NSU
NOVA SOUTHEASTERN UNIVERSITY
NSU Libraries

Nova Southeastern University
NSUWorks

CEC Theses and Dissertations

College of Engineering and Computing

2010

Early Detection of Online Auction Opportunistic Sellers Through the Use of Negative-Positive Feedback

Gregory J. Reinert
Nova Southeastern University, greinert@comast.net

This document is a product of extensive research conducted at the Nova Southeastern University College of Engineering and Computing. For more information on research and degree programs at the NSU College of Engineering and Computing, please click here.

Follow this and additional works at: http://nsuworks.nova.edu/gscis_etd

Part of the Databases and Information Systems Commons

Share Feedback About This Item

NSUWorks Citation

Gregory J. Reinert. 2010. *Early Detection of Online Auction Opportunistic Sellers Through the Use of Negative-Positive Feedback*. Doctoral dissertation. Nova Southeastern University. Retrieved from NSUWorks, Graduate School of Computer and Information Sciences. (285)
http://nsuworks.nova.edu/gscis_etd/285.

This Dissertation is brought to you by the College of Engineering and Computing at NSUWorks. It has been accepted for inclusion in CEC Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

Early Detection of Online Auction Opportunistic Sellers Through the Use of Negative-Positive Feedback

by

Gregory J. Reinert

A dissertation report paper submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Computer Information Systems

Graduate School of Computer and Information Sciences
Nova Southeastern University

2010

We hereby certify that this dissertation, submitted by Gregory J. Reinert, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

_____	_____
Maxine S. Cohen, Ph.D. Chairperson of Dissertation Committee	Date
_____	_____
Francisco J. Mitropoulos, Ph.D. Dissertation Committee Member	Date
_____	_____
Sumitra Mukherjee, Ph.D. Dissertation Committee Member	Date

Approved:

_____	_____
Leonidas Irakliotis, Ph.D. Dean	Date

Graduate School of Computer and Information Sciences
Nova Southeastern University
2010

An Abstract of a Dissertation Report Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Early Detection of Online Auction Opportunistic Sellers Through the Use of Negative-Positive Feedback

By
Gregory J. Reinert

November 2010

Apparently fraud is a growth industry. The monetary losses from Internet fraud have increased every year since first officially reported by the Internet Crime Complaint Center (IC3) in 2000. Prior research studies and third-party reports of fraud show rates substantially higher than eBay's reported negative feedback rate of less than 1%. The conclusion is most buyers are withholding reports of negative feedback.

Researchers Nikitov and Stone in a forensic case study of a single opportunistic eBay seller found buyers sometimes embedded negative comments in positive feedback as a means of avoiding retaliation from sellers and damage to their reputation. This category of positive feedback was described as "negative-positive" feedback. An example of negative-positive type feedback is "Good product, but slow shipping."

This research study investigated the concept of using negative-positive type feedback as a signature to identify potential opportunistic sellers in an online auction population.

As experienced by prior researchers using data extracted from the eBay web site, the magnitude of data to be analyzed in the proposed study was massive. The nature of the analysis required - judgment of seller behavior and contextual analysis of buyer feedback comments – could not be automated. The traditional method of using multiple dedicated human raters would have taken months of labor with a correspondingly high labor cost. Instead, crowdsourcing in the form of Amazon Mechanical Turk was used to reduce the analysis time to a few days and at a fraction of the traditional labor cost.

The research's results found that the presence of subtle buyer behavior in the form of negative-positive type feedback comments are an inter-buyer signal indicating that a seller was behaving fraudulently. Sellers with negative-positive type feedback were 1.82 times more likely to be fraudulent. A correlation exists between an increasing number of negative-positive type feedback comments and an increasing probability that a seller was acting fraudulently. For every one unit increase in the number of negative-positive type feedback comments a seller was 4% more likely to be fraudulent.

Acknowledgements

For their encouragement, understanding, and patience in my pursuit of a belated life-long dream – I would like to sincerely thank my wife, Linda, and my daughter, Caroline.

I would like to express my gratitude to Dr. Maxine Cohen, who has been an exceptional advisor. She was inspiring, supportive, and always helpful. Working with her was both productive and enjoyable. I greatly benefitted from her experience, knowledge, and kindness. Should I ever be in a position to serve as a mentor to a student, I can only hope to emulate her.

The supporting work of committee members is often underappreciated. I extend thanks to my dissertation committee consisting of Dr. Mitropoulos and Dr. Mukherjee. Their comments and suggestions helped in making this dissertation a much more complete and coherent piece of research.

The NSU Alvin Sherman Library staff repeatedly went over-the-top in their assistance to secure copies of exasperatedly elusive or obscure reference papers that I requested. After all this time it is a mystery how they can do it – but, I am grateful for their efforts.

A special thanks to my employer, Eastern Mountain Sports, for their support and flexibility.

Finally, my gratitude to Amazon Web Services for their research grant award which helped fund this effort.

Table of Contents

Abstract iii
List of Figures viii

Chapters

1. Introduction
Problem Statement and Goal 1
Relevance and Significance 11
Barriers and Issues 12
Definition of Terms 15
Summary 17

2. Review of Literature
Reputation Systems 18
Feedback 21
Trust 24
Fraud 25
Textual Analysis 26
Summary 27

3. Methodology
Preface 28
Introduction 28
Research Questions 29
Defining Fraud 30
Research Design 31
Selection of Research Design 32
Limitations of Correlational Research Design 34
Data Collection 35
How Much Data to Collect on Each Seller 37
Data Selection 38
Determining Data Sample Size 41
Data Validity 41
External Validity 42
Internal Validity 42
Reliability 43
Building the Prototype HITs for Amazon Mechanical Turk 44
Sources of the Variables 44
Independent Variable 45
Dependent Variable 46
Data Record Layout 46
Data Obfuscation 47

Identifying Fraudulent Sellers 48
Coding – Identifying Seller Behavior as Honest or Fraudulent 50
Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not 53
Population Size 56
Analysis of the Seller Workload Using Traditional Dedicated Raters 60
Analysis of the Feedback Workload Using Traditional Dedicated Raters 62
Analysis Summary of the Workload Using Traditional Dedicated Raters 62
Introduction to Amazon Mechanical Turk 63
Building the Prototype HITs for Amazon Mechanical Turk 69
Analysis of the Seller Workload Using Amazon Mechanical Turk 76
Analysis of the Feedback Workload Using Amazon Mechanical Turk 78
Analysis Summary for the Workload Using Amazon Mechanical Turk 80
Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not 81
Creating Gold Standard Sellers 81
Creating Gold Standard Feedbacks 83
Implementation of Production Seller HIT for Amazon Mechanical Turk 86
Implementation of Production Feedback HIT for Amazon Mechanical Turk 88
Data Analysis 90
Research Question 1 (RQ1) 91
Research Question 2 (RQ2) 92
Research Question 3 (RQ3) 93
Summary 93

4. Results
Introduction 95
Objective of the Study 95
Data Collection 95
Descriptive Statistics of the Collected Data 97
Amazon Mechanical Turk Processing – Sellers 98
Amazon Mechanical Turk Processing – Buyer Feedback Comments 99
Amazon Mechanical Turk – Quality Control 100
Analysis Delimitations 100
Research Question 1 (RQ1) 101
Research Question 2 (RQ2) 104
Research Question 3 (RQ3) 108
Summary of Results 109

5. Conclusions, Implications, Recommendations, and Summary
Conclusions 112
Research Question 1 (RQ1) 113
Research Question 2 (RQ2) 114
Research Question 3 (RQ3) 115
Limitations 115
Causal Direction 117
Implications and Recommendations 118
Summary 120

Appendixes
A. Methodology Overview 128
B. CSV Data File Schema 129
C. Data Collection Agent 131
D. Evaluator Worksheet 137
E. Coding: Identifying Seller Behavior as Honest or Fraudulent 138
F. Coding: Indentifying Buyer Feedback Comment as Negative-Positive 145
G. Research Qualifications Seller Test 148
H. Research Prototype Seller HIT 158
I. Research Qualifications Feedback Test 167
J. Research Prototype Feedback HIT With Instructions Hidden 169
K. Research Prototype Feedback HIT With Instructions Displayed 170
L. Research Production Feedback HIT 172

Reference List 174

List of Figures

Figures

1. eBay Feedback Score 4
2. eBay Member Feedback Profile 5
3. Example: Seller Using Decoying Deception Tactic in Response to Negative Feedback 8
4. eBay Sample Positive Feedback From Buyer 9
5. eBay sample negative feedback from buyer 9
6. eBay Sample Negative-Positive Feedback From Buyer 9
7. Prior Research on eBay Sales Categories 39
8. Sample Size Formula 41
9. Example Calculation Using Sample Size Formula 41
10. Flowchart of Coder Procedure 54
11. Data Extracts for Category – Computers & Networking: PC Laptops & Notebooks 57
12. Analysis of Data Extract for Week 3 58
13. Analysis of Sellers – Estimated Time and Cost 61
14. Analysis of Buyer Feedback Comments – Estimated Time and Cost 62
15. Sample Human Intelligence Task (HIT) 65
16. Sample HIT Template 67
17. Sample HIT for a Worker 68
18. Simplified and Annotated Example of HIT Results 69
19. Research Qualification Native English Speaker 70
20. Analysis of Sellers – Estimated Cost using AMT 78
21. Analysis of Buyer Feedback Comments – Estimated Cost Using AMT 80
22. Analysis of Extracted Production Data 97

