 321) notes that "in many cases, the truth is likely to be lie somewhere in between." Considering there are 24 time periods in the data, first-differencing estimation is necessary because assumptions under fixed-effect estimation are sensitive to violation when the number of time periods is large. The first-differencing estimation specification, corresponding to Equation (2), is:

$$\Delta y_{ij,t} = \beta_1 \Delta Y_{ij,t-1} + \beta_2 \Delta FB_{ij,t-1} + \beta_3 \Delta TW_{ij,t-1} + \gamma_{j1} + \gamma_{j2} \Delta t^2 + \Delta v_t + \Delta \epsilon_{ij,t} \quad (4)$$

Note that the deal fixed effect μ_i disappears in Equation (4). First-differencing estimation requires a different assumption of strict exogeneity (i.e., the first-differencing variables are strictly exogenous) and the corresponding disturbance terms $\Delta\epsilon_{ij,t}$ are serially uncorrelated. However, in Equation (4) the explanatory variable $\Delta Y_{ij,t-1} = Y_{ij,t-1} - Y_{ij,t-2}$ is endogenous because the dependent variable is $\Delta y_{ij,t} = y_{ij,t} - y_{ij,t-1}$. Hence, the estimated coefficient of $\Delta Y_{ij,t-1}$ may be biased due to the potential endogeneity issue. To address this concern, we use the dynamic Generalized Methods of Moments (GMM) to estimate Equation (4).

Specifically, we use Arellano-Bond/Blundell-Bover two-step robust system GMM estimation with orthogonal deviations. This estimation method instruments the lagged dependent variables and other endogenous variables using their second- or higher-order lags, while addressing the fixed effects using first-differencing. Two-step robust system GMM estimation with corrected standard errors is more efficient than difference GMM and using orthogonal deviations can deal with the unbalanced panel data (Arellano and Bover 1995, Blundell and Bond 1998). Considering T=24 is relatively large in our data set, we choose to use their 9th-order and deeper lags of $Y_{ij,t-1}$ as instruments,[27] because deeper lags are more likely to satisfy the IV assumptions of relevancy and exogeneity. We also treat $FB_{ij,t-1}$ and $TW_{ij,t-1}$ as endogenous variables and use their lags as instruments so that the potential endogeneity between $FB_{ij,t-1}$, $TW_{ij,t-1}$ and $y_{ij,t}$ can be addressed. The interaction terms are similarly instrumented with the corresponding lags.

Table 6 reports the system GMM estimates. In Column (1), the Arellano-Bond test for AR(2) in first differences cannot reject the null that there is no second-order serial correlation in the residuals of the first-differencing equation (p=0.32). Thus, serial correlation is not an issue in the GMM estimation. Neither Hansen J statistic (over-identification test) (p=0.14) nor difference-in-difference Hansen test (p=0.66) rejects the null that the instruments are uncorrelated with the disturbance terms, ensuring the validity of the instruments used in the GMM estimation (Roodman 2007). All these post-estimation diagnostics satisfy the criteria of system GMM estimation (Roodman 2007), indicating that the set of

[27] We also use other sets of lags and get qualitatively similar results.

instruments used in the analysis is valid. As we can see, the estimates in Column (1) of Table 6 are qualitatively similar to Column (5) of Table 3, although the magnitudes of the point estimates are different due to the different estimation specifications. Overall, the results in Column (1) of Table 6 support that both the effects of herding and Facebook-mediated WOM are positive and statistically significant while Twitter-mediated WOM is not.

Table 6. Robustness Checks using Dynamic GMM

	(1)	(2)	(3)	(4)
Past cumulative sales: $Y_{ij,t-1}$	21.31***	13.77***	19.14***	9.98***
	(8.21)	(3.59)	(7.44)	(2.80)
Past cumulative Facebook Likes: $FB_{ij,t-1}$	12.07***	-3.12	5.99**	-9.44
	(3.26)	(3.04)	(2.68)	(6.42)
Past cumulative tweets: $TW_{ij,t-1}$	8.45	10.43***	16.84**	14.75**
	(5.67)	(3.84)	(6.85)	(6.93)
$Y_{ij,t-1} \times FB_{ij,t-1}$		7.56***		11.57***
		(2.35)		(3.63)
$Y_{ij,t-1} \times TW_{ij,t-1}$			2.02	-2.91
			(2.89)	(3.12)
Deal fixed effects	Yes	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes	Yes
Number of instruments	390	393	393	483
Number of observations	10550	10550	10550	10550
Number of clusters	500	500	500	500

Notes: Dependent variable is hourly voucher sales $y_{ij,t}$. Variables in the interaction terms are centered (de-meaned). All standard errors are clustered at the deal level and reported in parentheses. Results are estimated using Arellano-Bond/Blundell-Bover two-step robust system GMM estimation with orthogonal deviations. *p < 0.10, **p < 0.05, ***p < 0.01

Column (2)-(4) of Table 6 report the estimates of the interaction effects between past cumulative sales and social media WOM. Again, all the post-estimation diagnostics satisfy the criteria of system GMM estimation. Take Column (4) as an example: the Arellano-Bond test for AR(2) in first differences cannot reject the null that there is no second-order serial correlation in the residuals of the first-differencing equation (p=0.36). Thus, serial correlation is not an issue in the GMM estimation. Neither Hansen J statistic (over-identification test) (p=0.25) nor difference-in-difference Hansen test (p=0.97) rejects the null that the instruments are uncorrelated with the disturbance terms. Therefore, these post-estimation diagnostics reveal that the set of instruments used in Column (4) are valid. According to Column (4), the interaction between past cumulative sales and Facebook Likes is positive and statistically significant, suggesting that herding and Facebook-mediated WOM are complements in affecting sales. Similar to Table 4, the interaction with Twitter-medicated WOM is not significantly different from zero, suggesting the positive interaction between Facebook-mediated WOM and herding is the indicative of true complements.

5.4.2 Revenue as Dependent Variable

In the main analysis based on Equation (2), we use the incremental sales occurring during the t^{th} hour, $y_{ij,t}$, as the dependent variable. Alternatively, we can use the incremental revenue $revenue_{ij,t}$ as the dependent variable, where $revenue_{ij,t} = y_{ij,t} \times voucherprice_i$. Using $revenue_{ij,t}$ as the dependent variable is another set of robustness checks and allows us to directly quantify the impacts of herding and social media WOM on financial metrics.

We report the results in Table 7. Column (1) shows that both past cumulative sales and Facebook Likes are positively associated with additional revenues in the next hour, suggesting both herding and Facebook-mediated WOM drive future revenues. Economically, all else equal, a 10% increase in the existing sales, on average, is associated with an increase in revenue of $30.5 in the next hour, and a 10% increase in the number of Facebook Likes on average leads to an increase in revenue of $19.1 in the next hour, after controlling for deal-specific heterogeneity, common time shocks and the nonlinear shape of sales over time. Yet, Twitter-mediated WOM has no statistically significant impact in affecting revenue. Columns (2)-(4) show that Facebook-mediated WOM positively interact with the herding effect in driving future revenues, while all the estimated coefficients related to Twitter-mediated WOM are insignificant. In general, the results in Table 7 based on using incremental revenues as an alternative dependent variable are consistent with our findings reported in Tables 5 and 6 that use sales as the dependent variable.

Table 7. Robustness Checks using Revenue as Dependent Variable

	(1)	(2)	(3)	(4)
Past cumulative sales: $Y_{ij,t-1}$	304.87***	353.90***	307.03***	356.81***
	(57.62)	(64.13)	(57.33)	(65.13)
Past cumulative Facebook Likes: $FB_{ij,t-1}$	191.05***	159.84***	188.81***	162.23***
	(52.20)	(53.79)	(53.18)	(54.40)
Past cumulative tweets: $TW_{ij,t-1}$	22.51	-60.79	16.25	-53.00
	(83.98)	(81.16)	(76.99)	(76.18)
$Y_{ij,t-1} \times FB_{ij,t-1}$		35.50***		45.37***
		(12.21)		(14.71)
$Y_{ij,t-1} \times TW_{ij,t-1}$			7.39	-36.55
			(19.08)	(21.90)
Deal fixed effects	Yes	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes	Yes
Number of observations	10550	10550	10550	10550
Number of clusters	500	500	500	500

Notes: Dependent variable is hourly incremetanl revenues $revenue_{ij,t} = y_{ij,t} \times voucherprice_i$. Variables in the interaction terms are centered (de-meaned). All standard errors are clustered at the deal level and reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

6. Discussion for Alternative Explanations

6.1 Ruling out Alternative Explanations for Herding

Besides controlling for deal-specific time-invariant heterogeneity, common time shocks at different hours over a day, and nonlinear shapes in sales over time, we need to rule out several alternative explanations so that the coefficient of past cumulative sales ($Y_{ij,t-1}$) on incremental sales ($y_{ij,t}$) can be interpreted as herding.

First, there is *social pressure* that induces adoptions so people conform to their peers (Young 2009). While this effect would have been a significant concern if peers can easily observe the adoption choice (e.g., fashion items), the deals in our context are largely personal experience/service goods, such as meals at restaurants, spas, massages, cleaning services. Because they are highly personalized and less observable or verifiable, social pressure is less likely to explain the herding behavior.

Network effects or payoff externalities could also influence sales because the value of a product and thus the likelihood of sales increases as more people are using it (Katz and Shapiro 1994). While network effects often occur with IT products (Brynjolfsson and Kemerer 1996), such as fax machines or microcomputers, they are less plausible for personal experience/service goods, because one's consumption does not directly increases others' utility of the goods. In fact, the opposite can happen for service goods due to capacity constraint; people may assume that the quality of the service goods would suffer when many vouchers have already been sold. Customers are likely to infer that the venue would be too crowded, especially when the expiration date is in the near future.

The third alternative explanation is *saliency* effect: when consumers are not aware of their entire choice sets, they tend to choose products that are prominently displayed (Cai, et al. 2009). Instead of herding, people may follow others' choice simply because the more salient or noticeable product is more likely to enter into their choice sets. Cai, et al. (2009) point out that saliency effect often confounds the empirical test for herding. For instance, when software on CNET.com are sorted by the number of downloads, the software with more downloads become more prominently displayed on the website. As a result, the herding effect estimated based on the software's downloading rank may be partly explained by the saliency effect. We purposefully collect only featured deals such that they are placed at the same location on Groupon's webpage. Thus, the saliency effect is reasonably controlled in our study.

Lastly, we expect the effect of herding to be strongest when the value of the good is hard to ascertain before consumption. In such case, the quality signal derived from herding is most useful to a potential buyer. If the effect of herding is similar across all goods, it is possible that unobserved heterogeneity is simultaneously driving existing sales and the sales in the next period, rendering the estimated effect of herding to be spurious. To rule out this possibility, we compare the effect of herding between experienced goods and search goods and show that the effect of herding is stronger for experienced goods whose values are relatively harder to ascertain before consumption than that for search goods.

6.2 Ruling out Alternative Explanations for Facebook-mediated WOM

Other online or offline promotions (such as TV advertising) could simultaneously affect voucher sales and Groupon shoppers' behaviors on social media platforms (Facebook Likes). Without ruling out this possibility, our estimation of the effect of Facebook-mediated WOM on sales may be biased. To address this alternative explanation, we first include time dummies in the estimation specifications to control for the effect of hourly time shocks. If the other online or offline promotions allow Groupon shoppers' purchasing behaviors to be more active in the late afternoon than in the morning, time dummies can reasonably control this confounding effect. Second, we further include linear and quadratic time trends that control for the possible effects that result from other promotions over the 24-hour period. Finally, while the effect of Facebook-mediated WOM on sales is significantly positive, we note that the estimated effect of Twitter-mediated WOM is far from statistically significant. Suppose the other online or offline promotions simultaneously affect voucher sales and also encourage consumers to share Groupon deals via social media, we expect them to bias our estimations of the effects of Facebook Likes and tweets similarly. However, we find that Facebook Likes are positively associated with future sales while Tweets are not. This suggests that the effect we find for Facebook-mediated WOM is likely to be causal.

7. Implications and Conclusion

This study yields several noteworthy implications for theory on herding and social media WOM. First, perhaps due to limited data availability, prior empirical works have measured either the effects of herding or online WOM, but not both; only recently a few exceptions have appeared in the literature (Chen, et al. 2011, Christy et al. 2012). Given the unique context of daily-deal sites, we collect accurate data of voucher sales, Facebook Likes and Twitter messages. Accordingly, we are able to precisely quantify the financial impacts of both herding and social media WOM using data from a real business setting. It is worth noting that although in theory social media WOM should have impacts on product sales (as we discussed in Section 3.2), a recent experimental study conducted by John et al. (2013) shows that when a person "likes" a brand, it has no impact on the person's own marketing outcomes that range from attitudes and word of mouth to advertisement choice and actual purchase. Our work complements their study by documenting significant effect of Facebook Likes on one's peers. That is, Facebook-mediated WOM can serve as an advertising tool to one's social circles, enlarging the pool of potential buyers and consequently generating product sales.

More importantly, we theorize and provide consistent empirical evidence that herding and Facebook-mediated WOM can positively interact with each other in driving product sales. Because social media WOM has the dual role of advertising the product through social media and improving consumers' perception about product quality, the findings of complementarity between herding and Facebook-mediated WOM suggest that perhaps Facebook Likes provide an informational signal about the product quality that is different from what is provided in herding. We also find the advertising role of social media WOM that increases the product awareness can drive the complementarities with herding.

Third, we find the effect of herding is much larger for experience goods than for search goods. The comparison supports the theoretical implication that the mechanism of herding effect is through signaling product quality when consumers have difficulty to infer the product quality prior to consumption (this is especially true in case of experience goods). Accordingly, the findings that the estimated association between cumulative sales and future sales is significantly larger for experience goods than search goods enhance our confidence that the effects we find can be interpreted as herding. We also find the effect of Facebook-mediated WOM does not significantly differ between experience goods and search goods, suggesting that advertising, as opposed to signaling, is the primary mechanism behind Facebook-mediated WOM in affecting sales. This finding also supports the complementarity results between herding and Facebook-mediated WOM. As more consumers become aware of the product through Facebook, they are more likely to buy when existing sales of the product are already high.

Finally, this study shows that appropriately implementing herding and social media WOM together could allow firms to achieve a better marketing outcome than implementing one or the other, because the two mechanisms are complements in generating additional product sales. The findings support the current practice of daily-deal sites where both mechanisms are implemented together. Besides the daily-deal sites, we expect online marketing firms in other industries could also recognize the complementarity between the two mechanisms and improve their marketing strategies by implementing both together.

References

Akerlof, G.A., R.E. Kranton. 2000. Economics and identity. *The Quarterly Journal of Economics* **115**(3) 715-753.

Aral, S., D. Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* **57**(9) 1623-1639.

Arellano, M., O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of econometrics* **68**(1) 29-51.

Banerjee, A.V. 1992. A Simple model of herd behavior. *The Quarterly Journal of Economics* **107**(3) 797-817.

Bapna, R., A. Gupta, S. Rice, A. Sundararajan. 2011. *Trust, Reciprocity and the Strength of Social Ties: An Online Social Network based Field Experiment*. Shanghai, China.

Berger, J., C. Heath. 2007. Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research* **34**(2) 121-134.

Bikhchandani, S., D. Hirshleifer, I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* **100**(5) 992-1026.

Blundell, R., S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics* **87**(1) 115-143.

Brynjolfsson, E., C.F. Kemerer. 1996. Network externalities in microcomputer software: An econometric analysis of the spreadsheet market. *Management science* **42**(12) 1627-1647.

Byers, J.W., M. Mitzenmacher, G. Zervas. 2012. *The Groupon Effect on Yelp Ratings: A Root Cause Analysis*.

Cai, H., Y. Chen, H. Fang. 2009. Observational learning: Evidence from a randomized natural field experiment. *American Economics Review* **99**(3) 864-882.

Carare, O. 2012. The Impact of Bestseller Rank on Demand: Evidence from the App Market. *International Economic Review* **53**(3) 717-742.

Chen, H., P. De, Y.J. Hu. 2011. *IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales*.

Chen, Y., Q. Wang, J. Xie. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research* **48**(2) 238-254.

Chevalier, J.A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345-354.