23. Negative-Positive Feedback Comments Predicting Evaluators' Consensus of Seller Fraudulent Behavior 102

24. Comparing Coders Negative-Positive Feedback Consensus to Seller Fraudulent Behavior by Evaluators 104

25. Number of Negative-Positive Feedback Comments Predicting Evaluators' Consensus of Seller Fraudulent Behavior 105

26. Chi-Square on Negative-Positive Feedback Comments Falling into a Cluster 108

Chapter 1

Introduction

Problem Statement and Goal

Willie Sutton the bank robber was asked why he robbed banks, his reported reply was "Because that's where the money is" (Sutton & Linn, 1976). In a similar case of criminals following the money, the 2009 IC3 Internet Crime Report found a 22% percent increase in Internet fraud complaints compared to 2008 (*2009 Internet crime report*, 2010). IC3 reported that monetary losses from Internet fraud increased over 210% rising from $264,600,000 in 2008 to $559,700,000 in 2009. The IC3 report found incidents of online auction fraud dropped to fourth place in the rankings for 2009, but still composed a significant 10.3% of the total monetary complaints. EBay – the largest online auction service – does not publicly release the total number of items listed for auction. A third-party vendor Medved that monitors eBay shows over 4,000,000 new listings per day are added to the over 106,000,000 active lists on eBay website (Medved, 2010). Even with thousands of eBay staff members monitoring the website around the clock; it is not possible to find all the potentially fraudulent auctions and immediately shut them down ("Consumer reports survey of eBay users," 2007).

An opportunistic seller is someone who attempts to negate online auction safeguards and exploit buyers for monetary gain. The exploitation is commonly manifested as

criminal activity in the online auction environment. Specifically it is exhibited in the forms of fraud, theft, and identity stealing (impersonating another user to shield criminal activity). Of these, fraud is the most prevalent (*2009 Internet crime report*, 2010).

Online auctions differ from traditional brick-and-mortar auctions. At a traditional auction, the bidder has a chance to examine the items up for auction. The auctioneer is a live person who controls the bidding. Identity of the bidders, buyers, and sellers is easy to ascertain. Online auctions are vulnerable to fraud more than are brick-and-mortar transactions due to increased information asymmetry between sellers and buyers (Kauffman & Wood, 2000). Online transactions rarely involve face-to-face contact; payment is made before goods can be inspected; repeat transactions between seller and buyer are unusual (Resnick & Zeckhauser, 2002); and no word-of-mouth reputation for the seller is available. Word-of-mouth is the most credible, objective, and influential means for exchanging feedback information and building trust since this type of communication among impartial buyers is unlikely to be biased or profit-driven (Kamins, Folkes, & Perner, 1977).

In order to compensate for these uncertainties, online auctions like eBay have instituted feedback systems that facilitate the collection and dissemination of information about seller past transaction behavior (Dellarocas, 2003a). By making publically available information about sellers' past transactions, an institutional feedback mechanism facilitates buyers' trust and reduce the risk from the community of sellers which enables buyer-seller transactions (Pavlou & Gefen, 2004). It is the culmination of feedback from buyers in prior transactions that builds the seller's reputation in an online auction.

A differentiator between online and traditional auctions is the type of reciprocity used. A traditional auction relies on direct reciprocity as in “I trust you because you were trustworthy with me before.” An online auction relies on indirect reciprocity as in “I trust you because you were trustworthy with others before.” In both cases past trustworthiness is a prerequisite for future transactions. It is the information about reputation that enables trust by inducing a reciprocal response (Dellarocas, 2006; Hendershott, 2006). Any undermining of the provided feedback’s validity or absence of negative feedback distorts the seller’s reputation and potentially exposes future buyers to exploitation by an opportunistic seller.

It is not easy to get feedback from buyers. Research on eBay’s feedback system shows buyers submits ratings on 41.8% to 52.1% of all transactions (Gregg & Scott, 2006; Resnick & Zeckhauser, 2002; Wood, Fan, & Tan, 2002). Buyers may not be motivated to report evaluations or to do so honestly. In a case where the seller’s capacity to provide a service or goods is limited, then it is not in the buyer’s self-interest to make the information public. An example is a serious collector’s reluctance to reveal a source for rare items. Buyers who want to be seen as “nice” may withhold negative evaluations in expectation of reciprocity. A seller’s threats of retaliation for negative feedback combined with explicit or implicit offers of rewards for positive feedback might lead buyers to submit reports that do not accurately reflect their experience. Clearly these factors are in effect as negative feedback for sellers by buyers on eBay occurs in less than 1% of all transactions (Zhang, 2006). This contrasts with the substantially higher fraud rates reported to external entities like the National Consumers League Internet Fraud Watch; suggesting buyers are hesitant to leave negative evaluations ("Watch out for

cyber scrooge this holiday season," 2006). When an eBay buyer does give negative feedback, the seller gives negative feedback 34% of the time which indicates that retaliation may be occurring (Miller, Resnick, & Zeckhauser, 2003).

On eBay for each transaction the buyers and sellers can opt to appraise the other party by leaving feedback. Feedback consists of a positive, negative or neutral rating with an optional short comment ("What is feedback and how does it affect my reputation?," 2010). The ratings are used to determine a member's *Feedback Score*. With some exceptions - feedback works like this:

- A positive rating increases the feedback score by one point.
- A neutral rating leaves the feedback score the same.
- A negative rating decreases the feedback score by one point.

Feedback score is a number used to measure a member's reputation on eBay. A high feedback score means that a member has received a high number of positive ratings from other members. Every member of eBay has a feedback score. It can be found in parentheses next to their eBay userid (see Figure 1). Identifying information was redacted in this and other figures to protect the privacy of the eBay members.

Figure 1. eBay Feedback Score

Clicking the feedback score enables access to a member's detailed *Feedback Profile* (see Figure 2). This includes recent feedback ratings, detailed seller rates, and feedback (rating plus optional comment) for each transaction with other members.

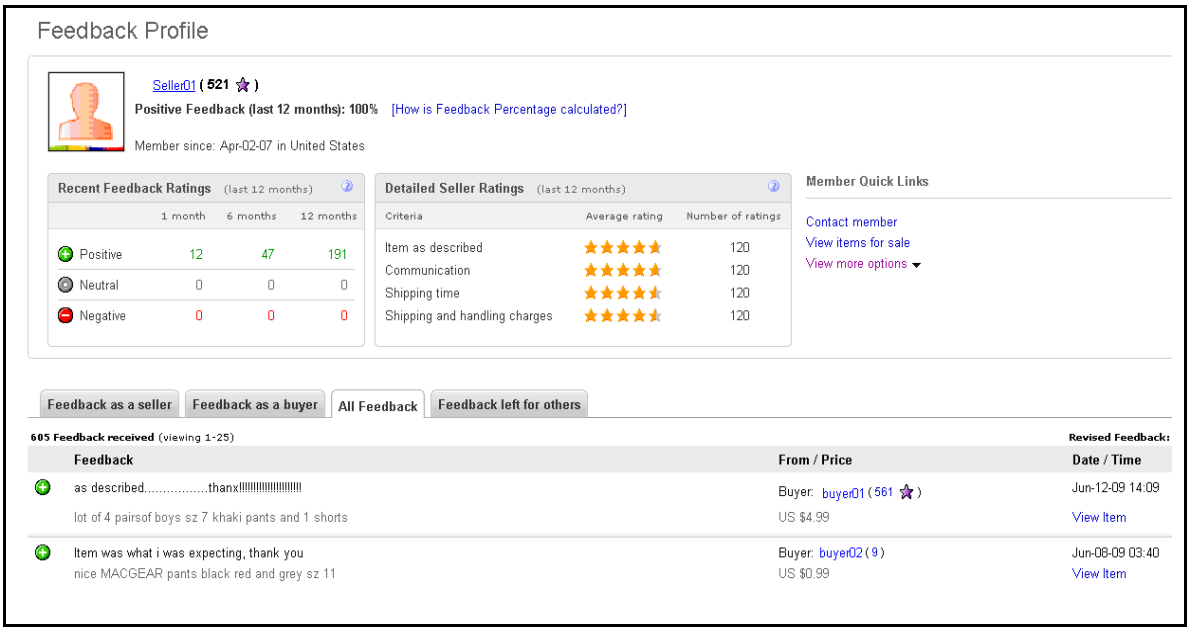


Figure 2. eBay Member Feedback Profile

Feedback is publically viewable immediately after it is posted by either party. Neither party can change a feedback rating after it has been posted. There are very limited circumstances when eBay will consider allowing a change or removing a feedback rating and/or comment based on the *eBay Feedback Abuse* policy ("Feedback abuse," 2010). Sellers and buyers are able to hold feedback hostage by refusing to leave feedback until the opposite party has provided a report. For fraudulent transactions, this behavior could result in false feedback reports or no feedback provided altogether based on fear of retaliation (McDonald & Slawson Jr., 2002). Thus important information to the online auction community about the seller's behavior can be lost.

Inexperienced eBay members are probably oblivious to the threat of feedback-retaliation, but members who are experienced with online trading are sensitive and protective of their reputation. Experienced members consider the possibility of retaliation and take this into account when they make their decision of what feedback type to provide. Other than the possibility of feedback-retaliation exactly why a buyer should

care about feedback is not obvious. It is too simple to assume that buyers remain buyers forever as most eBay participants switch back and forth between the role of buyer and seller. A buyer has to be sensitive to feedback because it may effect future income as a seller. Sellers with higher ratings (better reputations) are able to sell products at a higher price then sellers with lower ratings (Bajari & Hortacsu, 2003). Buyers with a good reputation will not risk finding their bids cancelled due to a low feedback score. Negative feedback can have an adverse effect not just on the seller, but on both parties.

Studying online auction deception is problematic when using conventional methods as with other deviant behaviors the successful perpetrators work hard to avoid detection. The degree of difficulty is compounded by the findings of Zhang (2006) that eBay buyers provide 99% positive comments and 0.7% negative comments. As prior research studies and third-party reports of fraud show rates substantially higher than the 0.7% rate reported, the conclusion can easily be drawn that most buyers are opting to withhold reports of negative feedback. The absence of negative feedback is problematic as it suggests a positive bias in feedback scores. This bias is borne out with the empirical observation that most eBay sellers have a reputation feedback scores that exceed 99%. Therefore analysis of numerical feedback scores for detection of opportunistic sellers is futile.

Building on the base issue of fraud, the research problem statement can be summarized as:

Online auction fraud represents a serious threat to e-commerce and undermines online trust. As fraud is pervasive, growing in use, and difficult to detect in online auctions; new techniques are needed for the early detection of opportunistic sellers.

Excluding the use of feedback scores in online auctions for the detection of opportunistic sellers raised the following issues:

- *Are there other signatures that could potentially identify deception in an online auction transaction?*
- *Can a new method be developed for the detection of opportunistic sellers by utilizing one of these signatures*?
- *For any new signature – What are its limitations and predictive reliability?*

An extensive forensic case study by Nikitov and Stone (2006) focused on modeling the behavior, attributes, and deception tactics of a single opportunistic seller who traded for eight years on eBay. Based on one of the case study findings, the concept of "negative-positive" type feedback appeared to be a candidate for a new signature to detect opportunistic sellers. The viability of the potential new signature along with determining its limitations and predictive reliability needed to be investigated; this investigation served as the premise for the research study.

Because of confidentiality rules, it was not possible for the Nikitov and Stone (2006) to acquire the case study's subject member data directly from eBay. Instead, publicly available data from the eBay website was gathered – transactions, feedback ratings, feedback comments, and seller replies (to buyer feedback comments). This was supplemented by e-mail surveys and follow up interviews with buyers who had interacted with the seller. The researchers acting anonymously in the role of buyer performed multiple transactions purchasing items to collect additional data on the seller's behavior.

Nikitov and Stone (2006) findings confirmed the lack of negative feedback by buyers even after having a problematic or fraudulent experience with an opportunistic seller. The

majority of sellers obliquely or explicitly stated fear of feedback-retaliation (i.e. tit-for-tat) as the reason for not leaving negative feedback on the seller. Several buyers (in escalated situations) indicated that the seller implied retaliation in e-mails should any complaint be made. Buyers that made negative feedback almost universally received retaliatory negative feedback from the seller. The most frequent response employed by the opportunistic seller to any communicated question or complaint was to use deception tactics to disarm, confuse or place the buyer on the defensive (see Figure 3). The deception tactics used were concealment strategies (masking, repackaging, dazzling, and red flagging) and simulation strategies (mimicking, inventing, decoying, and double-play) (Johnson, Grazioli, & Jamal, 1993).

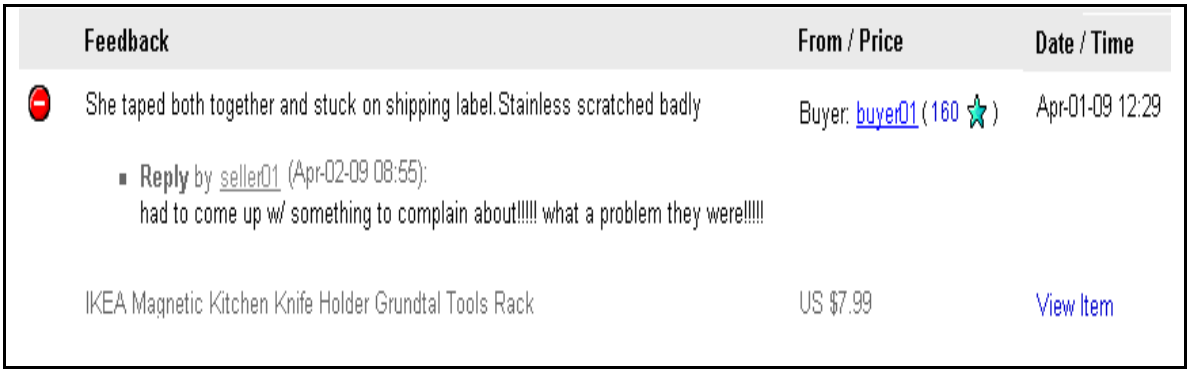
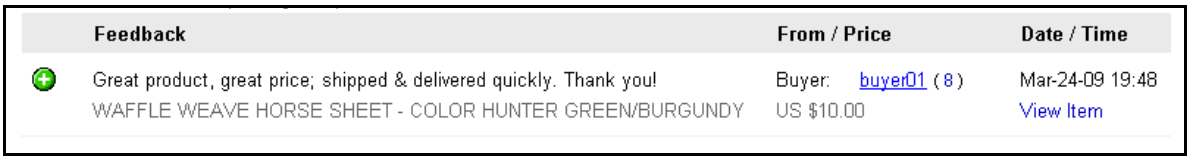


Figure 3. Example: Seller Using Decoying Deception Tactic in Response to Negative Feedback

The most interesting result from the forensic analysis of the opportunistic seller was a new discovery about buyers' feedback data (feedback rating and feedback comment). Nikitov and Stone (2006) found buyers sometimes embedded negative comments in positive feedback as a means of avoiding feedback-retaliation. They referred to this category of positive feedback as "negative-positive" type feedback. The concept of negative-positive type feedback is best understood by viewing a side-by-side comparison of positive, negative and negative-positive examples.

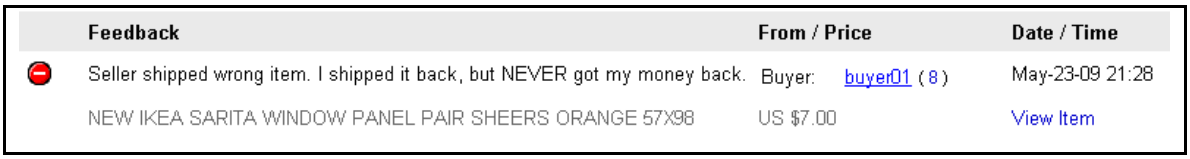
This is an example of a typical *positive* feedback from a buyer (see Figure 4).



	Feedback	From / Price	Date / Time
⊕	Great product, great price; shipped & delivered quickly. Thank you!	Buyer: buyer01 (8)	Mar-24-09 19:48
	WAFFLE WEAVE HORSE SHEET - COLOR HUNTER GREEN/BURGUNDY	US $10.00	View Item

Figure 4. eBay Sample Positive Feedback From Buyer

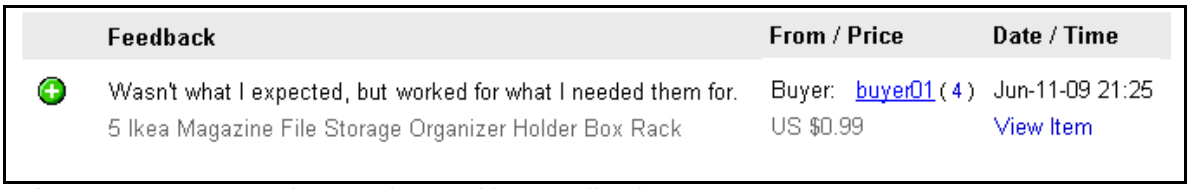
This is an example of a typical *negative* feedback from a buyer (see Figure 5).



	Feedback	From / Price	Date / Time
⊖	Seller shipped wrong item. I shipped it back, but NEVER got my money back.	Buyer: buyer01 (8)	May-23-09 21:28
	NEW IKEA SARITA WINDOW PANEL PAIR SHEERS ORANGE 57X98	US $7.00	View Item

Figure 5. eBay sample negative feedback from buyer

This is an example of a typical *negative-positive* feedback from a buyer (see Figure 6).



	Feedback	From / Price	Date / Time
⊕	Wasn't what I expected, but worked for what I needed them for.	Buyer: buyer01 (4)	Jun-11-09 21:25
	5 Ikea Magazine File Storage Organizer Holder Box Rack	US $0.99	View Item

Figure 6. eBay Sample Negative-Positive Feedback From Buyer

Nikitov and Stone (2006) found that negative-positive feedback postings contained hidden signals to the buyer community about a problematic or fraudulent seller. The composition of negative-positive feedback included both positive and negative aspects of a transaction. Negative-positive complaints were usually in the formats of "I was pleased with X, but unhappy about Y for the transaction" or "I was unhappy about Y, but was pleased with X for the transaction." Typical examples are "Good product, but slow shipping" and "Took 7 days and 2 messages before replying to my email, but product was well packaged."