Chintagunta, P.K., S. Gopinath, S. Venkataraman. 2010. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science* **29**(5) 944-957.

Christy, C., X. Bo, I.L.B. Liu. 2012. The Impact of Observational Learning and Electronic Word of Mouth on Consumer Purchase Decisions: The Moderating Role of Consumer Expertise and Consumer Involvement. *2012 45th Hawaii International Conference on System Science (HICSS)* 3228-3237.

Dholakia, U. 2011. *How Businesses Fare with Daily Deals: A Multi-Site Analysis of Groupon, Livingsocial, Opentable, Travelzoo, and Buywithme Promotions*. Available at SSRN: http://ssrn.com/abstract=1863466.

Duan, W., B. Gu, A.B. Whinston. 2009. Informational cascades and software adoption on the Internet: An empirical investigation. *Mis Quarterly* **33**(1) 23-48.

Edelman, B., S. Jaffe, S.D. Kominers. 2010. *To Groupon or Not to Groupon: The Profitability of Deep Discounts*.

Firat, A., S. Madnick, M. Siegel. 2000. *The cameleon web wrapper engine*.

Granovetter, M.S. 1973. The strength of weak ties. *American journal of sociology* **78**(6) 1360-1380.

Herzenstein, M., U.M. Dholakia, R.L. Andrews. 2011. Strategic Herding Behavior in Peer-to-Peer Loan Auctions. *Journal of Interactive Marketing* **25**(1) 27-36.

John, L.K., O. Emrich, M.I. North, S. Gupta. 2013. What Are Facebook "Likes" Really Worth? *Working Paper*.

Katz, M.L., C. Shapiro. 1994. Systems competition and network effects. *The Journal of Economic Perspectives* **8**(2) 93-115.

Kirmani, A., A.R. Rao. 2000. No pain, no gain: A critical review of the literature on signaling unobservable product quality. *The Journal of Marketing* 66-79.

Kumar, V., B. Rajan. 2012. Social coupons as a marketing strategy: a multifaceted perspective. *Journal of the Academy of Marketing Science* **40**(1) 120-136.

Li, X. 2013. How Does Online Reputation Affect Social Media Endorsements and Product Sales? Evidence from Regression Discontinuity Design. *Working Paper*.

Liu, Y. 2006. Word-of-mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing* **70**(3) 74-89.

Moretti, E. 2011. Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies* **78**(1) 356-393.

Muchnik, L., S. Aral, S.J. Taylor. 2013. Social Influence Bias: A Randomized Experiment. *Science* **341**(6146) 647-651.

Nelson, P. 1974. Advertising as information. *The Journal of Political Economy* **82**(4) 729-754.

Nisbett, R.E., L. Ross. 1980. *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall Englewood Cliffs, NJ.

Roodman, D. 2007. How to do xtabond2: An introduction to difference and system GMM in Stata.

Trusov, M., R.E. Bucklin, K. Pauwels. 2009. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing* **73**(5) 90-102.

Tucker, C. 2012. Social advertising. *Available at SSRN 1975897*.

Wojnicki, A., D. Godes. 2008. Word-of-mouth as self-enhancement. *HBS Marketing Research Paper*.

Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data*. The MIT press.

Wu, J., M. Shi, M. Hu. 2013. Threshold Effects in Online Group Buying. *Available at SSRN 2176554*.

Young, H.P. 2009. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *The American Economic Review* **99**(5) 1899-1924.

Zhang, J., P. Liu. 2012. Rational Herding in Microloan Markets. *Management Science* **58**(5) 892-912.

Essay Three

How Does Online Reputation Affect Social Media Endorsements and Product Sales? Evidence from Regression Discontinuity Design

Abstract

Despite the increasing importance of social media marketing, little research has explored what factors consumers would take into account in the decision-making of endorsing a product to their peers with established ties via social media. This paper examines if online reputation (restaurants' displayed Yelp ratings), which helps update consumers' perception of product value, is a causal factor that affects consumers' decisions of endorsing via Facebook and purchasing products (the restaurants' vouchers). We build a stylized Bayesian learning model and derive the hypotheses: (1) a higher online reputation leads to more social media endorsements and voucher sales, but only when it is built upon a sufficient amount of review ratings; (2) these effects are greater for restaurants with more reviews; and (3) these effects are greater for restaurants with a larger variance in the review ratings. Interestingly, the third hypothesis contrasts to the predictions by some established theories (e.g., cue diagnosticity theory). We test the hypothesis using data of Groupon and LivingSocial deals. To identify the causal effects of online reputation, we use a regression discontinuity design by exploiting the institutional feature that displayed Yelp ratings are rounded to the nearest half star. The empirical results largely support the hypotheses. In particular, we find the effect of displayed Yelp ratings on Facebook Likes is greater when the variance of ratings is larger, suggesting that perhaps consumers are risk averse when they consider endorsing a product to their peers. Yet, the effect on voucher sales does not significantly differ with the variance. This paper concludes with important implications for theory and practice.

1. Introduction

Most online review sites (e.g., Yelp, Amazon) calculate an overall rating score by averaging across all individual ratings of a product. The overall average rating becomes an indicator of online reputation signaling the product quality. Besides that, many review sites prominently display the total number of ratings and make the dispersion of ratings available by showing the numbers of individual ratings at each level (often from 1 to 5 star). The central hypothesis underlying such practice is that online reputation together with the number and dispersion of ratings could influence consumers' shopping behaviors.

Prior research (e.g., Chevalier and Mayzlin 2006, Luca 2011) has focused on establishing the casual impact of online reputation on product sales, but neglects consumers' social media endorsements which are also significantly meaningful to firms (Aral et al. 2013). For example, Facebook.com provides the "Like" button allowing users to share and endorse any product webpage. The activity that the users have "liked" the product page is immediately displayed to their friends via Facebook newsfeeds. Recent studies show that consumers' social media activities can increase product awareness (Aral and Walker 2011), drive additional sales (Chen et al. 2011), and enhance brand loyalty (Rishika et al. 2013). Chompon, an e-commerce platform company, estimates that each Facebook Like is worth $8 for its clients in terms of the immediate next sale.[28] Li and Wu (2013) find that a 10% increase in the number of Facebook Likes associated with a Groupon deal on average leads to an increase in revenue of $19 in the next hour. Consumers' social media activities also have a significant predictive power for firm equity value (Luo et al. 2013). Therefore, engaging with consumers through social media has become "a critical element of any organization's marketing strategy" (Malhotra et al. 2013) and the volume of social media endorsements (e.g., Facebook Likes) is a meaningful and increasingly important indicator to firms' business performance (Luo, et al. 2013, Miller and Tucker 2013). The importance of social media endorsements is also evident in the fact that there exists a commercial market for buying them.[29]

Consumers' social media endorsements are distinct from product sales, because the motive and cost of endorsing a product to one's peers with established ties via social media are different from buying it for own consumption. Consumers endorse a product via Facebook, perhaps because the product is interesting (special, unique) and they want to express their preferences for it publicly, or because it is a good deal and they want to inform their friends about it. In either case, consumers may expect to gain "social currency" if their friends appreciate the endorsement (Berger and Schwartz 2011). From the cost perspective, such an endorsement can be done with minimal involvement (i.e., a click on the Facebook "Like" button) and no monetary cost, but consumers may put their self-image at risk; endorsing a "bad" product to Facebook friends would probably damage one's self-image (Wojnicki and Godes 2008). Therefore, consumers' decision-making of endorsing a product is different from purchasing and deserves to be investigated separately.

Despite the importance and distinction, little research has explored what factors consumers would take into account in the decision-making of endorsing a product to their peers. Our study aims to fill this literature gap by investigating how online reputation, which helps update consumers' perception of

---

[28] See http://techcrunch.com/2011/02/18/facebook-shares-are-worth-almost-three-times-more-than-tweets-for-e-commerce/ (accessed on July 9, 2013)

[29] A simple search on Google using the keywords "buy Facebook Likes" provides a list of companies that sell Facebook Likes to paying customers, such as get-likes.com, buylikes.com, and fblikesmart.com.

product value, affects social media endorsements. For comparison, we also examine the effect of online reputation on product sales. Although psychological theory of consumer choice (Hansen 1976) suggests that the effect of a determinant (herein, online reputation) is often moderated by contextual factors, the moderating role played by the number and variance of individual ratings has not received much attention (Sun 2012). Therefore, we study the moderating effects of the number and variance of ratings, from which we show consumers' endorsing behaviors are quite different from their purchasing behaviors. Specifically, we seek to answer the following questions in this study:

(1) Does a higher online reputation increase consumers' social media endorsements and product sales?
(2) How does the number of ratings moderate the effect of online reputation?
(3) How does the variance of ratings moderate the effect of online reputation?

To answer the questions, we, based on the theory of word-of-mouth (WOM) as self-enhancement (Berger 2014, Wojnicki and Godes 2008), assume consumers' propensity to endorse a product via social media is dependent on their expected utility of the product (perception of the product value), so is their propensity to buy. Then, we develop a simple stylized Bayesian learning model and show the structural relationship between a product's review ratings and consumers' posterior expected utility of the product. The analytical results from the stylized model produce testable hypotheses.

Empirically, we examine the situation in which restaurants with review ratings on Yelp.com sell deal vouchers through Groupon.com and LivingSocial.com. Being influenced by the restaurants' displayed Yelp ratings, consumers can endorse the restaurant deals via Facebook and/or buy the vouchers. Assembling a data set from multiple sources (Groupon/LivingSocial, Facebook and Yelp), we are able to identify the causal impacts of displayed Yelp ratings on consumers' Facebook endorsements and voucher sales by using a regression discontinuity (RD) design (Hartmann et al. 2011, Imbens and Lemieux 2008, Lee and Lemieux 2010). In line with the recent econometric literature (Lee and Lemieux 2010), we carefully assess the validity of the RD design in our study using a number of robustness checks. The results show that a restaurant's higher displayed Yelp rating causes to increase consumers' endorsements (i.e., more Facebook Likes) and voucher sales, but only when the number of ratings is sufficiently large. Supporting that consumers' propensity to endorse a product depends on their perception of the product value, the empirical findings suggest expected utility (perception of product value) is a key factor that consumers would take into account in the decision-making of endorsing. The magnitudes of the estimated effects are practically significant. For restaurants with at least 20 Yelp reviews, an extra half-star displayed Yelp rating increases the aggregate volumes of Facebook Likes and voucher sales by 26.3% and 17.4%, respectively, after controlling for observed (and unobserved) characteristics of restaurant deals. However, these effects decrease significantly and even disappear for restaurants with fewer Yelp reviews.

More importantly, there seems to be no conclusive theoretical prediction for the moderating role of the variance of ratings. On the one hand, some established theories (Basuroy et al. 2006, Feldman and Lynch 1988, Sun 2012) predict that consumers' responsiveness to the average rating would decrease with the variance of ratings. For example, the cue diagnosticity theory (Feldman and Lynch 1988) suggests that consumers would reduce their reliance on the average rating as a quality signal when the variance of ratings is large, because they may find the quality signal is nondiagnostic (Basuroy, et al. 2006). Consequently, consumers would be less responsive to the average rating when the variance is larger. In a separate study, Sun (2012) develops an analytical model and shows the interaction effect between the average and variance of ratings on product sales is negative.

On the other hand, however, our stylized Bayesian learning model, which is built upon fairly general but different assumptions from the model of Sun (2012), shows that risk aversion could make consumers' posterior expected utility more responsive to the average rating when the variance of ratings is larger. Unlike the conventional Bayesian learning literature (Ching et al. 2011, Roberts and Urban 1988, Zhao et al. 2013), our model does not assume any explicit form for consumers' utility functions. Thus, the theoretical implications of our model hold true with a broad set of utility functions, including that for constant or decreasing absolute risk aversion (CARA / DARA) (Friend and Blume 1975).

Therefore, the competing predictions from our stylized model and alternative theories (Basuroy, et al. 2006, Feldman and Lynch 1988, Sun 2012) raise an interesting empirical question with important theoretical implications. Consistent with the results of our model, we find the effect of displayed Yelp ratings on Facebook Likes is greater when the variance of ratings is larger, implying that perhaps consumers are risk averse in endorsing restaurant deals via Facebook. Yet, we find the effect on voucher sales does not significantly differ with the variance, possibly because the mechanisms expounded by the competing theories offset each other in terms of purchasing. The fact that the variance of ratings moderates the effect of displayed Yelp ratings on Facebook Likes but not on voucher sales reveals that consumers exhibit different behaviors in endorsing versus purchasing. Understanding the difference would help managers design appropriate strategies for boosting consumers' social media activities and product sales.

Our study contributes to the emerging literature that focuses on demonstrating the importance of consumers' social media activities (Aral, et al. 2013). For example, Rishika, et al. (2013) find consumers' participation in a firm's social media effort leads to an increase in consumer visit frequency. Li and Wu (2013) find consumers' Facebook Likes drive additional product sales. Kosinski et al. (2013) show that

Facebook Likes can be used to predict sensitive personal attributes. Despite the emerging literature on the importance of Facebook Likes, little is known about the determinants of consumers' endorsing decisions via social media. Egebark and Ekström (2011) conduct one of the first studies in this research stream by showing that conformity could affect one's decision to "like" a Facebook status update. While Egebark and Ekström (2011) study the consumer-side influence on "liking" Facebook status updates, our research examines the seller-side influence on "liking" commercial products. Our study is also distinct from the works by Moe and Schweidel (2012) and Muchnik et al. (2013), because the motives and costs of endorsing a product to one's peers via social media are different from that of writing a positive online review (Moe and Schweidel 2012) or voting up a news article (Muchnik, et al. 2013).[30] To the best of our knowledge, we may be the first to establish and quantify the causal impact of a seller's online reputation (user-generated review ratings) on consumers' decisions of endorsing commercial products to their peers with established social ties. Our findings suggest that consumers take into account their perception of product value when they consider endorsing a product to their peers.

Our study also contributes to a large body of literature that examines the impact of review ratings on product sales. The existing literature, however, documents mixed empirical findings. While a considerable number of studies document a higher average rating could increase product sales (Chevalier and Mayzlin 2006, Chintagunta et al. 2010), it has been recognized that the average rating may not necessarily reveal the true product quality (Hu et al. 2009) or influence consumers' purchasing decisions (Eliashberg and Shugan 1997) due to at least two reasons. First, consumers may realize online reviews could be posted by biased consumers (Li and Hitt 2008). Second, firms have incentives to manipulate their online reputation by posting fake reviews (Dellarocas 2006). It is thus not surprising that Liu (2006) finds the valence of movie messages has little explanatory power for movie revenue. Duan et al. (2008) show similar findings and conclude "online user reviews have little persuasive effect on consumer purchase decisions". Therefore, whether a higher online reputation increases product sales or social media endorsements is still an open empirical question.

One way that could potentially reconcile the seemingly inconsistent empirical findings about the impact of review ratings is to examine the moderating role played by contextual factors. For example, Zhu and Zhang (2010) find the average rating has an influential impact on sales of video games only for less popular games. The literature in this research stream is still scant. Our study contributes to this

[30] The key distinction is that receivers of social media endorsements are consumers' peers with established social ties (usually real friends) and thus they may expect to gain social currency or risk their self-image from the endorsements, whereas users on online review or news sites are often anonymous and have no established social ties among them.

growing literature by reporting that the number and variance of ratings could moderate the effect of average ratings.

Perhaps the two prior studies by Luca (2011) and Anderson and Magruder (2012) are most closely related to our study from the perspective of empirical identification. Our study differs in the following important aspects. First, we seek to establish the causal impact of displayed Yelp ratings on consumers' social media endorsements, whereas the prior studies focus on the effect on product sales, such as restaurant revenues (Luca 2011) and table reservation availability (Anderson and Magruder 2012). Second, we investigate how such causal impacts are moderated by the number and variance of individual ratings. Last but not least, while Luca (2011) shows the impact of displayed Yelp ratings on restaurant revenues is larger for restaurants with more reviews, Anderson and Magruder (2012) find the opposite: the effect on table reservation availability is smaller for restaurants with more reviews. Our study contributes to the literature by providing new empirical evidence consistent with Luca's findings.