In their forensic analysis, Nikitov and Stone (2006) viewed negative-positive feedback as a hidden signal to the buyer community about a seller; utilizing feedback content analysis they were able to expose indicators that the seller was potentially problematic or fraudulent. Their research was limited to performing in-depth forensics analysis of a

single opportunistic seller. *The concept of using negative-positive feedback as a signature to identify potential opportunistic sellers in an online auction population was never explored.* This gap provided a narrowly scoped and tightly bounded area for research with a goal of *the early detection of online auction opportunistic sellers through the use of negative-positive feedback.* How to measure the success of using a negative-positive signature for indentifying opportunistic sellers is a little more problematic due to eBay confidentiality rules. The implications of this problem are explored in the *Methodology* chapter along with a verification rationale and implementation techniques.

Feedback-retaliation has been explored as noted in prior citations by a multitude of academic research studies since the inception of eBay in 1995 (Bolton, Greiner, & Ockenfels, 2009; Dellarocas & Wood, 2008; Resnick & Zeckhauser, 2002). The buyers at eBay have been vocal on issues about feedback policies through direct e-mail communication to the company and postings on discussion boards. In January 2008, eBay responded by announcing a fundamental change to the feedback system. Sellers could leave only positive or neutral ratings for buyers. That means buyers were free to leave negative feedback without fear of feedback-retaliation (Ambach, 2008).

Logically, buyers should have responded by providing negative feedback when appropriate. Although the new policy has been in effect for over two years, the status quo remains – eBay still reports less than 1% negative feedback; most members have a 99% or higher feedback rating; and the percentage of fraudulent transactions continues to rise (Gregg & Scott, 2008). Obviously the number of opportunistic sellers is increasing and buyers are still reluctant to provide explicit negative feedback. From this the conclusion can be drawn that buyers are continuing to use negative-positive feedback as a

means to signal the community about potentially opportunistic sellers. Ergo the ideal for early detection of online auction opportunistic sellers through the use of negative-positive feedback remains viable even under the modified feedback system.

Relevance and Significance

Understanding and identifying occurrences of online deception is critical for increasing participation in online auctions and other forms of e-commerce, as victims of fraud will leave the online auction market and potential new customers withhold participation based on fear of becoming a fraud victim (Nikitkov, 2006; Pennington, Wilcox, & Grover, 2003).

Investigating online deception is important as deception in any form is the enemy of trust and some degree of trust is required for all business transactions (Grazioli & Jarvenpaa, 2000). Opportunistic sellers use deception tactics to create an illusion of trustworthiness to the buyers' detriment. A goal of this research study was to help online buyers and online auction vendors to identify sellers who are unworthy of their trust.

According to the Federal Trade Commission, the number of consumer complaints about online auctions has been growing annually. Their latest report indicated that 89% of all Internet fraud complaints filed by the National Consumers League are related to online auctions ("Online auction fraud complaints still rising, says consumer watchdog," 2004). Losses due to fraud in online auctions range in the hundreds of millions of dollars annually. As with most type of frauds, a significant amount of fraudulent activity is never reported by the victims.

The size of the online auction market is immense, but difficult to pin down to a specific figure as many are privately held. An idea of its scale can be drawn from eBay's

2009 SEC Annual filing showing an income of $8,727,362,000 ("Form 10-K for eBay for 2009," 2010). The 2009 IC3 Internet Crime Report found a 22% percent increase in Internet fraud complaints compared to 2008 (*2009 Internet crime report*, 2010). IC3 reported that monetary losses from Internet fraud increased over 210% in the same time period rising from $264,600,000 in 2008 to $559,700,000 in 2009.

Barriers and Issues

No matter how successful the research study for early detection of opportunistic sellers, efforts to deter fraud by developing new detection techniques function like a new military stratagem. The advantage will shift back and forth between the offense and the defense, depending on the adoption of new behaviors and technologies driven by how much each side gains if it wins.

Detection of negative-positive feedback by buyers required the examination, interpretation, and categorization of each buyer's feedback comment text. As natural language communications are variable in form, subject to contextual use, can be incomplete, and prone to errors in spelling and/or grammar; it was necessary to transpose the relevant written text into a formatted and coded structure. A coded structure provides data uniformity and enables automated analysis. Normally, the difficulty is designing an appropriate structure to capture all the components that could be found when performing the contextual analysis (Krippendorff, 1980). In this case, the design of the structure was greatly simplified by use of just two categorical codes. The absence of negative-positive feedback in a buyer's feedback comment text was coded as N (No). The presence of negative-positive feedback in a buyer's feedback comment text was coded as Y (Yes).

Studying online auction deception is problematic as with other deviant behavior the successful perpetrators work hard to avoid detection. An opportunistic seller will employ deception tactics in order to mask his/her behavior and illicit activities. These deception tactics will include the use of concealment strategies (masking, repackaging, dazzling, and red flagging) and simulation strategies (mimicking, inventing, decoying, and double-play) (Johnson, et al., 1993). Although the objective of deception tactics is concealment or misdirection, the presence of deception tactics was used to advantage. The primary mode of communication between buyer and seller in an online auction is via written text. This text can take the form of internal correspondence - feedback comments and replies to feedback; or external correspondence via e-mail. Detecting the seller's usage of deception tactics by examining the written texts provided *corroborating evidence* supporting the identification of a potential opportunistic seller found by using a negative-positive signature. The textual communications were in natural language format with complex overtones and subtle nuances which precluded any easy method for representation in a coded structure. Automated textual analysis currently has limited capabilities and significantly less than a 100% rate of accuracy (Hijikata, Ohno, Kusumura, & Nishida, 2006; Lee, Jeong, & Lee, 2008). Therefore, processing of these types of textual communications required human review and interpretation. Reducing the subjectivity of interpretation required evaluation of each communication by multiple reviewers and creation of evaluation rules for uniform results.

Because of confidentiality rules, it was not possible to acquire data directly from eBay on any members. Mimicking the actions of Nikitov and Stone (2006), publicly available data from the eBay website was gathered – transactions, feedback ratings, feedback

comments, and seller replies (to buyer feedback comments). It was possible to automate the mechanics for the data gathering process by using a spider-like program to crawl the eBay website and extract publicly available data. This technique has already been used successfully by multiple prior researchers (Almendra & Schwabe, 2009; Lucking-Reiley, Bryan, Prasad, & Reeves, 2007; Pavlou & Dimoka, 2006; Zhang, 2006).

In a court of law the degree of difficulty and legal criteria to prove that a specific fraudulent action was performed is less than proving intention as in "intent to defraud." Similarly developing a new method which shifted through prior transactions to identify potentially fraudulent activity that had occurred was significantly easier than attempting to predict fraudulent intent for items being offered in auction. Most opportunistic sellers for practical reasons employ a long-term strategy of exploiting multiple buyers over an extended period of time, rather than use a one-time "take the money and run" strategy (Nikitkov & Stone, 2006). Two practical reasons are the increasing level of difficulty in setting up a new eBay userid and the time required to establish a "good" reputation. In order to deter fraud, eBay has continued to tighten the verification requirements for creation of new eBay userids and has improved detection of attempts to create multiple userids by one person. Therefore the new method took advantage of historical information and was forensic rather than predictive in design. Even when using a forensic method, definitive labeling of an online auction member as an opportunistic seller was not possible. This was because confirmation was not available from the sources with authority – eBay or court rulings. What could be stated was that the specific member exhibited the behaviors and actions characteristic of an opportunistic seller and therefore had a high probability of actually being an opportunistic seller.

Definition of Terms

Amazon Mechanical Turk (AMT) - A crowdsourcing system in which requesters post Human Intelligence Tasks (HITs) then workers do the HITs, submit the results, and receive a small payment ("Amazon Mechanical Turk," 2010).

Buyer – A member who buys an item from a seller using the online auction ("eBay glossary," 2010).

Category Listings – Items are organized by placement into predefined categories, subcategories, etc. Example category: Computers and Networking ("eBay glossary," 2010)

Feedback - For each transaction a buyer/seller can choose to leave an opinion about the other party's performance for the transaction. Feedback is composed of two parts – a rating (Feedback Rating) and an optional text comment (Feedback Comment). A rating can be positive, negative or neutral ("About feedback," 2010).

Feedback Comment – It is part of Feedback consisting of an optional text comment ("About feedback," 2010).

Feedback Profile - A webpage that shows all of a member's information – Feedback Score, Feedback Rating, Feedback Comments, list of items sold, etc. ("About feedback," 2010)

Feedback Rating – It is part of Feedback consisting of a rating which can be positive, negative or neutral ("About feedback," 2010).

Feedback Score - Feedback score is a number (from zero to infinity) used to measure a member's reputation on eBay based on the total number of previous sales or purchases

that were given feedback by the other party ("eBay feedback scores, stars, and your reputation," 2010).

Feedback Type – Also know as Feedback Rating. It can be positive, negative or neutral ("About feedback," 2010).

Feedback-Retaliation - Negative feedback that is left in response to negative feedback from the other party (Nikitkov & Stone, 2006).