The rest of this paper is organized as follows. In Section 2, we present a simple stylized Bayesian learning model and derive the hypotheses. We also discuss competing predictions by alternative theories. In Section 3, we describe the research setting and data. In Section 4, we present the identification strategy and estimation specifications. In Sections 5 and 6, we report the empirical results and robustness checks, respectively. Finally, we discuss the implications and conclude the paper in Section 7.

2. Theory

We develop a simple stylized model based on well-established assumptions from the classic Bayesian learning literature (Erdem and Keane 1996, Roberts and Urban 1988) to derive testable hypotheses. The simple stylized model results in a prediction about the moderating role of the variance of ratings which is in contrast to the predictions by some established theories.

2.1 A Simple Stylized Model

When consumers endorse a product to their peers via social media, they communicate not only information but also something about themselves (Berger and Schwartz 2011). Because people enjoy an enhanced self-image (identity) (Akerlof and Kranton 2000), consumers want their peers to think highly of them and often endeavor to associate themselves with superior products and brands (Berger and Heath 2007). Based on the theory of word-of-mouth (WOM) as self-enhancement (Berger 2014, Wojnicki and Godes 2008), we expect consumers are more likely to endorse good products to their peers. Accordingly, we assume consumers' propensity to endorse a product via social media is dependent on their expected utility of the product (perception of product value), so is their propensity to buy.

Herein, we develop a simple stylized Bayesian learning model to show the structural relationship between a product's review ratings and consumers' posterior expected utility. Following to the setup of the classic Bayesian learning model by Roberts and Urban (1988), we make the assumptions A1-A4:

- Assumption 1 (A1): A consumer i's prior belief about the value of a product j is X_{ij0}, where $X_{ij0} \sim N(\mu_{ij0}, \sigma_{ij0}^2)$ and σ_{ij0}^2 indicates the information uncertainty in consumer i's prior belief.
- Assumption 2 (A2): Each review rating is an unbiased[31] but imperfect signal of the value of product j, which is normally distributed with mean μ_j and variance σ_j^2. The random disturbance in the signal is normally distributed with zero mean and variance σ_j^2, which reflects "inherent product variability" (Roberts and Urban 1988) and "idiosyncratic perceptions" (Erdem and Keane 1996).
- Assumption 3 (A3): Consumers use a Bayesian updating rule to produce their posterior beliefs about the product value.
- Assumption 4 (A4): Consumers are risk averse (and prudent) with a utility function satisfying $u' > 0, u'' < 0, u''' > 0$.

A1-A3 are common assumptions in the Bayesian learning literature (Ching, et al. 2011, Roberts and Urban 1988, Zhao, et al. 2013). Besides the assumption of risk aversion, the Bayesian learning literature often assumes consumers are forward-looking (Ching, et al. 2011). Since this study aims to explore how online reputation would affect consumers' endorsing behaviors, we assume consumers are less likely forward-looking in this study, because endorsing a product via social media is not trial consumption and would not increase their information sets about the product value.

Note that unlike the conventional Bayesian learning literature (Ching, et al. 2011, Roberts and Urban 1988, Zhao, et al. 2013), A4 does not assume any explicit form for utility function. A4 is a fairly general assumption in that any utility function for either constant or decreasing absolute risk aversion (CARA / DARA)[32] (Friend and Blume 1975) implies A4. In fact, A4 is first introduced by Kimball (1990) in the economics literature as the notion of "prudence" - consumers are risk averse and have a positive

[31] Selection bias in online reviews (Li and Hitt 2008) may result in the fact that the average rating does not necessarily signal the true product value. We assume each review rating is an unbiased signal, because this study aims to explore how unbiased review ratings affect consumers' beliefs about the product value and their behaviors about the product.

[32] The coefficient of absolute risk aversion is defined as $A(w) = -\frac{u''(w)}{u'(w)}$. Constant absolute risk aversion (CARA) means $A(w)$ is constant and the exponential utility function $u(w) = -e^{-rw}, r > 0$, is unique in exhibiting CARA. Decreasing absolute risk aversion (DARA) means $A'(w) < 0$. Although experimental and empirical evidences are mostly consistent with DARA (Friend and Blume 1975), CARA is often assumed in the Bayesian learning literature for the sake of mathematical tractability.

precautionary saving motive; consumers' current savings increase with the uncertainty about their future incomes.[33] Subsequently, Eeckhoudt et al. (1995) introduce A4 in a management application.

Suppose there are n_j review ratings about product *j*. According to A2, we know the mean $\bar{x}_j$ of the n_j review ratings is normally distributed with $E(\bar{x}_j) = \mu_j$, $\sigma_{\bar{x}_j}^2 = \frac{\sigma_j^2}{n_j}$, where $\sigma_{\bar{x}_j}^2$ is the variance of the n_j review ratings and indicates the information uncertainty of the review ratings. Based on A1-A3, it can be shown that consumer *i*'s posterior belief about the product value, X_{ij}, is also normally distributed, $X_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$. As Roberts and Urban (1988) show, the mean and variance (information uncertainty) of consumer *i*'s posterior belief are given by

$$\mu_{ij} = \frac{\tau_i \mu_{ij0} + n_j \bar{x}_j}{\tau_i + n_j} \tag{1}$$

$$\sigma_{ij}^2 = \left(\frac{\tau_i}{\tau_i + n_j}\right)^2 \sigma_{ij0}^2 + \left(\frac{n_j}{\tau_i + n_j}\right)^2 \sigma_{\bar{x}_j}^2 \tag{2}$$

where τ_i is the relative strength / precision of consumer *i*'s prior belief, $\tau_i = \frac{\sigma_j^2}{\sigma_{ij0}^2}$.

Based on A1-A4, we can prove the following proposition (the proof is given in the appendix).

Proposition 1. Suppose consumers obey the von Neumann–Morgenstern axioms to produce the expected utility for decision-making. Given A1-A4, consumer *i*'s posterior expected utility of product *j* after learning product *j*'s review ratings, $E[U(X_{ij})]$, has the following properties:

(a) $E[U(X_{ij})]$ is increasing and concave w.r.t. the mean $\bar{x}_j$ of product *j*'s review ratings, i.e.,

$$\frac{\partial E[U(X_{ij})]}{\partial \bar{x}_j} > 0, \qquad \frac{\partial^2 E[U(X_{ij})]}{\partial \bar{x}_j{}^2} < 0;$$

(b) $E[U(X_{ij})]$ is decreasing w.r.t. the variance $\sigma_{\bar{x}_j}^2$ of product *j*'s review ratings, i.e.,

$$\frac{\partial E[U(X_{ij})]}{\partial (\sigma_{\bar{x}_j}^2)} < 0;$$

(c) The cross-partial derivative of $E[U(X_{ij})]$ w.r.t. $\bar{x}_j$ and $\sigma_{\bar{x}_j}^2$ is positive, i.e.,

$$\frac{\partial^2 E[U(X_{ij})]}{\partial \bar{x}_j \partial (\sigma_{\bar{x}_j}^2)} > 0.$$

[33] As Kimball (1990) explains, "prudence" is meant to suggest one's propensity to prepare and forearm oneself in the face of uncertainty in future income, whereas "risk aversion" simply indicates one dislikes facing uncertainty.

Given that consumers are risk averse (A4), properties (a) and (b) in Proposition 1 are intuitive. Risk-averse consumers would increase their posterior expected utility of a product if it has a higher average rating. The marginal posterior expected utility induced by the average rating diminishes with a higher average rating. When the variance of ratings is large, risk-averse consumers would decrease their posterior expected utility due to the large information uncertainty of the quality signal provided by the ratings.

By the same token, when the variance is large, it may be expected that risk-averse consumers are less responsive to the average rating because they may reduce their reliance on the review ratings due to the information uncertainty (Basuroy, et al. 2006). Somewhat counterintuitively, property (c) in Proposition 1 shows the opposite: consumers' posterior expected utility is more responsive to the average rating when the variance is larger. Although property (c) seems counterintuitive, it is intuitively understandable. Figure 1 illustrates the intuition. The expected utility is increasing and concave w.r.t. the certainty equivalent which is a function of the average and variance of ratings. When the variance rises, risk-averse consumers reduce their posterior expected utility to a lower level where the marginal expected utility induced by an incremental increase in the average rating (i.e., the first-order derivative of $E[U(X_{ij})]$ w.r.t. $\bar{x}_j$) is greater, because consumers are risk averse and the utility function is concave. It is exactly risk aversion (A4) that results in property (c).

According to property (a) in Proposition 1, a higher average rating increases a consumer's posterior expected utility of the product. However, when the number of reviews n_j goes to zero, Equation (1) shows the weight of the average rating in consumers' posterior beliefs reduces to zero, suggesting that the positive marginal expected utility w.r.t. the average rating may be minimal and empirically undetectable when n_j is too small. After all, if a product only has a few review ratings, consumers may doubt the representativeness of the only few ratings and simply ignore the quality signal of the average rating. On the other hand, when the average rating is calculated based upon a larger sample of reviews, the weight of the average rating, compared to the prior beliefs, increases and the information uncertainty in review ratings reduces. Consequently, the effect of the average rating would increase. By assuming consumers' propensities to endorse and buy a product are dependent on their posterior expected utility, we therefore formulate the following hypotheses.

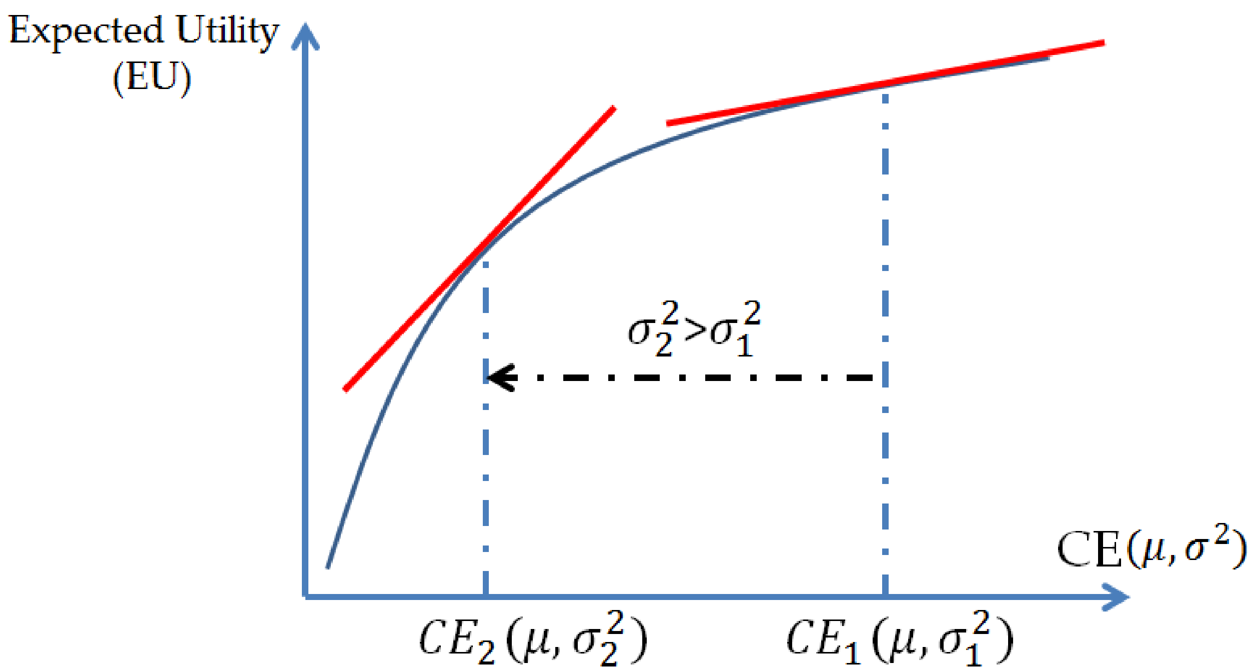

Figure 1: Illustration for the intuition about property (c) in Proposition 1

Notes: The expected utility is increasing and concave w.r.t. the certainty equivalent (CE) which is a function of the average and variance of ratings. A large variance makes the slope of expected utility w.r.t. the average rating steeper so that the expected utility is more responsive to the average rating. Consequently, the cross-partial derivative of the expected utility w.r.t. the average and variance of ratings is positive.

Hypothesis (H1): *A restaurant's higher online reputation (displayed Yelp rating) increases consumers' social media endorsements and voucher sales, but only when the restaurant has enough reviews.*

Hypothesis (H2): *The effects of a restaurant's online reputation (displayed Yelp rating) on consumers' social media endorsements and voucher sales are greater for restaurants with more reviews.*

According to property (c) in Proposition 1, consumers' posterior expected utility is more responsive to the average rating when the variance of ratings is larger. We hypothesize

Hypothesis (H3A): *The effects of a restaurant's online reputation (displayed Yelp rating) on consumers' social media endorsements and voucher sales are greater for restaurants with a larger variance of ratings.*

2.2 Predictions by Alternative Theories

In regard to the moderating role of the variance of ratings, two established theories in the literature would predict the opposite of H3A. First, the cue diagnosticity theory (Feldman and Lynch 1988) suggests that when the variance of a product's ratings is large, consumers' reliance on the average rating as a specific cue signaling product quality may reduce, because they may find it nondiagnostic and turn to alternative quality signals other than review ratings (Basuroy, et al. 2006). Second, Sun (2012) develops an analytical model, which incorporates consumer preference heterogeneity and mismatch costs, and shows the interaction effect between the average and variance of ratings on product sales is negative. The intuition behind her model lies in that a large variance of ratings could improve consumers' perception of the product quality only if the average rating is low. When the average rating rises, the dominant role played by a large variance would change to signal a high mismatch cost and reduce quality perception. Based on the two alternative theories, we hypothesize

Hypothesis (H3B): *The effects of a restaurant's online reputation (displayed Yelp rating) on consumers' social media endorsements and voucher sales are smaller for restaurants with a larger variance of ratings*.

In sum, our simple stylized model based on well-established assumptions from the Bayesian learning literature shows that the effects of online reputation as indicated by the average rating increase with the variance of ratings when consumers are risk averse, while alternative theories (Basuroy, et al. 2006, Feldman and Lynch 1988, Sun 2012) suggest mechanisms for the competing prediction. Therefore, whether the effects of online reputation increase or decrease with the variance of ratings is an empirical question which has important theoretical implications and will be answered in this study.

3. Research Setting and Data

3.1 Setting

We choose the daily-deal businesses as our research setting because of its practical importance and theoretical relevancy. The popularity of using daily deals as a new marketing vehicle has dramatically increased in recent years (Dholakia 2012). As of April 2012, consumers in North America have spent approximately $7 million a day (more than $2.5 billion a year) on daily deals[34] and it is projected to reach

[34] See http://savvr.com/2012/04/top-10-highest-grossing-daily-deals-of-all-time/, (accessed on March 10, 2013)

$5.5 billion a year by 2016.[35] Many restaurants have been selling vouchers through daily-deal sites. While some have attracted thousands of consumers, others have acquired only a few. It is thus important to understand what factors would affect consumers' response to these deals.

Besides the practical importance, the leading daily-deal sites (Groupon, LivingSocial) provide an ideal context for us to identify the causal effects of online reputation on consumers' social media endorsements and product sales.

First, leading daily-deal sites provide the setting where we are able to accurately collect the two outcome variables (i.e., aggregate numbers of consumers' endorsements via Facebook and voucher sales for each deal) so that we could quantify the effects precisely. Figure 2 shows a screenshot of a typical restaurant deal from Groupon. On the deal page, consumers can see the characteristics of the deal, such as restaurant name, discounted voucher price, and the displayed star rating of the restaurant from third-party reputation sites (most likely Yelp.com). Consumers can buy the deal and/or endorse it by clicking the Facebook "Like" button (as circled in Figure 2).

Second, to identify the causal effect of displayed Yelp ratings, it is required to prevent consumers from "interfering" Yelp ratings so that the possible reverse causality is avoided. Based on our inspection, the Yelp star ratings displayed on Groupon deal pages are hard-coded and fixed during the deal promotion. Moreover, since most restaurant deals are only sold for one or two days and the vouchers are often valid for redemption within six months, it is less likely for consumers who buy vouchers to redeem them immediately and post review ratings on Yelp.com when the deal is still on sale. Therefore, displayed Yelp ratings in this setting are largely exogenous.

For data about online reputation, we choose Yelp.com as the data source, because it is perhaps the most well-known and widely-used third-party site providing user-generated reviews about restaurants. Particularly, in most cases where Groupon deals are related to restaurants, their overall Yelp star ratings are prominently displayed on the deal pages (as circled in Figure 2). Thus, consumers are likely to be influenced by the restaurants' Yelp ratings when they look at the restaurant deals, which is supported by the survey conducted by Kimes and Dholakia (2011). Consumers may further go to the restaurants' Yelp profiles through the hyperlinks and check detailed information about the reviews, such as the number and variance of ratings. Correspondingly, we focus on the category of restaurant deals about which Yelp review ratings are most often available.

[35] See http://streetfightmag.com/2012/09/17/forecast-consumer-daily-deals-spending-to-reach-5-5-billion-by-2016/, (accessed on May 13, 2013)

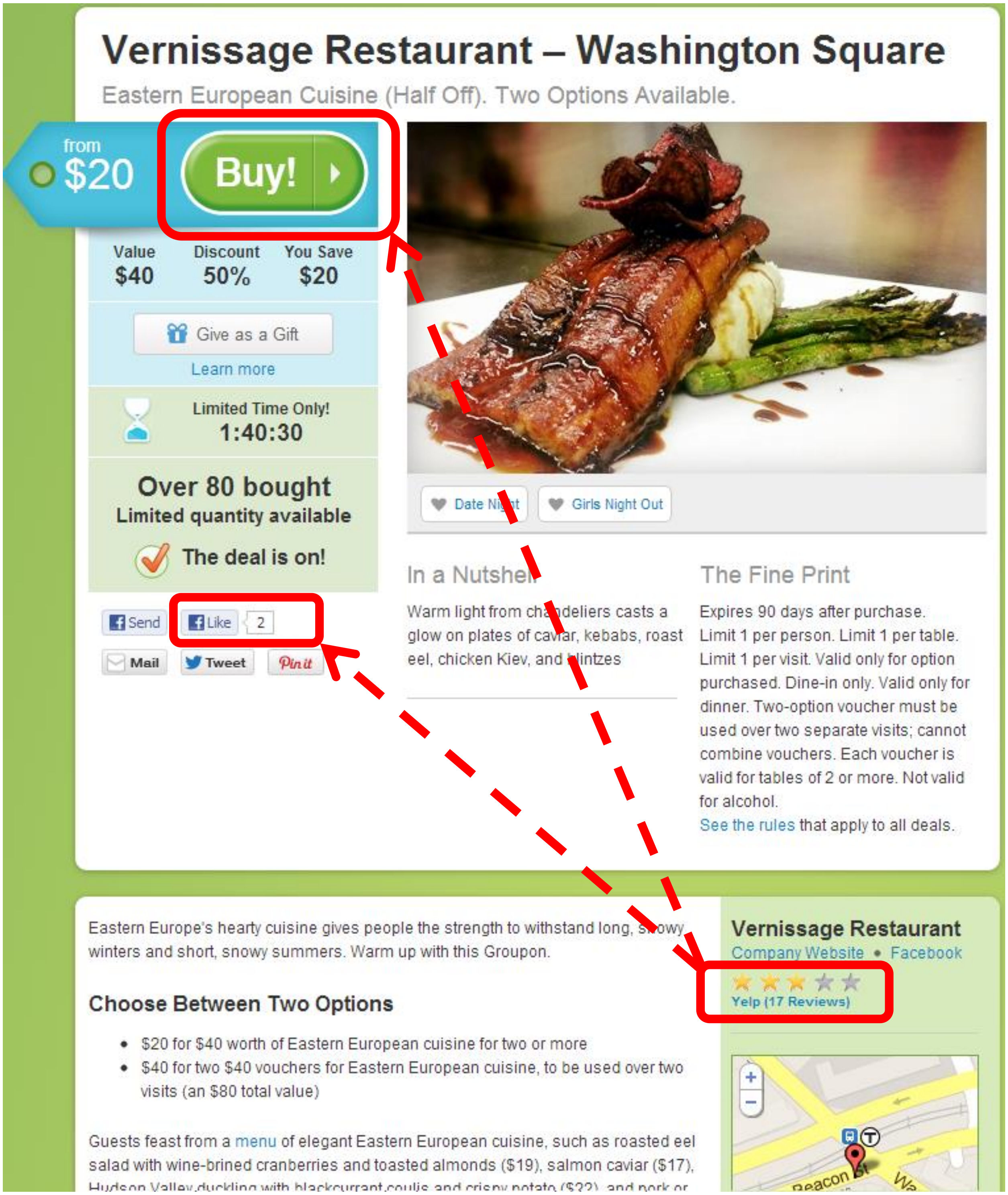

Figure 2: Screenshot of a typical restaurant deal from Groupon.com

Notes: Restaurants' overall Yelp ratings (if any) are often prominently displayed as well as the number of reviews, which could potentially influence consumers to endorse and/or buy the deal.

3.2 Data

We collect the data about restaurant deals from two sources: one is from the dataset provided by Byers et al. (2012) (named as BMZ) and the other is from Yipit.com, an aggregator of daily deals. The BMZ dataset contains a nationwide sample of deals distributed from 19 major cities across the US.[36] The BMZ dataset includes the characteristics (e.g., vendor, discounted voucher price) and accurate voucher sales of each deal. Besides that, the BMZ dataset contains the accurate number of Facebook Likes associated with each deal. In the BMZ dataset, Groupon deals are collected between January 3rd and July 3rd of 2011, and LivingSocial deals are collected between March 21st and July 3rd of 2011. Thus, we turn to Yipit.com and additionally collect LivingSocial deals between January 3rd and March 20th of 2011 (for the same 19 cities). For those LivingSocial deals from Yipit, all relevant deal characteristics are collected but not Facebook Likes. In total, we assemble 3,311 restaurant deals from the 19 US cities between January 3rd and July 3rd of 2011.

For each restaurant deal, we manually check if the restaurant has a profile on Yelp.com.[37] Since our study aims at identifying the impact of a restaurant's Yelp ratings, we exclude restaurants for which we could not confidently find their Yelp profiles[38] and those with no reviews on their Yelp profiles. For each remaining restaurant's Yelp profile, we use a computer program to automatically extract all individual reviews (including numeric ratings, textual contents, and dates) that are posted by the date of the deal promotion. Ultimately, we assemble a cross-sectional dataset consisting of 2,545 restaurant deals and 129,129 individual review ratings (from 1 to 5 star).

In this study, we have two outcome variables: the number of Facebook Likes (*Likes*) and number of voucher sales (*Sales*). *Likes* measure the total number of "Likes" that consumers endorse for a restaurant deal via Facebook. *Sales* measure the total number of vouchers purchased for a restaurant deal. We collect explanatory and control variables at two aspects. One is about the deal characteristics, including voucher price, discount rate, the number of days that a deal promotion lasts, and a dummy indicating whether it is from Groupon (coded as 1) or LivingSocial. The other is about the restaurant characteristics, including the displayed overall Yelp ratings, number of individual ratings, the mean and variance of individual ratings. We code a proxy variable for a restaurant's business age by calculating the

[36] The 19 US cities are Atlanta, Boston, Chicago, Dallas, Detroit, Houston, Las Vegas, Los Angeles, Miami, New Orleans, New York, Orlando, Philadelphia, San Diego, San Francisco, San Jose, Seattle, Tallahassee, and Washington DC.

[37] We use the restaurant information (e.g., name, street address, zip code, and phone number) to search the restaurants' profiles on Yelp.com.

[38] In order for a restaurant's Yelp profile to be confidently identified, we require (a) the restaurant deal must have a single physical location for redemption, and (b) the restaurant must only have one single Yelp profile.

number of days from when the restaurant's earliest Yelp review was posted to the promotion date. Table 1 reports the summary statistics of the key variables in our dataset.

Table 1. Summary Statistics

	N	Mean	S.D.	Min	Median	Max
Dependent Variables:						
Facebook Likes (*Likes*)	2459	50.35	73.93	0	31	1126
Voucher sales (*Sales*)	2545	910.93	994.77	0	660	26560
Explanatory / Control Variables:						
Voucher price ($)	2545	14.86	30.74	1	12	1500
Original value ($)	2545	30.34	61.67	4	25	3000
Discount rate	2545	50.89	3.03	0	50	83
Is Groupon or LivingSocial deal?	2545	0.84	0.37	0	1	1
Promotion duration (days)	2545	1.72	0.83	0	1	5
Restaurant's displayed Yelp rating	2545	3.62	0.60	1	3.5	5
No. of reviews per restaurant	2545	50.74	80.27	1	24	1186
Restaurant's true average rating	2545	3.61	0.59	1	3.63	5
Variance of a restaurant's ratings	2464	1.19	0.63	0	1.16	8
Proxy of restaurant age (days)	2545	993.64	656.91	1	943	2450

Notes: The notion of unbiased sample variance is used to calculate the variance of ratings, while it is undefined for restaurants with only one review.

4. Identification

In non-experimental studies, identifying the causal effects of online reputation is a challenging task due to the potential endogeneity problem; online reputation (e.g., the average review rating) is often correlated with unobserved heterogeneity that affects consumers' responses. For example, unobserved marketing expenditure is likely correlated with both online reputation and product sales. Without reasonably controlling for such unobserved heterogeneity, online reputation may just serve as a predictor of consumers' preference rather than an influencer (Eliashberg and Shugan 1997). To identify the causal effect of displayed Yelp ratings, we need variation in Yelp ratings that is uncorrelated with any deal or restaurant characteristics (e.g., unobserved marketing expenditure). Only the changes in consumers' responses produced by such variation in Yelp ratings could allow us to identify the causal effect.

Fortunately, Yelp's institutional feature of displaying the overall average ratings provides an opportunity for the identification strategy. For a restaurant with multiple review ratings (each ranging from 1 to 5 star), Yelp calculates the average of these ratings and rounds it up or down to the nearest half-star. For example, one restaurant with an average rating of 3.74 is rounded down and displayed as 3.5-star Yelp rating, while the other with an average rating of 3.76 is rounded up and displayed as 4-star. As a result, there is a half-star difference between the displayed Yelp ratings of the two restaurants, although their true average ratings are fairly close. The rounded average rating is prominently displayed on the restaurant's Yelp profile (and Groupon's deal page as shown in Figure 2), while the true average rating is not displayed.

For restaurants whose true average ratings fall in a small "window" centered on a threshold (in the above case 3.75), whether one gets rounded up or down is likely to be merely subject to random chance such that they appear to be randomly assigned around the threshold.[39] The only difference between the restaurants on the left and right of a threshold, on average, would be a half-star difference in the displayed Yelp ratings. Therefore, any possible discontinuity in consumers' responses to the restaurant deals (e.g., social media endorsements, voucher sales) could be attributed to the extra half-star displayed Yelp rating. The discontinuity induced by Yelp's displaying rule allows us to identify the causal effect of displayed Yelp ratings by implementing a regression discontinuity (RD) design (Hartmann, et al. 2011, Imbens and Lemieux 2008, Lee and Lemieux 2010).

Let r_i be the true average Yelp rating of restaurant deal i which may fall in a small (e.g., 0.2-star) bandwidth of a certain threshold c. The value of c can be 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25 or 4.75. Each of the thresholds with a bandwidth of 0.25 star corresponds to one rating range, such as (3.75$\pm$0.25). In total, there are 8 rating ranges between 1 and 5 stars. We pool the data from the 8 rating ranges and use the 0.2-star bandwidth in the main analysis (we also use different bandwidths in robustness checks and get similar results). We use local linear regression (Imbens and Lemieux 2008) as specified in Equation (3) to estimate the causal effects of displayed Yelp rating

$$y_i = \alpha_0 + \beta \times I(r_i \geq c) + \alpha_1 \times (r_i - c) + \alpha_2 \times (r_i - c) \times I(r_i \geq c) + \gamma X_i + \epsilon_i \tag{3}$$

where the outcome variable y_i is the natural log of deal i's *Likes* or *Sales*. $I(r_i \geq c)$ is the indication function. If $r_i \geq c$, then $I(r_i \geq c) = 1$ and the restaurant's average rating is rounded up to the nearest half-star; otherwise, $I(r_i \geq c) = 0$ and it is rounded down. The displayed Yelp ratings of rounded-up restaurants are, on average, half-star higher than that of the rounded-down restaurants. Because the discontinuity in outcome y_i is likely to be merely induced by the indication function $I(r_i \geq c)$, the coefficient β estimates the causal effect of an extra half-star displayed Yelp rating. In a valid RD design, including control variables is not necessary for estimating the causal effect, because restaurants around a

[39] This is the key identification assumption of regression discontinuity (RD) design, so-called local randomization around the threshold. In this study we conduct a number of robustness checks that assess the validity of the assumption.

threshold is “locally” randomized (Hartmann, et al. 2011, Imbens and Lemieux 2008, Lee and Lemieux 2010). Still, we include a vector of baseline covariates X_i about deal and restaurant characteristics (determined prior to the deal promotion) as controls to improve precision of the estimation. The full set of baseline controls include city, promotion duration, weekday, log of voucher price, log of number of reviews, whether it is a Groupon or LivingSocial deal, log of restaurant age proxy, and a categorical variable indicating the restaurant’s rating range.

To examine the differential impacts of displayed Yelp ratings for restaurants with more or less reviews, we follow the median split method which is commonly used in the current literature (Demers and Lewellen 2003, Efendi et al. 2012, Rishika, et al. 2013). That is, we create a dummy dn_i indicating if the number of restaurant i’s reviews is above or equal to the median of the sample.[40] We include the interaction term between dn_i and the indication function $I(r_i \geq c)$. Since the local linear regressions on the left and right of the threshold may have different slopes, we include interaction terms between dn_i and $(r_i - c)$, $(r_i - c) \times I(r_i \geq c)$. Accordingly, we estimate the moderating effect of dn_i using Equation (4) in which the coefficient β_2 identifies the difference between the RD estimates (differential effects) for restaurants with more or less reviews

$$\begin{aligned} y_i = \alpha_0 &+ \beta_1 \times I(r_i \geq c) + \beta_2 \times I(r_i \geq c) \times dn_i + \beta_3 \times dn_i \\ &+ \alpha_1 \times (r_i - c) + \alpha_2 \times (r_i - c) \times I(r_i \geq c) \\ &+ \alpha_3 \times (r_i - c) \times dn_i + \alpha_4 \times (r_i - c) \times I(r_i \geq c) \times dn_i + \gamma X_i + \epsilon_i \end{aligned} \tag{4}$$

Similarly, to examine the differential impacts of displayed Yelp ratings for restaurants with a large or small variance of individual ratings, we create a dummy dv_i indicating if the variance of restaurant i’s ratings is above or equal to the median of the sample. We estimate the moderating effect of dv_i using Equation (5) in which β_2 identifies the difference between the RD estimates for restaurants with a large or small variance of ratings. If $\beta_2 > 0$ in Equation (5), then H3A is confirmed; otherwise, if $\beta_2 < 0$ in Equation (5), then H3B is supported.

[40] Using the dummy variable (the median split method) makes it easy to interpret the moderating effect. Nevertheless, we also use the method of continuous variables: we include the interaction terms with the continuous variables of the number (variance) of ratings to estimate the moderating effects. The results are qualitatively similar.

$$
\begin{aligned}
y_i = \alpha_0 + \beta_1 \times I(r_i \geq c) + \beta_2 \times I(r_i \geq c) \times dv_i + \beta_3 \times dv_i \\
+\alpha_1 \times (r_i - c) + \alpha_2 \times (r_i - c) \times I(r_i \geq c) \\
+\alpha_3 \times (r_i - c) \times dv_i + \alpha_4 \times (r_i - c) \times I(r_i \geq c) \times dv_i + \gamma X_i + \epsilon_i
\end{aligned} \tag{5}
$$

5. Results

5.1 Balance Check on Baseline Covariates

Before presenting the results of RD estimation, we first show balance check on baseline covariates to support the "local randomization" assumption of RD design. We report more robustness checks in Section 6.