Fraud - Any act of deception carried out for the purpose of unfair, undeserved and/or unlawful financial gain. This term has been broadened for the purpose of the study as when the seller imposes a cost on the buyer for which other potential buyers should be aware of when considering purchasing from that seller (author).

Gold Standard Data - Collection of preselected data that have a known set of answers produced by one or more individuals who are trusted and a domain expert (Sorokin & Forsyth, 2008).

Human Intelligence Test (HIT) - A task that a human requester asks a human worker to complete that is simple for a human to do and inherently difficult for a computer to do ("Amazon mechanical turk requester best practices guide," 2010)

Member – A person who has created a profile on the online auction website. A member has a userid and password for providing secure access to the online auction functions like buying or selling, review or leave feedback, or updating personal information ("eBay glossary," 2010).

Negative-Positive Feedback – The use of embedded negative comments in positive feedback by a buyer as a means of avoiding retaliation from the seller (Nikitkov & Stone, 2006).

Online auction – Is a business model in which members bid for products and services over the Internet. Example: eBay (Bajari & Hortacsu, 2004)

Opportunistic seller - A person who attempts to negate online auction safeguards and exploit buyers for monetary gain (Nikitkov & Stone, 2006).

Reputation – The culmination of feedback that a member receives in an online auction (Resnick, Zeckhauser, Swanson, & Lockwood, 2006).

Seller – A member who sells an item using an online auction ("eBay glossary," 2010).

Transaction – Either a sale or purchase made by a member ("eBay glossary," 2010).

Userid - A unique moniker or name used to identify a member of the online auction. Most online auctions allow the person to choice his/her own userid ("eBay glossary," 2010).

Summary

Researching online auction deception is problematic as with other deviant behavior the successful perpetrator works hard to avoid detection. An opportunistic seller will employ deception tactics in order to mask his/her behavior and illicit activities. The research study investigated if the presences of subtle buyer behavior in the form of negative-positive type feedback comments are an inter-buyer signal indicating that a seller is behaving fraudulently.

Chapter 2

Review of the Literature

Reputation Systems

The corpus of this research concentrates on two areas of literature. The first is asymmetric information and second is reputation system design. Asymmetric information is a situation in which the seller knows relevant information about a product that the buyer does not know (Akerlof, 1970). This creates an imbalance of power in transactions which can sometimes cause a transaction to go awry (adverse selection) or make a buyer reluctant to risk engaging in a transaction (moral hazard). Reputation systems are used in online communities when a member has no prior knowledge or experience interacting with another member. In this type of situation, it is often helpful to make a decision whether or not to interact with a member based on the prior experiences of other members. Reputation system design as the name implies is the process of creating appropriate mechanisms to enable a reputation system to function effectively.

Asymmetric information regarding products or sellers has a major impact on market exchange which can result in a market collapsing or failing (Kauffman & Wood, 2000). Reputation and reputation mechanisms play an important role in reducing information asymmetry. These mechanisms facilitate buyer's trust and reduce the risk from the community of sellers which enables buyer-seller transactions (Levine & Martinelli, 1998).

Each of the two asymmetric information models takes into consideration adverse selection and moral hazard. Reputation mechanisms have different roles in each of the two models. For adverse selection the role of a reputation mechanism is in helping the community to learn the initially unknown character (i.e. honesty) of a member (Dellarocas, 2003b). In a moral hazard setting, the objective of reputation mechanisms is promoting cooperative and honest behavior among sellers and buyers by the threat of future punishment (Shapiro, 1983). As Cabral (2004) stated, typical reputation mechanism models that incorporate reputations are based on Bayesian updating of beliefs. In other reputation models trust is modeled through repeated interaction and by the possibility of punishing inappropriate actions in a moral hazard setting (Diamond, 1989).

In an online auction, a reputation system is the primary means to induce sellers and buyers to behave cooperatively. A reputation system's mechanism enables future buyers to condition behavior on a seller's current actions. A reputation system can work as a feasible and less costly substitute for legal enforcement for online auctions (Bakos & Dellarocas, 2003). A reputation system serves as a proxy for the transactional history that would be developed between buyers and sellers over the succession of repeated interactions (Resnick & Zeckhauser, 2002).

Resnick and Zeckhauser (2002) demonstrated the problems of low feed-back rates and potential reporting biases. Based on their work other researchers proposed mechanisms to solve these problems. One technique employed a monitoring mechanism. Ba, Whinston, and Zhang (2002) suggested a Trusted Third-party (TTP) mechanism which entailed issuing certificates to sellers and buyers. Dellarocas (2003b) proposed charging a listing

fee contingent on a seller's announced expected quality then rewarding the seller based on the announced quality compared to the posted rating by the buyer. Both mechanisms were designed to discourage sellers from lying about the true quality of a product.

A second mechanism design attempts to promote honest behavior and facilitate online auction transactions between sellers and buyers through peer-provided feedback. Miller, Resnick, and Zeckhauser (2005) proposed a peer-prediction technique by comparing the likelihood assigned to a reference rater's possible ratings to the reference rater's actual rating. Jurca and Faltings (2007), Papaioannou and Stamoulis (2005), and others proposed reward and punishment systems that induced both sellers and buyers to report truthfully. Two drawbacks to the feedback concept is failure of peers to respond truthfully and positive bias caused by the missing negative feedback as in Dellarocas and Wood (2008).

The third kind of mechanism accounted for the missing reports through a computational mechanism. Dellarocas and Wood (2008) designed a sophisticated computational mechanism to remedy distortions introduced by reporting bias. Their mechanism required buyers to take missing feedback into consideration.

A reputation system must meet three challenges. First, it must provide information that allows buyers to distinguish between trustworthy and non-trustworthy sellers. Second, it must encourage sellers to be trustworthy. Finally, it must have a mechanism to discourage participation from those who are not trustworthy (Resnick, Kuwabara, Zeckhauser, & Friedman, 2000). A number of empirical studies of eBay’s reputation mechanism have been conducted almost entirely focused on buyers’ response to published feedback. Multiple studies have estimated the regression of sale prices based

on seller feedback characteristics. Surveys containing these results can be found in Bajari and Hortaçsu (2004) and Resnick, Zeckhauser, Swanson, and Lockwood (2006). The contributing factor of these studies is their tangential remarks about negative feedback – lack thereof, effects on sellers, effects on buyers, hints about retaliation, etc.

Analyzing eBay's imperfect reputation mechanism has been the subject of much research. McDonald and Slawson (2002) noted that eBay's reputation system revealed only a portion of a member's private information due to some members' unwillingness to provide feedback. eBay members have little incentive to leave feedback once a transaction has been completed and often they do not bother to do so. Members have incentives not to provide negative feedback when appropriate for fear of retaliatory feedback.

Cabral and Hortaçsu (2004) created a basic theoretical model of eBay's reputation mechanism that featured both adverse selection and moral hazard. Their model suggested when in equilibrium a seller's reputation was positively correlated with seller effort (honest sellers rewarded and opportunistic sellers punished). The authors' model also suggested that sellers, specifically opportunistic sellers, had incentives to "buy" a reputation by engaging in purchases rather than sales. Cabral and Hortaçsu also noted that eBay's feedback system though functional was not optimal.

Feedback

Feedback comments from an online auction should be viewed as a narrative-textual representation of a user's reputation. A single feedback type rating cannot capture all the information about a transaction as the impressions of buyers and sellers are typically nuanced. Assume for the moment that there are two buyers - one only moderately

satisfied regarding the purchase and another buyer ecstatic. Both buyers would normally select positive for feedback type because the transaction would have been perceived as positive. The feedback type of positive does not truly capture the essence of the transaction. A better understanding of the experiences of the buyers could be found by examining the text of their feedback comments. For example – moderately happy Buyer A might write "Product OK, but delivery slow." While the ecstatic Buyer B might write "Great product and shipped fast!" With a traditional numerical reputation system, Buyer A and Buyer B would be deemed identical in terms of their purchasing experience which not the case. Research into feedback comments provides insights into online auction transactions (Pavlou & Dimoka, 2006).

Prior research for online auctions include studies focused on the buyer response to published feedback. The interaction of sale price with buyers' feedback types and feedback comments has been reviewed by multiple researchers including McDonald and Slawson (2002), Melnik and Alm (2002), Resnick and Zeckhauser (2002), and Resnick, Zeckhauser, Swanson and Lockwood (2006). Whether quantitative aggregate summary ratings (feedback score), feedback type (i.e. negative, positive or neutral) or feedback comment (detailed text reviews), the consistent recommendation for managing reputation in online auctions is maximize the positive and minimize the negative for feedback type and comments (Melnik & Alm, 2002; Pavlou & Dimoka, 2006; Resnick & Zeckhauser, 2002; Resnick, et al., 2006). It has been shown that negative information has a greater impact than positive information on buyers. This bias of focusing on negative comments and giving much greater weight to negative information in decision making has been well documented (Ofir & Simonson, 2001; Poortinga & Pidgeon, 2004; Weinberg & Davis,

2005). The finding of a negative bias only further emphasizes the importance of feedback in online auctions as feedback types and comments tend to be permanent or very long term. A typical example is eBay's feedback policy which clearly states that feedback ratings and comments are generally a permanent part of a member's Feedback Profile ("About feedback," 2010).

One basic tenet of social psychology is people look to others for guidance in resolving uncertainty in their judgments (Festinger, 1954). Theoretically under the right circumstances individual judgment can be improved by listening to others. One of the most ancient techniques in human society to gather additional information from others is the use of word-of-mouth. Word-of-mouth is the most credible, objective, and influential means for exchanging feedback information and building trust since this type of communication among impartial buyers is unlikely to be biased or profit-driven (Kamins, et al., 1977). Reputation systems incorporate feedback to build artificial word-of-mouth networks in which individuals can share opinions and experiences (Resnick, et al., 2000).

The feedback mechanisms found in the reputation systems are changing people's behavior in subtle but important ways. Based on anecdotal evidence, people are now increasingly relying on opinions posted on reputations systems in order to make decision on selecting an honest seller, financial investments, and entertainment choices (Shirky, 2008). Even if buyers have slightly different understandings of what constitutes honest seller behavior, it is possible to identify a broad set of feedback comments that a majority of buyers would agree conveys honest seller behavior (Pavlou, 2002). Evidence from prior research studies suggest people tend to rely on the opinions of others, even in the presence of their own personal information (Banerjee, 1992). A traditional auction relies

on direct reciprocity as in “I trust you because you were trustworthy with me before.” An online auction relies on indirect reciprocity as in “I trust you because you were trustworthy with others before.” In both cases past trustworthiness is a prerequisite for future transactions. It is the information about reputation that enables trust by inducing a reciprocal response (Dellarocas, 2006; Hendershott, 2006).

Trust

Trust is an essential element in forming and maintaining commercial relationships (Nah & Davis, 2002). Trust is particularly challenging to develop in an online context like an online auction (Cofta, 2006). The converse of trust in the online auction environment is fraud. As a result trust and fraud have become important topics in online auction research. Lansing and Hubbard (2002) and Albert (2002) examined possible techniques to mitigate fraud through regulation. Bywell and Oppenheim (2001) recommended bidders be more aggressive in pursing fraud complaints against sellers. While fraudulent behaviors like competitive shilling, reserve price shilling, buy-back shilling, and false bidding have been investigated by researchers like Kauffman and Wood (2005) and Dong, Shatz, and Xu (2009).

For online auctions, trust translates to a good reputation in the form of positive feedback ratings and feedback comments. A seller’s poor reputation can deter buyers from participating in an auction (Brinkmann & Seifert, 2001). There is conflicting research results on the effect of reputation on price paid. Melnik and Alm (2002) and Ba and Pavlou (2002) showed a correlation of reputation score increasing with the price paid by a seller. While the latest research from Kauffman and Wood (2006) could not find any

significant effect of reputation on price. The conclusion is that there are other unknown factors which are increasing or reducing the effect of reputation on price.

Fraud

The number of people being victimized by deceptive practices over the Internet continues to rise (Grazioli & Wang, 2001). Auction fraud is a problem that has been getting increasingly serious. The anonymity provided by online auctions may be fostering deception as the deceiver is able to disassociate himself/herself from the deceiving message (Bowker & Tuffin, 2003). On the Internet, high anonymity is possible making it difficult to assess identity and accountability regarding deception. On average the number of Internet frauds grew more than 250% annually (Grazioli & Jarvenpaa, 2003). The Internet Crime Complaint Center (IC3) which was created by the Federal Bureau of Investigation and the National White Collar Crime Center has received an increasing number of complaint submissions each year.

The limited research that has been conducted has been unable to suggest systematic approaches in detecting or preventing online auction fraud. Some researchers have categorized online auction fraud into different types, but they have not constructed any formalized methods to deal with them (C. Chua & Wareham, 2004). Work has been done in other research areas related to online auction fraud detection - reputation systems (Melnik & Alm, 2002; Resnick, et al., 2000; Resnick, et al., 2006), graph mining (Zacharia, Moukas, & Maes, 1999), and trust (Gyongyi, Garcia-Molina, & Pedersen, 2004).

Research into feedback text comments is arguably more important than aggregate feedback ratings or scores because it can provide greater insights into the behavior and

character of sellers and buyers. However, it is only recently that research has been undertaken specifically on feedback text comments and their impact on reputation systems in online auctions (Bolton, Katok and Ockenfels 2004; Bolton, Loebbecke and Ockenfels 2008; Dellarocas 2003; Resnick and Zeckhauser 2002; Resnick et al. 2006). Consumers read and place significant weight on detailed reputation system elements in the feedback text comments found in a seller's reputation feedback (Weinberg & Davis, 2005). This finding was supported by Pavlou and Dimoka (2006) who reported that buyer feedback text comments in online auctions had a greater impact on a seller's credibility and benevolence than did aggregate "crude numerical" measures. They advised online auction members to attract outstanding (i.e. extremely positive) feedback text comments to avoid receiving abysmal (i.e. extremely negative) feedback text comments. These research studies confirm the importance of feedback text comments and provide supporting evidence on the continued use, role, and value of negative-positive feedback comments to buyers in online auctions.

Textual Analysis

Contextual analysis is a systematic method for analyzing data in a standardized way (Weber, 1990). Contextual analysis can be applied to classify key ideas in any communication media – written, audio, and visual. The term *textual analysis* is used when contextual analysis is applied to written communication. What makes the textual analysis technique powerful and effective is its use of coding and categorizing of the data (Krippendorff, 1980). Coding is the marking of words or text passages with alphanumeric codes. The codes are used to create categorical variables representing the original textual information. The resulting categorical variables can be analyzed using standard statistical

methods. One problem experienced by prior researchers working with feedback comments was finding a technique to extract nuances, inferences, and information from the provided textual data. The technique of choice by prior researchers to solve this problem was textual analysis.

Summary

For online auctions a feedback system is the reputation mechanism used to facilitate buyer's trust and reduce the risk from the community of sellers which enables buyer-seller transactions. Identifying online deception is important as deception in any form is the enemy of trust and some degree of trust is required for all business transactions. Opportunistic sellers use deception tactics to create an illusion of trustworthiness to the buyers' detriment. The problem is that identifying sellers that exhibit fraudulent behavior is difficult as they constitute only a very small percentage of the entire online auction population and are elusive adapting their behavior to avoid detection. The issue with online auction fraud is further compounded as number of occurrences and resulting monetary losses has increased every year. As a result trust and fraud have become important topics in online auction research.

Research in online auction fraud is primary based on three methodologies – economic modeling, legal analysis, and analyses of online auction lists (Wood, 2004). Analyzing the feedback ratings and comments provided by buyers on a seller in online auction lists is a common track taken by many prior researchers. Each succeeding group of researchers has applied ever more varied and sophisticated techniques using the feedback ratings and comments provided by buyers and sellers to analyze user interactions, user behavior, and attempt to identify potentially criminal activity.

Chapter 3

Methodology

Preface

When conducting the research study, the initial plan was to use dedicated raters as evaluators and coders for the duration of the work to be done. Based on best practices, a pilot test was performed to estimate the time required and cost of performing the research. The results of the pilot test indicated using the traditional method of dedicated raters was not viable due to the excessive time of 175 days and estimated cost of $37,152. An alternative method of crowdsourcing was found, determined to be viable, and used to perform the required work for the research study.

The pilot test was based on the methodology details for the dedicated raters. Significant portions of the alternative method of crowdsourcing were based on the methodology details for the dedicated raters. As a result, the Methodology chapter contains details for both methods which are referenced accordingly as “initial plan” for using dedicated raters and “alternative plan” for crowdsourcing using Amazon Mechanical Turk.

Introduction

Research in online auction fraud is primary based on three methodologies – economic modeling, legal analysis, and analyses of online auction lists (Wood, 2004). The research

study focused on analyzing the feedback comments provided by buyers on a seller in online auction lists. The objective of the research was to determine if the presence of negative-positive type feedback comments by buyers is a predictor that a seller is behaving fraudulently. A diagram showing an overview of the research methodology can be found in Appendix A.

Research Questions

The research study focused on the determining if negative-positive type feedback comments by buyers are a predictor that a seller is behaving fraudulently. Three research questions were used in framing an answer for this primary question.

There is a need to determine if the presence of negative-positive type feedback comments by buyers is a predictor that a seller is behaving fraudulently:

Research Question 1 (RQ1): Does negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

If the presence of negative-positive type feedback comments by buyers is a predictor that a seller is behaving fraudulently per RQ1, then need to determine if the number of negative-positive type feedback comments found for a given seller is a basis for the strength of the predictive relationship. The form of the predictive relationship could be linear or non-linear:

Research Question 2 (RQ2): Does the number of negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Any patterns found in the presence of negative-positive type feedback comments in a seller's transaction history could provide additional insights into seller and/or buyer

behavior; or be used to augment the accuracy of negative-positive type feedback comments as a predictor per RQ1:

Research Question 3 (RQ3): For each seller will negative-positive type feedback comments from buyers fall into a pattern?