If restaurants are truly "locally" randomized around threshold, we expect all observed baseline covariates of the restaurant deals on the left and right of threshold would appear to be balanced, just like in a true randomized controlled experiment. In addition to the dummies of cities and weekdays, we collect 17 baseline covariates about deal and restaurant characteristics, such as voucher price, number of reviews, and the true average rating. Table 2 reports the results of balance check and show that the means of the 17 covariates of restaurant deals above and below threshold are all balanced. The balance check on the observed covariates also enhances our confidence that restaurant deals around threshold are comparable even in terms of unobserved heterogeneity. For example, unobserved marketing expenditure is likely correlated with some observed covariates, such as the number of reviews. Table 2 shows that the natural logs of number of reviews for the restaurants above and below threshold are quite close (3.30 *vs*. 3.37), suggesting that unobserved marketing expenditures between the two groups are plausibly comparable. Moreover, we use Kolmogorov–Smirnov test to examine the equality of distributions of the covariates and find their distributions between the two groups are also comparable. Figure 3 plots the density distributions of four covariates for restaurant deals above and below threshold and shows the distributions are fairly comparable. Therefore, the balance check gives us confidence that the RD design in our study is valid and the estimated effects of displayed Yelp ratings can be interpreted as causal.

Table 2. Balance Check on Baseline Covariates of Deal and Restaurant Characteristics

	Mean		Diff. in Means	*t*-statistic
	Above Threshold (rounded-up)	Below Threshold (rounded-down)		
Deal Characteristics				
log(Voucher price)	2.55	2.51	0.037	1.50
Value saved ($)	16.68	14.55	2.14	1.28
Full value ($)	32.71	28.51	4.20	1.26
Discount rate (%)	50.90	50.78	0.12	0.83
Is a deal from Groupon?	0.83	0.85	-0.026	-1.52
Promotion duration (days)	1.73	1.71	0.026	0.67
Restaurant Characteristics				
log(Number of reviews)	3.30	3.37	-0.073	-1.35
log(Restaurant age proxy)	6.65	6.63	0.023	0.53
True average rating	3.61	3.60	0.007	0.29
Variance of ratings	1.19	1.22	-0.027	-1.06
Percent of 5-star ratings	0.246	0.250	-0.003	-0.42
Percent of 4-star ratings	0.371	0.366	0.005	0.73
Percent of 3-star ratings	0.198	0.196	0.002	0.41
Percent of 2-star ratings	0.113	0.111	0.001	0.30
Percent of 1-star ratings	0.072	0.077	-0.005	-1.26
Number of reviews in the past month	2.67	2.71	-0.042	-0.24
Average number of reviews in the past three months	2.57	2.57	0	0

Notes: Balance check compares the baseline covariates of restaurant deals on the left and right of the threshold within a bandwidth of 0.2 star. Dummies of cities and weekdays are also checked but not reported in the table. The balance check confirms that restaurant deals that are above and below a threshold are comparable. *p < 0.10, **p < 0.05, ***p < 0.01

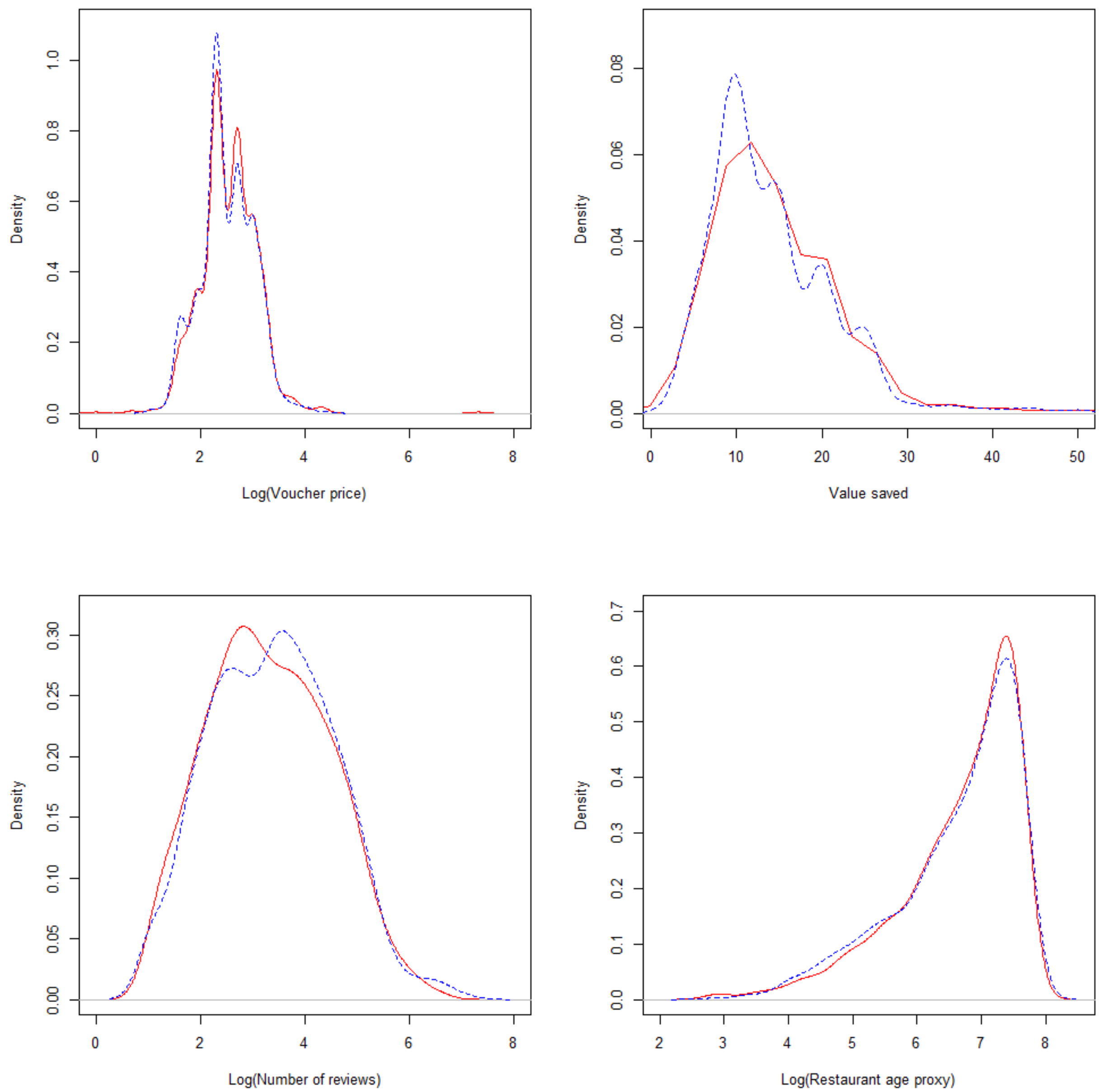

Figure 3. Density distributions of four covariates of restaurant deals above and below threshold.

Notes: The red solid distributions are for restaurant deals above threshold (rounded-up), while the blue dashed distributions are for restaurant deals below threshold (rounded-down). The density distributions of the two groups are fairly comparable.

5.2 Main Effects When Number of Reviews is Sufficiently Large

If a restaurant has only a few reviews on Yelp, consumers may not believe in the displayed Yelp rating and simply ignore it. In order for a restaurant's displayed Yelp rating to have an influential impact, the restaurant needs to have a sufficient amount of review ratings. Thus, we first focus on restaurants with at least 20 reviews[41] and estimate the main effects of display Yelp rating on Facebook Likes and voucher sales.

Table 3 presents the OLS estimates with log(*Likes*) as the dependent variable.[42] In Column (1), we only use a set of categorical baseline covariates as controls, including dummies of cities, weekdays, Yelp rating ranges and promotion duration. The significantly positive estimated coefficient of the indication function (discontinuity) $I(r_i \geq c)$ suggests that being rounded up (i.e., an extra half-star displayed Yelp rating) increases consumers' Facebook Likes. In Columns (2) to (5), more baseline covariates are included as controls in the estimation. Including baseline covariates as controls in a valid RD design helps improve the estimation precision but would not reduce bias (if any). As shown in Columns (1)-(5), the point estimates of the discontinuity remain fairly stable and become more precise and significant when additional covariates are included. The stable RD estimates increase our confidence that the RD design is valid; whether a restaurant falls above and below threshold is "locally" randomized. Table 3 shows consistent evidence that displayed Yelp ratings affects consumers' endorsements via Facebook for restaurants with enough reviews, supporting H1. The magnitude of the estimated effect is also practically significant. Column (5) suggests that for those restaurants with at least 20 reviews, an extra half-star displayed Yelp rating increases the total number of consumers' Facebook Likes by 26.3%.

Table 4 presents the OLS estimates with log(*Sales*) as the dependent variable. In Column (1), we only use a set of categorical covariates as controls and the estimated coefficient of the discontinuity is positive but not significant. When additional covariates are included, the positive coefficient estimates of the discontinuity become more significant. Again, the estimates of the discontinuity in Columns (1)-(5) are fairly stable, enhancing our confidence about the validity of the RD design. Thus, Table 4 shows consistent evidence that displayed Yelp ratings also affects voucher sales, supporting H1. Economically, Column (5) suggests that for those restaurants with at least 20 reviews, an extra half-star displayed Yelp rating increases voucher sales by 17.4%.

---

[41] In our sample, 45.3% (1154) out of the 2545 restaurants have less than 20 Yelp reviews and thus 20 is a substantive cutoff. Qualitative evidence from our interviews confirms that 20 reviews are often enough to make consumers believe the displayed Yelp ratings are meaningful. Other cutoffs (e.g., 15 or 25) provide qualitatively similar results.

[42] There are 77 deals with zero Facebook Likes, accounting for 3% of the sample. To include these deals in the regression, we also use log(*Likes*+0.5) or log(*Likes*+1) as alternative dependent variables and get similar results.

Table 3. RD Estimates of Displayed Yelp Effect on Facebook Likes

	(1)	(2)	(3)	(4)	(5)
Discontinuity $I(r_i \geq c)$	0.218*	0.273**	0.270**	0.241**	0.263**
	(0.13)	(0.12)	(0.12)	(0.11)	(0.11)
Distance $(r_i - c)$	-1.32	-1.34	-1.33	-0.721	-0.939
	(0.85)	(0.81)	(0.81)	(0.72)	(0.72)
$(r_i - c) \times I(r_i \geq c)$	0.944	0.680	0.682	0.034	0.264
	(1.10)	(1.06)	(1.06)	(0.95)	(0.95)
log(Number of reviews)		0.470***	0.466***	0.500***	0.543***
		(0.046)	(0.046)	(0.040)	(0.042)
log(Voucher price)			0.038	0.155**	0.160***
			(0.063)	(0.060)	(0.062)
Is a deal from Groupon?				1.35***	1.35***
				(0.067)	(0.067)
log(Restaurant age proxy)					-0.124***
					(0.040)
Promotion duration	Yes	Yes	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes	Yes	Yes
Number of Observations	1017	1017	1017	1017	1017
R^2	0.101	0.195	0.196	0.382	0.389

Notes: Dependent variable is log(*Likes*). OLS estimates of the effect of one extra half-star displayed Yelp rating on the number of Facebook Likes. All regressions use restaurants with at least 20 Yelp reviews and a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

Table 4. RD Estimates of Displayed Yelp Effect on Voucher Sales

	(1)	(2)	(3)	(4)	(5)
Discontinuity $I(r_i \geq c)$	0.126	0.157*	0.161*	0.156*	0.174**
	(0.091)	(0.084)	(0.084)	(0.083)	(0.082)
Distance $(r_i - c)$	-1.04*	-1.00*	-0.992*	-0.847	-1.06*
	(0.59)	(0.56)	(0.55)	(0.55)	(0.55)
$(r_i - c) \times I(r_i \geq c)$	0.888	0.679	0.649	0.464	0.708
	(0.81)	(0.77)	(0.77)	(0.76)	(0.75)
log(Number of reviews)		0.379***	0.388***	0.397***	0.448***
		(0.032)	(0.032)	(0.031)	(0.032)
log(Voucher price)			-0.102	-0.074	-0.066
			(0.067)	(0.070)	(0.073)
Is a deal from Groupon?				0.267***	0.276***
				(0.056)	(0.055)
log(Restaurant age proxy)					-0.145***
					(0.034)
Promotion duration	Yes	Yes	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes	Yes	Yes
Number of Observations	1087	1087	1087	1087	1087
R^2	0.137	0.234	0.238	0.250	0.263

Notes: Dependent variable is log(*Sales*). OLS estimates of the effect of one extra half-star displayed Yelp rating on the number of voucher sales. All regressions use restaurants with at least 20 Yelp reviews and a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

For comparison, we also use simple OLS regressions to estimate the effect of displayed Yelp rating on voucher sales. Table A.1 in the Appendix A reports the simple OLS estimates, suggesting that displayed Yelp rating has no effect or even significantly negative effect on voucher sales. The results in Table A.1 are similar to the simple OLS estimates reported by Byers, et al. (2012). The comparison between the estimates in Table 4 and Table A.1 reveals that simple OLS regressions without an appropriate identification strategy may produce misleading results, while results from the RD design would be more convincing.

It is worth commenting on the estimates of the key covariates in Tables 3 and 4. The coefficient estimates of number of reviews in both tables are positive and significant, suggesting that restaurants with more reviews are likely to receive more Facebook Likes and voucher sales. This is consistent with prior research (e.g., Liu 2006, Duan et al. 2008) that shows the volume of reviews has a significant predictive power for product sales. More interestingly, while the coefficient estimates of voucher price in Table 4 are all negative (though not significant), the estimates of voucher price in Table 3 are positive and significant. Perhaps a high voucher price is correlated with some unobserved factors (e.g., high quality, specialty) that encourage consumers to endorse the deal via Facebook, but meanwhile it decreases consumers' propensity to buy. The opposite signs of the estimates of voucher price in Tables 3 and 4 reveal that consumers do behave differently in endorsing versus purchasing the deals. The coefficient estimates of restaurant age proxy in both tables are negative and significant, indicating that a younger restaurant is associated with more Facebook Likes and voucher sales. This finding suggests that deal shoppers may favor relatively newer restaurants. Lastly, we find Groupon deals receive more Facebook Likes and voucher sales. This is not surprising, because Groupon as the industry leader has more subscribers than LivingSocial. In general, these findings are consistent with our intuition and enhance our confidence about the credibility of the dataset.

5.3 Moderating Effect of Number of Reviews

To examine the differential impacts of displayed Yelp ratings for restaurants with more or less reviews, we create a dummy dn_i indicating if the number of restaurant i's reviews is above or equal to the sample median. We first estimate the two subsamples separately and then use Equation (4) to estimate if the difference between the RD estimates is significant. The results are reported in Tables 5 and 6.

Column (1) of Table 5 shows the coefficient estimate of the discontinuity is positive and significant, indicating a strong positive effect of displayed Yelp rating on Facebook Likes for restaurants with above-median reviews. By contrast, Column (2) shows the coefficient estimate of the discontinuity is negative but insignificant, indicating that the effect of displayed Yelp rating is minimal for restaurants with below-median reviews. Column (3) confirms that the difference between the RD estimates is positive and significant. Therefore, the results in Table 5 suggest that a higher displayed Yelp rating increases consumers' Facebook Likes only when the restaurants have enough reviews. The effect of displayed Yelp rating on Facebook Likes decreases and even disappears when restaurants have few reviews.

Column (1) of Table 6 shows that coefficient estimate of the discontinuity is positive and significant, indicating a positive effect of displayed Yelp rating on voucher sales for restaurants with above-median reviews. Column (2) shows the coefficient estimate of the discontinuity is negative but insignificant, indicating the effect of displayed Yelp rating is minimal for restaurants with below-median reviews. Column (3) confirms that the difference between the RD estimates is positive and significant. Thus, the results in Table 6 suggest that a higher displayed Yelp rating increases voucher sales only when the restaurants have enough reviews. The results in Tables 5 and 6 support H1 and H2.