Defining Fraud

This raises the question – What is fraudulent? The definition of fraudulent per the Merriam-Webster dictionary is "characterized by, based on, or done by fraud" ("Merriam-Webster's collegiate dictionary," 2005). The online auction company eBay defines “fraud” as the seller’s failure to deliver the sold merchandise or the delivery of the item in physically bad condition ("eBay buyer protection plan," 2010). For this research study, fraud was defined in broader terms than eBay does. The terms "fraudulent" and "problematic” transaction were used interchangeably as any breach of the *eBay User Agreement* (contract) that comes at a cost to the buyer ("Your user agreement," 2010). If the seller ships an item later than agreed upon without reimbursing the buyer for the delay, late shipping constituted fraud. If merchandise differs from the item’s auction description in make, model or condition (i.e. used vs. new), the seller committed fraud. If the seller does not explicitly state that the item is not genuine (i.e. a copy), the seller committed fraud. If any deficit attributes of the product are not explicitly stated (i.e. headphones with a six-inch cord rather than the standard three to six foot cord), the seller committed fraud. A fraudulent transaction does not exclusively mean that a seller collected the buyer’s money and then failed to ship the item. *Fraud was viewed as the seller imposing a cost on the buyer for which other potential buyers should be aware of when considering purchasing from that seller*.

The logic behind broadening the definition of fraud as committed by an opportunistic seller becomes obvious once the content of feedback comments and anecdotal evidence of postings on eBay's discussion boards are reviewed. Although a seller's action may not be a breach under the legal terms of the *eBay User Agreement*, buyers have shown that they are sensitive to any questionable action (or lack of action) by a seller. Broadening the definition of fraud was also supported as most complaints filed with the FTC as Internet auction fraud report problems are with sellers who fail to send the merchandise; send something of lesser value than advertised; fail to deliver in a timely manner; or fail to disclose all relevant information about a product or terms of the sale ("Online auction fraud complaints still rising, says consumer watchdog," 2004). Similar to Nikitov and Stone (2006), the preliminary evaluation of feedback comments and postings on eBay's discussion boards indicated that buyer complaints could be categorized as – product, shipping, communication, and other (non-specific).

Research Design

The research study implemented a correlational research design using an automated data collection agent (Creswell, 2002). The research study required the extraction and analyzing of data that met predefined qualifying conditions from immense data sets. Manually sifting through data sets of this magnitude was not practical due to the time and labor required to extract the qualified data. Instead customized software in the form of an automated data-collection agent was used to search, locate, and extract the qualified data from the data set (Allen, Burk, & Davis, 2006). The objective of the research was to determine if the presence of negative-positive type feedback comments by buyers (independent variable) is a predictor that a seller is behaving fraudulently (dependent

variable). The correlational research design provided for discovering relationships between variables, measuring the degree and direction of relationships, and from the discovered relationships predictions could be made.

Selection of Research Design

There are two basic types of research - experimental research and non-experimental research. Each type of research answers different research questions and uses different research designs to collect data (Creswell, 2002). Experimental research designs are composed of true experimental and quasi-experimental. Non-experimental research designs are composed of observational and correlational. The selection of the non-experimental correlative research design was primarily due to constraints which eliminated alternative research designs.

In the research study, a true experiment would violate ethical standards. The researcher wanted to determine if a buyer will leave negative-positive type feedback comments as an indicator that an opportunistic seller had behaved fraudulently. In the hypothetical true experiment, one would start with a sample population of sellers and divide them randomly into a treatment group (asked to make only fraudulent sales) and a control group (asked to make only honest sales). After a period of time making sales to the unaware buyers, the researcher would conduct a review of the buyer feedback comments for both seller groups. Needless to say, such an experiment would violate common ethical principles and criminal statutes.

A quasi-experimental design is one that looks like a true experimental design but lacks the key ingredients of manipulation and random assignment. The most commonly used quasi–experimental design is non-equivalent groups design. Due to the source of data

(extracted from website pages) and type of data (historical transaction logs), it is impossible to perform the required pre-test, treatment, and post-test for this research design. Other researchers like Bajari and Hortacsu (2005), Brown, Forin and Rhodes (2009), and Kauffman and Lee (2009) have used crawlers to collect data from website pages, performed online auction focused research, and explicitly declared their research design as quasi-experimental. However, upon closer examination the term quasi-experimental could only be loosely applied as all the required components – pre-test, treatment, and post-test were not present.

Non-experimental designs are used to describe, differentiate, or examine associations, as opposed to direct relationships, between or among variables, groups, or situations. There is no random assignment, control groups, or manipulation of variables, as these designs use observation only. The most common non-experimental designs are observational and correlational studies.

The observational design is based on gathering detailed information about behavior. Typically this is done by direct or indirect visual observation by the researcher of the study subjects. As the data source was website pages and type of data was historical transaction logs, there was no observable behavior rendering this research design moot.

A correlational research design focuses on investigating the existence and the degree of a relationship between two or more quantitative variables. If two variables are highly related, values of one variable could be used to predict values on other variable. The objective of the research was to determine if the presence of negative-positive type feedback comments by buyers is a predictor that a seller is behaving fraudulently. The definition and functionality of the correlational research design made it the optimum

choice for the research study. Selecting and combining the correlational research design with a data collection agent provided another advantage as analyzing only a subset of all available data increases the validity of the resulting conclusions, provided that the subset of data is based on tightly defined and narrow conditions. The extracted subsets can provide evidence for stronger conclusions regarding causality than uninformed analysis of the entire data set (Creswell, 2002).

Limitations of Correlational Research Design

The correlational research design does provide the ability to detect patterns or relationships among variables (i.e. Is X related to Y?). Relationships between variables are discovered through the use of correlational statistics. These relationships could be linear or non-linear in form. The correlation coefficient can provide a measure of the degree and direction of relationship. From the discovered relationships predictions can be made.

Correlational research design will not identify the causes or reasons for the observed behavior. This is because a correlational relationship between variables could be the result of an outside source. Based on this possibility, it must be understood that the correlation does not necessarily explain cause and effect. Hence the maxim – Correlational does not equal causation (Aldrich, 1995).

Under certain conditions, it may be possible to have a high degree of confidence that there is causality between two variables. Determining the direction of causality can be difficult or impossible to quantify. Casual direction can be hinted if information about time is available. This is because a cause must precede its effects under classic Newtonian physics and natural laws. The type of data to be used is time-stamped

historical transaction logs which provide the possibility of indicating the direction of causality.

Data Collection

Prior published research investigating online auction fraud generally started by identifying two groups of sellers based on their historical behavior pattern - fraudulent or honest. In the research study, the sample population was obtained by using an automated data collection agent crawling over the eBay website.

The optimum means to secure data for research would be having it directly supplied by the company which is the source for the study – eBay. Unfortunately, eBay will not provide data upon request to researchers. Prior researchers have also experienced this problem and resorted to either manually collecting the data or using an automated data collection agent (i.e. web crawler or spider).

An Application Programming Interface (API) is an interface implemented by a software program to enable interaction with other software or a website. It is not uncommon for commercial companies to provide APIs to allow other companies to interact with their website for product availability, pricings queries, place purchase requests, etc. Providing APIs allows the target company to control access, optimize usage, and throttle dataflow. APIs are a recently available option for eBay, but have several issues that precluded their use ("Advanced research API," 2010). Although the eBay API software is free, usage based on number of API calls is metered and charged appropriately. The second issue is that the eBay APIs are limited in functionality as to what data can be retrieved. As the data collection process would require hundreds of

thousands of API calls and possibly need to be repeated multiple times, the cost would be prohibitable.

The option most frequently chosen by prior researchers like Bapna, Goes, Gupta and Jin (2004); Clemons, Hann and Hitt (2002); Easley and Tenorio (2004); Palmer (2002); and Pavlou and Gefen (2004) was using a web crawler. A web crawler is a software program that accesses a website and traverses through the site by following the links present on the web pages. Although commercial web crawlers are available, their cost and limited functionality forces most researchers to build a custom web crawler.

The custom automated web crawler used in this research study was written in Java object-oriented programming language. The web crawler was specifically designed for the eBay website to retrieve web pages, parse the webpage to find the required data, determine if the found data met the selection criteria, and store the qualified data for later analysis in a Comma Separated Variable (CSV) ASCII file. Details on the web crawler design for this research study can be found in Appendix B.

There are distinct advantages and disadvantages to using an automated data collection agent compared to performing the task manually. An advantage of using an automated data collection agent is the reduction of human error in the data collection process. Agents collect more qualified data in a significantly shorter period of time then possible manually. One disadvantage is that large quantities of superfluous or irrelevant data can be collected – this was avoided by defining very specific constraints for qualifying data. Constructing a custom automated data collection can be a complex and time consuming programming task depending on the data to be collected and the dispersion of data over multiple linked web pages. There are potential legal issues of copyright in collecting data

(Winn, 2005) or having an agent cause the equivalent of denial-of-service attack on a website due to its processing demands (Mierzwa, 2005). eBay has pressed legal suits against commercial companies for using automated data collection agents, but to date has not restricted personal or research based use of automated data collection agents ("eBay, Inc. v. Bidder's Edge, Inc," 2000).