Table 5. Number of Reviews Moderates Displayed Yelp Effect on Facebook Likes

	(1) # of Reviews >= 24	(2) # of Reviews < 24	(3) Full Sample
Discontinuity $I(r_i \geq c)$	0.260**	-0.072	-0.102
	(0.11)	(0.13)	(0.13)
$I(r_i \geq c) \times dn_i$ (more reviews)			0.315*
			(0.17)
Distance $(r_i - c)$	-1.11	-0.496	0.117
	(0.76)	(0.93)	(0.92)
$(r_i - c) \times I(r_i \geq c)$	0.466	0.975	0.178
	(0.98)	(1.16)	(1.14)
dn_i (more reviews)			-0.363**
			(0.15)
$(r_i - c) \times dn_i$			-1.18
			(1.19)
$(r_i - c) \times I(r_i \geq c) \times dn_i$			0.311
			(1.51)
log(Number of reviews)	0.612***	0.190**	0.456***
	(0.046)	(0.064)	(0.038)
log(Voucher price)	0.152**	0.065	0.125**
	(0.064)	(0.078)	(0.049)
Is a deal from Groupon?	1.39***	1.27***	1.33***
	(0.070)	(0.097)	(0.056)
log(Restaurant age proxy)	-0.132***	-0.076**	-0.102***
	(0.043)	(0.033)	(0.026)
Promotion duration	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes
Number of Observations	940	792	1732
R^2	0.401	0.263	0.340

Notes: Dependent variable is log(*Likes*). OLS estimates show the effect of displayed Yelp rating on the number of Facebook Likes is greater for restaurants with more Yelp reviews. The coefficient estimate of $I(r_i \geq c) \times dn_i$ in Column (3) indicates the difference between the RD estimates of the effects of displayed Yelp rating for restaurants with more and less reviews. All regressions use a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

Table 6. Number of Reviews Moderates Displayed Yelp Effect on Voucher Sales

	(1) # of Reviews >= 24	(2) # of Reviews < 24	(3) Full Sample
Discontinuity $I(r_i \geq c)$	0.160*	-0.082	-0.081
	(0.089)	(0.090)	(0.091)
$I(r_i \geq c) \times dn_i$ (more reviews)			0.221*
			(0.128)
Distance $(r_i - c)$	-0.977*	0.091	0.208
	(0.58)	(0.70)	(0.70)
$(r_i - c) \times I(r_i \geq c)$	0.608	-0.364	-0.855
	(0.79)	(0.86)	(0.86)
dn_i (more reviews)			-0.341***
			(0.110)
$(r_i - c) \times dn_i$			0.221*
			(0.13)
$(r_i - c) \times I(r_i \geq c) \times dn_i$			1.43
			(1.18)
log(Number of reviews)	0.482***	0.303***	0.416***
	(0.035)	(0.044)	(0.028)
log(Voucher price)	-0.095	-0.032	-0.060
	(0.076)	(0.060)	(0.051)
Is a deal from Groupon?	0.293***	0.431***	0.336***
	(0.057)	(0.072)	(0.045)
log(Restaurant age proxy)	-0.125***	-0.106***	-0.111***
	(0.036)	(0.024)	(0.020)
Promotion duration	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes
Number of Observations	1005	837	1842
R^2	0.263	0.266	0.307

Notes: Dependent variable is log(*Sales*). OLS estimates show the effect of displayed Yelp rating on the number of voucher sales is greater for restaurants with more Yelp reviews. The coefficient estimate of $I(r_i \geq c) \times dn_i$ in Column (3) indicates the difference between the RD estimates of the effects of displayed Yelp rating for restaurants with more and less reviews. All regressions use a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

5.4 Moderating Effect of Variance of Ratings

To examine the differential impacts of displayed Yelp ratings for restaurants with a large or small variance of ratings, we create a dummy dv_i indicating if the variance of restaurant i's ratings is above or equal to the sample median. We estimate the two subsamples separately and then use Equation (5) to estimate if the difference between the RD estimates is significant. The results are reported in Tables 7 and 8 where unbiased sample variance of ratings is used.

Column (1) of Table 7 shows the coefficient estimate of the discontinuity is positive and significant, indicating a positive effect of displayed Yelp rating on Facebook Likes for restaurants with a large variance of ratings. By contrast, Column (2) shows the coefficient estimate of the discontinuity is negative and insignificant, indicating the effect of displayed Yelp rating is minimal for restaurants with a small variance of ratings. Column (3) confirms that the difference between the RD estimates is positive and significant. Therefore, the results in Table 7 suggest that the effect of displayed Yelp rating on Facebook Likes is greater for restaurants with a larger variance of ratings, supporting H3A.

The coefficient estimate of the discontinuity in Column (1) of Table 8 is positive and smaller than the counterpart estimate in Column (2), indicating that the effect of displayed Yelp rating on voucher sales might be smaller for restaurants with a larger variance of ratings. Yet, neither of the coefficient estimates is precise or significant. Accordingly, Column (3) shows that the difference between the RD estimates is negative but not significant. Despite the insignificance, the negative sign directionally suggests that the effect of displayed Yelp rating on voucher sales might be smaller for restaurants with a larger variance of ratings, consistent with H3B.

To summarize the findings about the moderating effects of number and variance of ratings, we provide RD estimates of the effects of displayed Yelp ratings on Facebook Likes and voucher sales for restaurants with above and below median reviews and variance of ratings in Table 9. The upper panel of Table 9 shows the effect of displayed Yelp rating on Facebook Likes is largest for restaurants with more reviews and a larger variance of ratings, whereas the lower panel of Table 9 shows the effect on voucher sales is largest for restaurants with more reviews but a smaller variance of ratings. Therefore, the findings indicate that consumers exhibit different behaviors when they consider endorsing or purchasing the deals.

Table 7. Variance of Ratings Moderates Displayed Yelp Effect on Facebook Likes

	(1) Large Variance	(2) Small Variance	(3) Full Sample
Discontinuity $I(r_i \geq c)$	0.200*	-0.119	-0.127
	(0.11)	(0.13)	(0.13)
$I(r_i \geq c) \times dv_i$ (large variance)			0.329*
			(0.17)
Distance $(r_i - c)$	-1.46*	0.180	0.183
	(0.78)	(0.92)	(0.90)
$(r_i - c) \times I(r_i \geq c)$	1.55	-0.444	-0.512
	(0.99)	(1.15)	(1.14)
dv_i (large variance)			-0.464***
			(0.14)
$(r_i - c) \times dv_i$			-1.48
			(1.18)
$(r_i - c) \times I(r_i \geq c) \times dv_i$			1.86
			(1.50)
log(Number of reviews)	0.354***	0.403***	0.385***
	(0.037)	(0.036)	(0.026)
log(Voucher price)	0.167***	0.107	0.139***
	(0.063)	(0.075)	(0.048)
Is a deal from Groupon?	1.29***	1.39***	1.34***
	(0.079)	(0.083)	(0.057)
log(Restaurant age proxy)	-0.087**	-0.128***	-0.110***
	(0.036)	(0.040)	(0.026)
Promotion duration	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes
Number of Observations	898	834	1732
R^2	0.328	0.387	0.345

Notes: Dependent variable is log(*Likes*). OLS estimates show the effect of displayed Yelp rating on the number of Facebook Likes is greater for restaurants with a large variance of ratings. The coefficient estimate of $I(r_i \geq c) \times dv_i$ in Column (3) indicates the difference between the RD estimates of the effects of displayed Yelp rating for restaurants with large and small variances of ratings. All regressions use a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses.

*p < 0.10, **p < 0.05, ***p < 0.01

Table 8. Variance of Ratings Moderates Displayed Yelp Effect on Voucher Sales

	(1) Large Variance	(2) Small Variance	(3) Full Sample
Discontinuity $I(r_i \geq c)$	0.0019	0.026	0.041
	(0.087)	(0.093)	(0.091)
$I(r_i \geq c) \times dv_i$ (large variance)			-0.031
			(0.13)
Distance $(r_i - c)$	-0.762	-0.055	-0.138
	(0.62)	(0.64)	(0.62)
$(r_i - c) \times I(r_i \geq c)$	1.26	-1.14	-1.15
	(0.80)	(0.86)	(0.85)
dv_i (large variance)			-0.250**
			(0.098)
$(r_i - c) \times dv_i$			-0.584
			(0.86)
$(r_i - c) \times I(r_i \geq c) \times dv_i$			2.32**
			(1.16)
log(Number of reviews)	0.323***	0.391***	0.357***
	(0.029)	(0.028)	(0.020)
log(Voucher price)	-0.045	-0.043	-0.048
	(0.077)	(0.056)	(0.050)
Is a deal from Groupon?	0.289***	0.416***	0.348***
	(0.064)	(0.062)	(0.045)
log(Restaurant age proxy)	-0.092	-0.145***	-0.117***
	(0.025)	(0.031)	(0.020)
Promotion duration	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes
Number of Observations	954	888	1842
R^2	0.291	0.354	0.310

Notes: Dependent variable is log(*Likes*). OLS estimates show the effect of displayed Yelp rating on the number of voucher sales is smaller for restaurants with a large variance of ratings. The coefficient estimate of $I(r_i \geq c) \times dv_i$ in Column (3) indicates the difference between the RD estimates of the effects of displayed Yelp rating for restaurants with large and small variances of ratings. All regressions use a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses.

*p < 0.10, **p < 0.05, ***p < 0.01

Table 9. RD Estimates of Effects of Displayed Yelp Ratings for Different Subsamples

	Above Median Variance	Below Median Variance
Dependent Variables: log(Likes)		
Above Median Reviews	0.273	0.175
	(0.15)	(0.17)
Below Median Reviews	0.041	-0.155
	(0.18)	(0.20)
Dependent Variables: log(Sales)		
Above Median Reviews	0.015	0.179
	(0.12)	(0.13)
Below Median Reviews	-0.059	-0.135
	(0.13)	(0.13)

Notes: RD estimates of the effects of displayed Yelp ratings on Facebook Likes and voucher sales for different subsamples. The upper panel is produced using the model in Column (5) of Table 2. The bottom panel is produced using the model in Column (5) of Table 3. Robust standard errors are reported in parentheses.

6. Robustness Checks

Besides the balance check reported in Section 5.1, we conduct a number of additional robustness checks to verify if the RD design in our study is valid and the findings are robust.

6.1 Inspection of Possible Review Manipulation

The key identification assumption of a valid RD design is that the restaurants could be considered as "locally" randomized around the threshold. If some restaurants could precisely manipulate their average ratings (e.g., through posting fake review ratings) and therefore are more likely to be rounded up, the identification assumption would be invalidated (Hartmann, et al. 2011). Although it is difficult to directly observe restaurants' review manipulation (Mayzlin et al. 2012), prior studies (Anderson and Magruder 2012, Luca 2011) provide both qualitative arguments and empirical evidence that restaurants' incentives to manipulate Yelp ratings is less likely an issue for the RD design used in our study.

Herein, we add two additional arguments. First, Yelp.com has been actively fighting with possible fake reviews by using advanced detection algorithms and punishment policies.[43] Second, in order for the possible review manipulation to invalidate the RD design in our study, the manipulation has to be sufficiently precise such that the true average rating is shifted from the left to the right of a threshold, e.g., from 3.74 to 3.76. Shifting the average rating from 3.74 to 4.24 would not invalidate the RD design in our study, because it is still on the left of a threshold and in the rounded-down group.

[43] See http://bits.blogs.nytimes.com/2012/10/18/daily-report-yelp-fights-fake-reviews-with-shaming/ (accessed on June 1, 2013)

In this study we provide additional empirical evidence to further reduce the concern about restaurants' possible manipulation of Yelp review ratings. If some restaurants could precisely manipulate Yelp ratings and shift their average ratings from the left to the right of a threshold, we expect the aggregate distribution of the distance from threshold would be discontinuous at zero and sorting toward the right of threshold. Figure 4 shows the frequency distribution of the distance from a threshold for all restaurants with distance less than 0.25 (N=2129). The distribution appears symmetric about the threshold; the skewness coefficient is -0.021, far from significantly different from zero (p=0.69). The symmetry of the distribution reduces the concern of restaurants' possible review manipulation. Although there seems to be a peak exactly at the zero point, we provide further evidence (reported in Appendix B) indicating that the peak at zero alone may not necessarily suggest precise review manipulation.

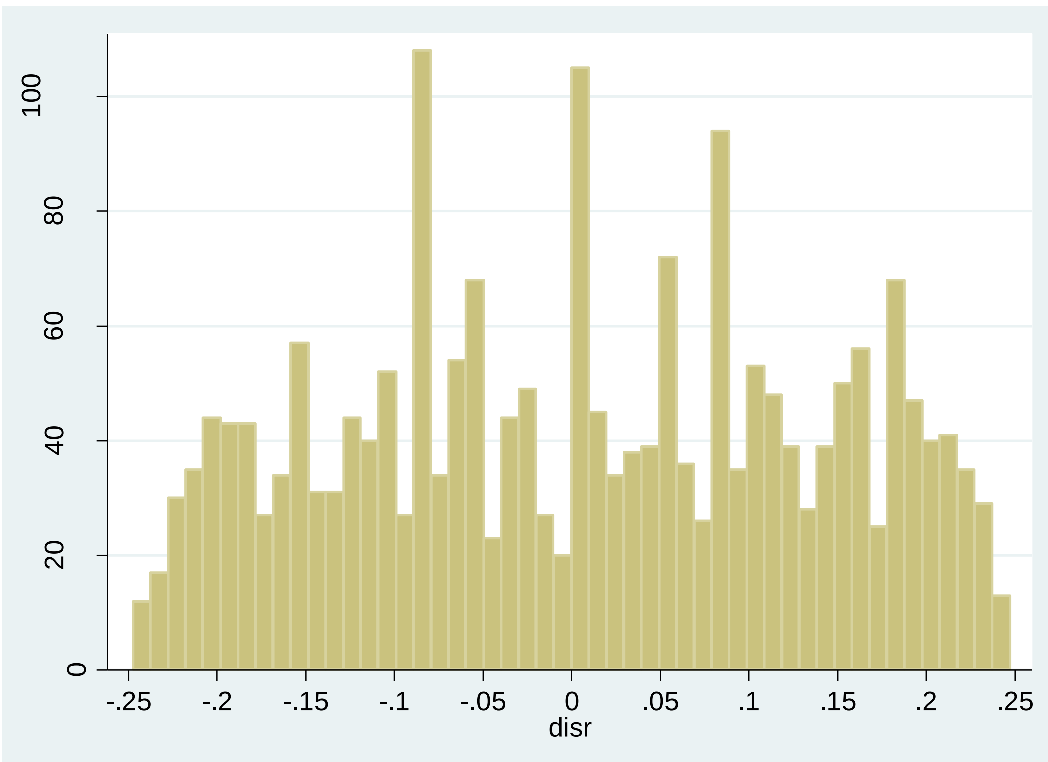

Figure 4: Histogram of the frequency distribution of distance from threshold.

Notes: The histogram plots the frequency distribution of the distance from a threshold for all restaurants with distance<0.25 (N=2129). Note that the distribution appears symmetric about the threshold (skewness coefficient=-0.021, p=0.69), but there is a peak exactly at the zero point.

6.2 Different Bandwidths

A narrow bandwidth would make the RD estimates more convincing in terms of "local randomization" and "local linearity", but it reduces the sample size substantially and may lead to an insignificant estimate even if the true effect exists. On the other hand, a wide bandwidth allows more observations in the analysis but may make "local randomization" or "local linear regression" less likely to be valid. Thus, RD estimates may be sensitive to bandwidth selection (Imbens and Lemieux 2008, Lee and Lemieux 2010). To verify if the estimated effect of displayed Yelp ratings is robust, we choose a number of different bandwidths to analyze the data. If the RD estimates are relatively stable with the selection of different bandwidths, the findings would be more credible.

Using different bandwidths from the smallest (0.05-star) to the widest (0.25-star), Table 10 reports the RD estimates of the effect of displayed Yelp rating on Facebook Likes for restaurants with at least 20 reviews. As Columns (2)-(5) show, the RD estimates are all positive and significant when the bandwidth increases from 0.10 to 0.25. Even though the RD estimate in Column (1) is insignificant with the smallest bandwidth of 0.05 (in this case only 253 observations are used in the analysis), the point estimate is still comparable with those in Columns (2)-(5). Considering the small number of observations used in Column (1), the positive effect on Facebook Likes is likely there. Similarly, Table 11 reports the RD estimates of the effect of displayed Yelp rating on voucher sales for restaurants with at least 20 reviews. All the RD estimates in Table 11 are positive and significant. In sum, the results in Tables 10 and 11 suggest that the findings about the effects of displayed Yelp ratings are robust with the selection of different bandwidths.

6.3 Placebo Effects on Baseline Covariates

Since baseline covariates are predetermined deal and restaurant characteristics, they would not be affected by displayed Yelp ratings. Thus, we conduct another set of robustness checks to test if any placebo effects of displayed Yelp ratings on baseline covariates could be detected by the RD design. Specifically, we perform the same procedure of RD estimation as we have done with the true outcome variables (Facebook Likes and voucher sales), but instead use a baseline covariate as the dependent variable. If the RD estimates of placebo effect on any baseline covariate are detected as significant, it may raise some concern about the credibility of the estimated effect on the true outcomes. Table 12 reports the RD estimates of placebo effects on four different baseline covariates. None of the RD estimates is significant, suggesting that no placebo effect on the baseline covariates is detected. Therefore, the RD design in our study is valid in the sense that it only allows us to detect the effects on the true outcomes.

Table 10. Displayed Yelp Effect on Facebook Likes with Different Bandwidths

	(1) BW=0.05	(2) BW=0.10	(3) BW=0.15	(4) BW=0.20	(5) BW=0.25
Discontinuity $I(r_i \geq c)$	0.228	0.374**	0.313**	0.263**	0.177*
	(0.21)	(0.15)	(0.12)	(0.11)	(0.098)
Distance $(r_i - c)$	-2.93	-2.74	-1.28	-0.939	-0.288
	(6.04)	(2.04)	(1.07)	(0.72)	(0.57)
$(r_i - c) \times I(r_i \geq c)$	6.52	2.28	0.375	0.264	-0.080
	(6.96)	(2.67)	(1.44)	(0.95)	(0.73)
log(Number of reviews)	0.465***	0.587***	0.577***	0.543***	0.512***
	(0.069)	(0.056)	(0.050)	(0.042)	(0.038)
log(Voucher price)	0.016	0.078	0.157**	0.160***	0.134**
	(0.11)	(0.084)	(0.070)	(0.062)	(0.056)
Is a deal from Groupon?	1.31***	1.36***	1.38***	1.35***	1.35***
	(0.12)	(0.091)	(0.074)	(0.067)	(0.063)
log(Restaurant age proxy)	-0.203***	-0.219***	-0.182***	-0.124***	-0.131***
	(0.071)	(0.047)	(0.042)	(0.040)	(0.037)
Promotion duration	Yes	Yes	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes	Yes	Yes
Number of Observations	253	516	764	1017	1250
R^2	0.521	0.455	0.427	0.389	0.365

Notes: Dependent variable is log(*Likes*). OLS estimates of the effect of one extra half-star displayed Yelp rating on the number of Facebook Likes using different bandwidths. All regressions use restaurants with at least 20 Yelp reviews. Robust standard errors are reported in parentheses.

*p < 0.10, **p < 0.05, ***p < 0.01

Table 11. Displayed Yelp Effect on Voucher Sales with Different Bandwidths

	(1) BW=0.05	(2) BW=0.10	(3) BW=0.15	(4) BW=0.20	(5) BW=0.25
Discontinuity $I(r_i \geq c)$	0.313**	0.218**	0.157*	0.174**	0.152**
	(0.15)	(0.11)	(0.092)	(0.082)	(0.073)
Distance $(r_i - c)$	-3.51	-0.891	-0.694	-1.06*	-0.976**
	(4.18)	(1.54)	(0.87)	(0.55)	(0.38)
$(r_i - c) \times I(r_i \geq c)$	0.577	-0.618	0.454	0.708	0.709
	(5.08)	(2.13)	(1.16)	(0.75)	(0.52)
log(Number of reviews)	0.482***	0.477***	0.481***	0.448***	0.430***
	(0.053)	(0.044)	(0.036)	(0.032)	(0.028)
log(Voucher price)	-0.154*	-0.128	-0.037	-0.066	-0.067
	(0.084)	(0.098)	(0.084)	(0.073)	(0.067)
Is a deal from Groupon?	0.387***	0.282***	0.298***	0.276***	0.255***
	(0.094)	(0.077)	(0.062)	(0.055)	(0.049)
log(Restaurant age proxy)	-0.128**	-0.173***	-0.174***	-0.145***	-0.152***
	(0.063)	(0.044)	(0.040)	(0.034)	(0.030)
Promotion duration	Yes	Yes	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes	Yes	Yes
Number of Observations	279	560	824	1087	1336
R^2	0.388	0.315	0.288	0.263	0.259

Notes: Dependent variable is log(*Sales*). OLS estimates of the effect of one extra half-star displayed Yelp rating on the number of voucher sales using different bandwidths. All regressions use restaurants with at least 20 Yelp reviews. Robust standard errors are reported in parentheses.

*p < 0.10, **p < 0.05, ***p < 0.01

Table 12. RD Estimates of Placebo Effect on Baseline Covariates

	(1) log(Number of reviews)	(2) log(Voucher price)	(3) Is a deal from Groupon?	(4) log(Restaurant age proxy)
Discontinuity $I(r_i \geq c)$	-0.116	0.037	0.017	0.124
	(0.082)	(0.059)	(0.045)	(0.086)
Distance $(r_i - c)$	0.328	0.011	-0.516	-1.47***
	(0.51)	(0.38)	(0.28)	(0.53)
$(r_i - c) \times I(r_i \geq c)$	0.056	-0.143	0.661	1.66**
	(0.69)	(0.51)	(0.38)	(0.74)
log(Number of reviews)		Yes	Yes	Yes
log(Voucher price)	Yes		Yes	Yes
Is a deal from Groupon?	Yes	Yes		Yes
log(Restaurant age proxy)	Yes	Yes	Yes	
Promotion duration	Yes	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes	Yes
Number of Observations	1090	1090	1090	1090
R^2	0.280	0.139	0.147	0.183

Notes: OLS estimates of the placebo effect of displayed Yelp rating on baseline covariates. None of the placebo tests on covariates is significant. All regressions use restaurants with at least 20 Yelp reviews and a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

6.4 Alternative Measures for Dispersion of Ratings

A key finding of this study is that the effect of displayed Yelp ratings on Facebook Likes is greater for restaurants with a larger variance. In Section 5.4, we use the notion of unbiased sample variance, but this measure is undefined for restaurants with only one review. Herein, we verify if our findings are robust to a number of alternative measures for dispersion of ratings.

First, since unbiased sample variance is undefined for restaurants with only one review, we use the notion of biased sample variance for them and define it as zero. In such way, the restaurants with only one review (accounting for 3.2% of the full sample) are included in the analysis. Second, entropy, a concept from information theory (Shannon 2001), is an alternative measure of dispersion and uncertainty in a random variable (Ebrahimi et al. 1999). For a discrete random variable X, each possible value x_i is realized with a probability p_i, then the entropy of X is defined as: $H(X) = -\sum_i p_i \log(p_i)$. Entropy is maximized if p_i is equal across all possible realizations, and it is minimized as zero for a deterministic value. In recent literature, entropy has been used for measuring the dispersion of different opinion groups (Dellarocas et al. 2007) and for mining online product reviews (Zhang and Tran 2008). Third, the Herfindahl–Hirschman index (HHI) in the economic literature is a measure of market share concentration and has been used to capture the consensus in movie critics' reviews by summing up the squares of proportions of pro, con, and mixed opinions (Basuroy, et al. 2006). Since HHI is a measure of opinion consensus, we use the inverse of HHI as an alternative measure of dispersion of review ratings.

Table 13 shows the Pearson correlations between the alternative measures of dispersion of ratings. Since we only additionally define the variance of restaurants with a single review as zero, it is not surprising that the augmented variance is perfectly correlated with unbiased sample variance. On the other hand, entropy and inverse HHI are both positively but not perfectly correlated with unbiased sample variance, suggesting that both are meaningful alternative measures of dispersion.

Table 13. Pearson Correlation between Alternative Measures of Dispersion of Ratings

Variable	Mean	S.D.	(1)	(2)	(3)
(1) Unbiased sample variance	1.19	0.63	1.00		
(2) Variance including restaurants with only one review	1.16	0.65	1.00	1.00	
(3) Entropy	1.07	0.34	0.435	0.522	1.00
(4) Inverse HHI	2.82	0.75	0.451	0.525	0.947

Table 14 reports the estimates of the differential effects of displayed Yelp ratings on Facebook Likes using the three alterative measures and different bandwidths. All the estimates of the interaction term between the discontinuity and large-variance dummy $I(r_i \geq c) \times dv_i$ are positive and significant. The results in Table 14 are consistent with Table 7, suggesting that the effect of displayed Yelp rating on Facebook Likes is greater for restaurants with a larger variance of ratings. On the other hand, using these alternative measures of dispersion does not produce any significant differential effects of displayed Yelp ratings on voucher sales, which is also consistent with the results in Table 8. Therefore, we conclude that the dispersion of ratings moderates the effect of displayed Yelp ratings on Facebook Likes, but not on voucher sales.

6.5 Controlling Confounding Factors for Variance of Ratings

The empirical findings that the effect of displayed Yelp ratings on Facebook Likes is greater for restaurants with a larger variance of ratings may result from confounding factors other than variance or dispersion. For example, voucher price may be associated with the variance of ratings and perhaps it is voucher price that results in the moderating effects of the variance, rather than the variance itself. Therefore, we need control the possible confounding factors to reveal that consumers respond to the variance (dispersion) of ratings.

We compare the restaurants with large (dv_i=1) and small (dv_i=0) variances and report the results in Table 15. While restaurants with large and small variances are similar in terms of the number of reviews and restaurant age, the voucher price and true average rating of the two groups are different. Thus, we control these confounding factors by including the interaction terms with them; the results are reported in Table 16. We find that the interaction terms with dv_i (large variance) are all positive and significant. Note that in all columns the point estimates of the interaction terms with dv_i (large variance) are fairly stable, suggesting that the estimates are not biased by the confounding factors (voucher price, true average rating). Therefore, the results in Table 16 suggest that the moderating effect of the variance of ratings on Facebook Likes truly exists, after controlling the observed confounding factors.

Table 14. Displayed Yelp Effect on Facebook Likes Increases When Variance of Ratings is Larger
Using Alternative Measurements of Dispersion and Different Bandwidths

	Including restaurants w/ only one review		Entropy		Inverse HHI	
	(1) BW=0.20	(2) BW=0.15	(3) BW=0.20	(4) BW=0.15	(5) BW=0.20	(6) BW=0.15
Discontinuity $I(r_i \geq c)$	-0.184	-0.110	-0.140	-0.135	-0.101	-0.106
	(0.13)	(0.15)	(0.13)	(0.15)	(0.13)	(0.15)
$I(r_i \geq c) \times dv_i$ (large variance)	0.412**	0.352*	0.374**	0.446**	0.306*	0.403**
	(0.17)	(0.19)	(0.17)	(0.20)	(0.17)	(0.20)
Distance $(r_i - c)$	0.600	-0.506	0.834	0.449	0.725	0.266
	(0.91)	(1.38)	(0.90)	(1.51)	(0.90)	(1.50)
$(r_i - c) \times I(r_i \geq c)$	-1.07	0.247	-1.02	-0.155	-0.999	0.301
	(1.14)	(1.84)	(1.13)	(1.96)	(1.14)	(1.95)
dv_i (large variance)	-0.553***	-0.465***	-0.319**	-0.362**	-0.302**	-0.331**
	(0.14)	(0.16)	(0.14)	(0.17)	(0.14)	(0.17)
$(r_i - c) \times dv_i$	-2.14*	-1.13	-2.56**	-2.91	-2.40**	-2.71
	(1.17)	(1.87)	(1.18)	(1.92)	(1.18)	(1.92)
$(r_i - c) \times I(r_i \geq c) \times dv_i$	2.80*	1.57	2.74*	2.33	2.76*	1.71
	(1.49)	(2.42)	(1.51)	(2.50)	(1.51)	(2.49)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	1732	1289	1732	1289	1732	1289
R^2	0.347	0.359	0.339	0.351	0.339	0.351

Notes: Dependent variable is log(*Likes*). OLS estimates show the effect of displayed Yelp rating on the number of Facebook Likes is greater for restaurants with a large variance/dispersion of ratings. The coefficient estimate of $I(r_i \geq c) \times dv_i$ indicates the difference between the RD estimates of the effects of displayed Yelp rating for restaurants with a large and small variance of ratings. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

Table 15. Comparison between Restaurants with Large and Small Variances

	Mean		Diff. in Means	*t*-statistic
	Large Variance	Small Variance		
log(Number of reviews)	3.38	3.29	0.086	1.58
log(Voucher price)	2.59	2.46	0.13***	5.44
log(Restaurant age proxy)	6.67	6.61	0.058	1.33
True average rating	3.38	3.84	-0.46***	-20.8
Is a deal from Groupon?	0.85	0.83	0.012	0.72
Discount rate	50.82	50.87	-0.050	-0.34

Notes: The comparison indicates that restaurants with large and small variances are different in terms of voucher price and true average rating. *p < 0.10, **p < 0.05, ***p < 0.01

Table 16. Controlling Confounding Factors for Variance of Ratings

	(1)	(2)	(3)	(4)
Discontinuity $I(r_i \geq c)$	-0.184	-0.419	-0.904*	-1.15
	(0.13)	(0.29)	(0.47)	(0.92)
$I(r_i \geq c) \times dv_i$ (large variance)	0.412**	0.403*	0.384**	0.405**
	(0.17)	(0.17)	(0.17)	(0.19)
$I(r_i \geq c) \times$ log(Number of reviews)		0.072	0.054	0.055
		(0.075)	(0.077)	(0.077)
$I(r_i \geq c) \times$ log(Voucher price)			0.222	0.224
			(0.18)	(0.18)
$I(r_i \geq c) \times$ True average rating				0.062
				(0.19)
Other controls	Yes	Yes	Yes	Yes
Number of Observations	1732	1732	1732	1732
R^2	0.347	0.347	0.349	0.349

Notes: Dependent variable is log(*Likes*). The variance of the restaurants with a single review is defined as zero. When additional confounding factors are controlled, the interaction terms with dv_i (large variance) are all positive and significant. All regressions use a bandwidth of 0.2 stars. Robust standard errors are reported in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01

7. Conclusion

7.1 Summary of Findings

Little extant research has studied what factors consumers would take into account in the decision-making of endorsing a product to their peers with social ties. We investigate if and how a seller's online reputation affects consumers' social media endorsements and product sales. We develop a stylized Bayesian learning model to derive the testable hypotheses.

Empirically, we examine the situation in which restaurants with review ratings on Yelp sell deal vouchers through Groupon and LivingSocial. We identify the causal impacts of displayed Yelp ratings on consumers' Facebook Likes and voucher sales of restaurant deals. To establish the causal relationships, we implement a RD design and conduct a number of robustness checks to ensure the validity of the RD design. We find a restaurant's higher displayed Yelp rating increases the aggregate number of Facebook Likes and voucher sales, but only for restaurants with enough reviews. The effects of displayed Yelp ratings decrease and even disappear for restaurants with fewer reviews. More interestingly, we find the effect of displayed Yelp ratings on Facebook Likes is greater when the variance of ratings is larger, but the effect on voucher sales does not significantly change with the variance.

7.2 Implications

Our study yields several important implications for theory and practice.

First, social media endorsements, as an increasingly important indicator of firms' business performance (Aral, et al. 2013), are distinct from product sales, because the motive and cost of endorsing a product are different from purchasing. Therefore, consumers' decision-making of endorsing via social media deserves to be investigated separately. Our empirical findings suggest that online reputation could affect not only product sales, but also consumers' social media endorsements. Our study is perhaps the first to establish the causal relationship between sellers' online reputation and consumers' social media endorsements for commercial products. Our Bayesian learning model provides a plausible theoretical explanation for the mechanism of the effects of online reputation, that is, through signaling product quality and updating consumers' perception of product value. The results suggest that consumers seem to incorporate their perception of product value into their decision-making of endorsing a product to their peers via social media. The results also show that consumers' social media endorsing behaviors can be predicted well by using a simple Bayesian learning model.

Second, we show that the effects of online reputation are moderated by the number and variance of review ratings. Ignoring the moderating role played by the two contextual factors may lead to misleading results. For example, we find the positive effects of displayed Yelp ratings could only be detected for restaurants with enough reviews, but not for those with few reviews. Our results provide a plausible explanation for the seemingly inconsistent empirical findings about the effects of the valence of online reviews (Chevalier and Mayzlin 2006, Chintagunta, et al. 2010, Duan, et al. 2008, Liu 2006). The findings also offer insights on when and for which restaurants the effects of online reputation would be more salient.

Third and more interestingly, our stylized model based on well-established assumptions from the Bayesian learning literature shows that risk aversion makes consumers' posterior expected utility of a product more responsive to the average rating when the product has a larger variance of ratings, whereas the cue diagnosticity theory (Feldman and Lynch 1988) suggests consumers may reduce their reliance on the average rating and become less responsive to it. Consistent with the prediction of the stylized model, our empirical findings show that the effect of displayed Yelp ratings on consumers' social media endorsements is greater when the variance of ratings is larger, in contrast to the predictions from the alternative theories (Basuroy, et al. 2006, Feldman and Lynch 1988, Sun 2012). The results suggest that perhaps consumers are risk averse in endorsing restaurant deals via Facebook. Yet, we find the effect on voucher sales does not significantly change with the variance. The different moderating roles of the variance of ratings on Facebook Likes and voucher sales reveal that consumers exhibit different behaviors in endorsing versus purchasing products. One possible explanation is that perhaps consumers are relatively less risk averse in purchasing products for their own consumption than they are in endorsing to their peers with social ties and the mechanisms expounded by the competing theories may offset consumers' risk aversion in purchasing.

Fourth, our study reveals that the true causal effect of displayed Yelp ratings is more likely to be detected using a valid RD design, while simple OLS regressions without an appropriate identification strategy may produce misleading results. What's more, beyond the prior studies (Anderson and Magruder 2012, Luca 2011), we provide some new procedures (see Sections 5.1 and 6.1) to inspect if restaurants manipulate the review ratings in a way that may invalidate the RD design. Empirical evidence from our inspection does not support that restaurants in our dataset have precisely manipulated their Yelp ratings in this research setting. The procedures for inspection of possible review manipulation that we use in this study can be applied in other contexts of using the RD design.

Last but not least, we show that the average rating, the number and variance of ratings are all important predictors for consumers' responses to restaurant deals. Managers (e.g., restaurant owners, daily-deal sites, and movie studios) may use these simple descriptive statistics of online review ratings in forecasting models for consumers' social media activities and product demand.

7.3 Future Work

We note some limitations of our study for future work. First, the modeling assumption (A2) that the random disturbance in the review signal is normally distributed may not reflect the empirical reality of online review ratings; prior empirical findings show that review ratings often follow a binomial distribution (Hu, et al. 2009) and the mean and variance of review ratings are correlated due to the bounding nature of 1 to 5 stars. Although A2 does not perfectly capture the empirical reality, we use A2 because of its mathematical tractability and believe it does not compromise the key theoretical implications from the stylized model. Future researchers may be able to relax this assumption. The other limitation is that we cannot control unobserved confounding factors for identifying the moderating effect of the variance of ratings, although we find robust results by using alternative measures for dispersion (Section 6.4) and controlling observed confounding factors (Section 6.5). Future work could consider experimental methods to further examine the moderating effect of the variance of ratings.

References

Akerlof, G.A., R.E. Kranton. 2000. Economics and identity. *The Quarterly Journal of Economics* **115**(3) 715-753.

Anderson, M., J. Magruder. 2012. Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database. *The Economic Journal* **122**(563) 957-989.

Aral, S., C. Dellarocas, D. Godes. 2013. Social Media and Business Transformation: A Framework for Research. *Information Systems Research* **24**(1) 3-13.

Aral, S., D. Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* **57**(9) 1623-1639.

Basuroy, S., K.K. Desai, D. Talukdar. 2006. An empirical investigation of signaling in the motion picture industry. *Journal of Marketing Research* **43**(2) 287-295.

Berger, J. 2014. Word-of-Mouth and Interpersonal Communication: An Organizing Framework and Directions for Future Research. *Working Paper*.

Berger, J., C. Heath. 2007. Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research* **34**(2) 121-134.

Berger, J., E.M. Schwartz. 2011. What Drives Immediate and Ongoing Word of Mouth? *Journal of Marketing Research* **48**(5) 869-880.

Byers, J.W., M. Mitzenmacher, G. Zervas. 2012. Daily deals: Prediction, social diffusion, and reputational ramifications. *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM'12)* 543-552.

Chen, H., P. De, Y.J. Hu. 2011. IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales. *Working Paper*.

Chevalier, J.A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345-354.

Ching, A., T. Erdem, M. Keane. 2011. Learning models: An assessment of progress, challenges and new developments. *Working paper available at SSRN*.

Chintagunta, P.K., S. Gopinath, S. Venkataraman. 2010. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science* **29**(5) 944-957.

Dellarocas, C. 2006. Strategic manipulation of Internet opinion forums: Implications for consumers and firms. *Management Science* **52**(10) 1577-1593.

Dellarocas, C., X.M. Zhang, N.F. Awad. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing* **21**(4) 23-45.

Demers, E., K. Lewellen. 2003. The marketing role of IPOs: evidence from internet stocks. *Journal of Financial Economics* **68**(3) 413-437.

Dholakia, U. 2012. How Businesses Fare with Daily Deals as They Gain Experience: A Multi-Time Period Study of Daily Deal Performance. *Working Paper at SSRN*.

Duan, W., B. Gu, A.B. Whinston. 2008. Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems* **45**(4) 1007-1016.

Ebrahimi, N., E. Maasoumi, E.S. Soofi. 1999. Ordering univariate distributions by entropy and variance. *Journal of Econometrics* **90**(2) 317-336.

Eeckhoudt, L., C. Gollier, H. Schlesinger. 1995. The risk-averse (and prudent) newsboy. *Management Science* **41**(5) 786-794.

Efendi, J., M. Kinney, K. Smith, M. Smith. 2012. Marketing Supply Chain Using B2B Buy-Side E-Commerce Systems: Does Adoption Impact Financial Performance? *Academy of Marketing Studies Journal (forthcoming)*.

Egebark, J., M. Ekström. 2011. Like What You Like or Like What Others Like? Conformity and Peer Effects on Facebook. *IFN Working Paper No. 886*

Eliashberg, J., S.M. Shugan. 1997. Film critics: Influencers or predictors? *The Journal of Marketing* **61**(2) 68-78.

Erdem, T., M.P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* **15**(1) 1-20.

Feldman, J.M., J.G. Lynch. 1988. Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology* **73**(3) 421-435.

Friend, I., M.E. Blume. 1975. The demand for risky assets. *The American Economic Review* **65**(5) 900-922.

Hansen, F. 1976. Psychological theories of consumer choice. *Journal of Consumer Research* **3**(3) 117-142.

Hartmann, W., H.S. Nair, S. Narayanan. 2011. Identifying causal marketing mix effects using a regression discontinuity design. *Marketing Science* **30**(6) 1079-1097.

Hu, N., J. Zhang, P.A. Pavlou. 2009. Overcoming the J-shaped distribution of product reviews. *Communications of the ACM* **52**(10) 144-147.

Imbens, G.W., T. Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* **142**(2) 615-635.

Kimball, M.S. 1990. Precautionary Saving in the Small and in the Large. *Econometrica* **58**(1) 53-73.

Kimes, S.E., U.M. Dholakia. 2011. Customer Response to Restaurant Daily Deals. *Available at SSRN 1925932*.

Kosinski, M., D. Stillwell, T. Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* **110**(15) 5802-5805.

Lee, D.S., T. Lemieux. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature* **48**(2) 281-355.

Li, X., L.M. Hitt. 2008. Self-selection and information role of online product reviews. *Information Systems Research* **19**(4) 456-474.

Li, X., L. Wu. 2013. Observational Learning and Social-Network Word-of-Mouth: Evidence from Groupon. *Working paper available at SSRN*.

Liu, Y. 2006. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing* **70**(3) 74-89.

Luca, M. 2011. Reviews, reputation, and revenue: The case of Yelp. com. *Harvard Business School NOM Unit Working Paper* (12-016).

Luo, X., J.J. Zhang, W. Duan. 2013. Social Media and Firm Equity Value. *Information Systems Research* **24**(1) 146-163.

Malhotra, A., C. Kubowicz Malhotra, A. See. 2013. How to Create Brand Engagement on Facebook. *MIT Sloan Management Review* **54**(2) 18-20.

Mayzlin, D., Y. Dover, J.A. Chevalier. 2012. Promotional reviews: An empirical investigation of online review manipulation.

Miller, A.R., C. Tucker. 2013. Active Social Media Management: The Case of Health Care. *Information Systems Research* **24**(1) 52-70.

Moe, W.W., D.A. Schweidel. 2012. Online product opinions: Incidence, evaluation, and evolution. *Marketing Science* **31**(3) 372-386.

Muchnik, L., S. Aral, S.J. Taylor. 2013. Social Influence Bias: A Randomized Experiment. *Science* **341**(6146) 647-651.

Rishika, R., A. Kumar, R. Janakiraman, R. Bezawada. 2013. The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation. *Information Systems Research* **24**(1) 108-127.

Roberts, J.H., G.L. Urban. 1988. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science* **34**(2) 167-185.

Shannon, C.E. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* **5**(1) 3-55.

Sun, M. 2012. How Does the Variance of Product Ratings Matter? *Management Science* **58**(4) 696-707.

Wojnicki, A., D. Godes. 2008. Word-of-mouth as self-enhancement. *HBS Marketing Research Paper*.

Zhang, R., T. Tran. 2008. An entropy-based model for discovering the usefulness of online product reviews. *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* 759-762.

Zhao, Y., S. Yang, V. Narayan, Y. Zhao. 2013. Modeling consumer learning from online product reviews. *Marketing Science* **32**(1) 153-169.

Zhu, F., X. Zhang. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing* **74**(2) 133-148.

Appendix A

Table A.1. Simple OLS Estimates of Displayed Yelp Effect on Voucher Sales

	(1) Full Sample	(2) # of Reviews >=20	(3) # of Reviews <20
Displayed average Yelp rating	-0.104	0.012	-0.308***
	(0.064)	(0.083)	(0.10)
Log(Number of reviews)	0.360***	0.429***	0.266***
	(0.018)	(0.028)	(0.051)
Log(Voucher price)	-0.065	-0.067	-0.062
	(0.047)	(0.066)	(0.063)
Is a deal from Groupon?	0.324***	0.260***	0.477***
	(0.042)	(0.050)	(0.077)
Log(Restaurant age proxy)	-0.110***	-0.147***	-0.089***
	(0.019)	(0.030)	(0.025)
Promotion duration	Yes	Yes	Yes
Dummies of rating ranges	Yes	Yes	Yes
Dummies of cities	Yes	Yes	Yes
Dummies of weekdays	Yes	Yes	Yes
Number of Observations	2126	1336	790
R^2	0.299	0.256	0.254

Notes: Dependent variable is log(*Sales*). Simple OLS estimates of the effect of displayed average Yelp rating on number of voucher sales. Robust standard errors are reported in parentheses.
*p < 0.10, **p < 0.05, ***p < 0.01

Appendix B

Herein, we provide further evidence indicating that the peak at zero alone may not necessarily suggest precise review manipulation. This is because when the number of reviews is 4**k* (*k* is an integer), the average ratings may “naturally” fall exactly at a threshold without any manipulation. For example, for restaurants with four reviews, (5,5,4,1) and (5,5,3,2) both result in an average rating of 3.75. In fact, the combinatorial math suggests that for the case of four review ratings, there are 625 ($=5^4$) different combinations in total, out of which 312 fall at a threshold. That is, even if restaurants are truly randomized and assigned to each of the 625 rating combinations with an equal probability, the aggregate probability that the average rating falls at a threshold is 49.92% (=312/625).[44] By using this simple case, we show that when the number of reviews is 4*k (especially for a small k), the average ratings are very likely to fall at a threshold even without any manipulation.

Figure B.1 shows the frequency distribution of the distance from a threshold for restaurants with number of reviews not as 4*k, called non-4k-type restaurants (N=1631). Note that the peak at the zero point disappears in Figure B.1. Also, the distribution appears symmetric about the threshold (skewness coefficient is -0.017, p=0.78), suggesting that for non-4k-type restaurants (accounting for 76.6%), there seems to be no sorting toward the right of the threshold.

Figure B.2 shows the frequency distribution of the distance from a threshold for restaurants with number of reviews as 4*k, called 4k-type restaurants (N=498). They account for 23.4% of the total restaurants with distance less than 0.25. Consistent with the results from the simple case of four reviews, there is a striking peak at the zero point in Figure B.2. On the other hand, the distribution is also symmetric about the threshold (skewness coefficient is -0.043, p=0.69). Interestingly, given the fact that 4k-type restaurants are likely to fall at a threshold "by nature", restaurants in our dataset are not more likely to be 4k-type, compared to the case in a true randomization (23.4% *vs*. 25%). That is, there seems no evidence that restaurants manipulate the number of reviews to be 4*k so that they could have a higher chance of being rounded up.

In sum, by inspecting the aggregate distribution of the distance from threshold, we find that neither non-4k-type restaurants sort toward the right of a threshold nor restaurants manipulate the number of reviews to be 4*k. The findings suggest that restaurants' incentive to manipulate Yelp ratings is not a concern and enhance our confidence about the validity of the RD design used in our study.

[44] Intuitively, the average of four ratings can only end with 0, 0.25, 0.5, and 0.75. Two of the four endings fall at a threshold, suggesting that the probability is nearly 50%.

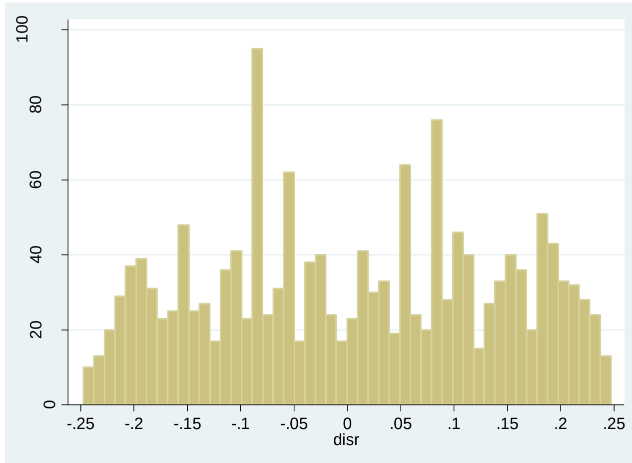

Figure B.1: Histogram of the frequency distribution of distance from threshold, for restaurants with number of reviews not as 4**k* (non-4k-type restaurants).

Notes: The histogram plots the frequency distribution of the distance from a threshold for restaurants with distance<0.25 but the number of reviews not as 4**k* (*N*=1631). Note that there is no peak at the zero point and the distribution appears symmetric about the threshold (skewness coefficient=-0.017, *p*=0.78).

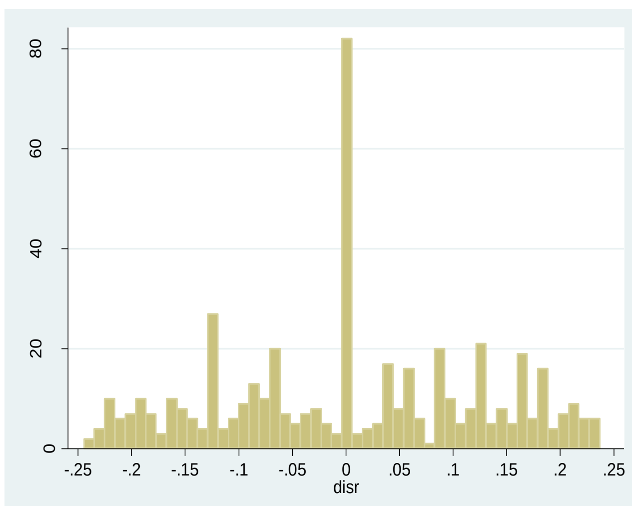

Figure B.2: Histogram of distribution of distance from threshold, only for restaurants with number of reviews as 4**k* (4k-type restaurants).

Notes: The histogram plots the frequency distribution of the distance from a threshold for restaurants with number of reviews as 4**k* (*N*=498). They account for 23.4% of the total restaurants with distance<0.25. Note that there is a striking peak at the zero point and the distribution appears symmetric about the threshold (skewness coefficient=-0.043, *p*=0.69).