How Much Data to Collect on Each Seller

Prior research shows that recent feedback is the most influential on online auction buyers and also indicated buyers rarely examine feedback text comments beyond the first webpage (Dellarocas, 2003b). Nikitkov and Stone (2006) found that opportunistic sellers for practical reasons employ a long-term strategy of exploiting multiple buyers over an extended period of time. Based on these two behavior patterns, it should be possible to predict buyers will repeatedly be "caught" by opportunistic sellers as evidence of previous fraudulent actions are "hidden" from any prospective buyer's view as they roll off the first webpage. From this it could be surmised that one characteristic for identifying a typical opportunistic seller is multiple occurrences of negative-positive feedback in his/her transaction history. Extrapolating on above suppositions, the multiple occurrences of negative-positive type feedback comments should result in a "bunching" or "clustering" pattern. The actual existence of a pattern and its construct was investigated per the previously stated RQ3: *For each seller will negative-positive type feedback comments from buyers fall into a pattern?* The forensic method of the research study required the examination of a seller's entire transaction history in order to identify any pattern. Therefore, the entire transaction history was collected for each qualified seller.

Data Selection

The correlational research design provides for discovering relationships between variables, measuring the degree and direction of relationships, and from the discovered relationships predictions can be made. In the case of the research study – it was used to explore if there is a relationship between the presence of negative-positive type feedback comments by eBay buyers (independent variable) and eBay sellers identified as behaving fraudulently (dependent variable). The collected data was separated into two groups based on the characteristic of the seller's behavior - honest and fraudulent.

The data sets behind the eBay website contain immense quantities of data currently reported to exceed two petabytes ("eBay's two enormous data warehouses," 2010). The most recent numbers for eBay are from 2009 and show active registered users currently total 90,000,000 ("Form 10-K for eBay for 2009," 2010). Combine this with the fact that fraudulent sellers constitute a minuscule number of the active registered users, raises some obvious questions. What size sample population is needed? How can the probability be increased that the sample population includes multiple fraudulent sellers?

In order to build a sample population that contains sellers that behave honestly and fraudulently, choosing the sellers randomly would not work as the probability of finding even a single seller that behaves fraudulently (i.e. opportunistic seller) would be very small. Exactly how small can been seen by the 0.01 percent officially reported by eBay (B. Cox, 2003; Konrad, 2005). The number only rises to 0.20 percent based on a research study of eBay fraud by Gregg and Scott (2008). Nor does it appear that fraudulent sellers are evenly distributed across the thousands of sales categories available on the eBay

auction site. The distribution of fraudulent sellers appears to be skewed and focused on specific categories.

Prior researchers have determined which specific eBay sales categories have the highest incidents of fraudulent sales (See Figure 7). Of particular interest is the *Computers and Networking: PC Laptops and Notebooks* category where one research study found three-quarters of the survey respondents did not receive their computer or it arrived damaged (Gavish & Tucci, 2008).

Category	**Researchers**
Baseball Cards	Jin, G. and A. Kato (2008). "Price, quality and reputation: Evidence from an online field experiment." *RAND Journal of Economics* 47(4): 983-1005.
Camcorders	IFCC (2001). "Internet Auction Fraud. Washington, DC: National White Collar Crime Center and Federal Bureau of Investigation, 2001, ." Retrieved 6/1/10, 2010, from www.ifccfbi.gov/strategy/AuctionFraudReport.pdf.
Coins	Finch, B. J. (2006). "Customer expectations in online auction environments: An exploratory study of customer feedback and risk " *Journal of Operations Management* 25(5): 985-997.
Computers and Networking: PC Laptops and Notebooks	Gavish, B. and C. L. Tucci (2008). "Reducing Internet auction fraud." *Communications of the ACM* 51(5): 89-97. Finch, B. J. (2006). "Customer expectations in online auction environments: An exploratory study of customer feedback and risk " *Journal of Operations Management* 25(5): 985-997. IFCC (2001). "Internet Auction Fraud. Washington, DC: National White Collar Crime Center and Federal Bureau of Investigation, 2001, ." Retrieved 6/1/10, 2010, from www.ifccfbi.gov/strategy/AuctionFraudReport.pdf.
Jewelry	Gavish, B. and C. L. Tucci (2008). "Reducing Internet auction fraud." *Communications of the ACM* 51(5): 89-97.

Figure 7. Prior Research on eBay Sales Categories

Which raises the question - Why is the skewed distribution of fraudulent sellers of any interest? A brief analogy will help answer this. Imagine hunting for a single needle in a very large haystack. Odds are you either will not find the needle or have to invest considerable time and effort to find it. How can you improve your odds of finding a needle? The optimum answer requires adopting two strategies. First – search a smaller haystack that purportedly has a needle in it (i.e. reduced solution space). Second - increase your odds by finding a smaller haystack purportedly with multiple needles in it (i.e. increase probability). Substituting needle with fraudulent seller and haystack with sample population, the solution becomes obvious. Target the data selection process on

extracting a sample population from a given eBay sales category that has been demonstrated to contain a high number of potentially fraudulent sellers. For the research study, the targeted eBay sales category used was *Computers and Networking: PC Laptops and Notebooks*.

Dr. Floyd, a fictional character in the book *2010: Odyssey Two* by Author C. Clark (1983), said "Once is an accident; twice is a coincidence; three times is a conspiracy." Based on a similar sentiment, one final step needed to be done to refine the data selection process. A seller with a sales history showing a single sale in the *Computers and Networking: PC Laptops and Notebooks* category was more likely cleaning out a closet rather than engaging in fraud. Repeated sales transactions by a seller in the category demonstrate the difference between a casual seller and being in the business of selling laptops either legitimately or fraudulently. A seller needs to have a track record in the form of a sufficiently sized feedback history to provide for an accurate categorization of the seller's behavior as honest or fraudulent. Using the same initial data selection criteria as that of Finch (2006), the initial plan was for sellers with a feedback score lower than 600 be excluded. A feedback score of 600 means that a seller had a minimum of 600 sales in all categories, but given the feedback response rate of 48.9% to 59.2% will have a higher actual number of sales (Gregg & Scott, 2006; Resnick & Zeckhauser, 2002; Wood, et al., 2002).

Should the resulting retrieved population size proved too small compared to the required data sample size, the initial plan was to rerun the automated data collection agent after adjusting the feedback score threshold filter. This process would be repeated

as often as necessary until an appropriately sized data sample population size was obtained. See Appendix A for a diagram of the research methodology.

Determining Data Sample Size

The initial plan's sampling method was representational. Yamane (1967) provides a simplified formula to calculate sample sizes (See Figure 8). Where *n* is the sample size, *N* is the population size, and *e* is the level of precision:

$$n = \frac{N}{1 + N(e)^2}$$

Figure 8. Sample Size Formula

An example of how this sample size formula would be used is shown in Figure 9. For demonstration purposes, let it be said that 2000 unique sellers were found listed in the *Computers and Networking: PC Laptops and Notebooks* classification. A 95% confidence level and $p = 0.5$ are assumed.

$$n = \frac{N}{1 + N(e)^2} = \frac{2000}{1 + 2000(.05)^2} = 333 \text{ Sellers}$$

Figure 9. Example Calculation Using Sample Size Formula

Data Validity

There are two major threats to validity – internal threats and external threats. Internal validity threats are experimental procedures, treatments or subject experiences that threaten a researcher's ability to correctly draw inferences from study population. External validity threats are the result of the researcher incorrectly drawing inferences

from the data to other settings (conditions) or apply it to past or future events. Each of these threats to validity were addressed in the methodology for the research study. Correlational studies are higher than true or quasi-experiments on external validity but lower on internal validity (Creswell, 2002).

External Validity

External validity refers to the extent to which the results of a study can be generalized to a larger population. While true experiments have higher internal validity as they are internally consistent what is sacrificed is the ability to generalize to the real world. The non-experimental correlational research design achieves external validity through the generalization to the studied population which in this case was the large (in the millions) eBay auction site membership. As the auctions collected were selected on product category, the auction sellers and buyers could not be selected a priori.

Internal Validity

Internal validity of a study establishes that the data or findings are true or measures what is purported to be measured (Borg & Gall, 2006). Measurement error must be minimized and the instruments for data collection must be trusted to ensure internal validity.

Measurement error is the discrepancy between the observed value of a measurement and the true value due to the error contained in the measuring instrument. Any measurement error would be analyzed using statistical calculations. As a web crawler was the instrument for data collection, the data collection procedure could be repeated and results compared to prove replication and reliability.

Internal validity can also refer to the extent which variation in the dependent measure can be attributed exclusively to the independent variable. This is especially true in the case of the experimental research designs where the independent variable is directly manipulated in the treatment group, but not changed in the control group. In the research study, the initial plan's focus was on locating sellers with a large number of sales in order to have the maximum number of buyer feedback comments to evaluate. Sampling would be random based on "n" sellers with a feedback score greater than or equal to the filter threshold number where "n" will be the suggested sample size for the given population. The initial plan's sampling technique would duplicate that previously used by Finch (2006). Assignment to group – based on seller's behavior (honest or fraudulent) – would take place in a post-selection process when the seller was categorized by the evaluators. Thus the selection of sellers would be blind as to group.

Reliability

Inter-rater agreement, inter-rater reliability, or concordance is the degree of agreement among raters. Inter-rater agreement is used to measure reliability. Inter-rater agreement is estimated based on the correlation of scores in the ratings of two or more observers (raters) assigned to reviewing each behavior or observation. Two independent groups of raters were used – evaluators and coders. Details for each type of rater are specified in the appropriate sections describing the purpose, function, training, etc. In the initial plan Cohen's Kappa was to be used to assess inter-coder agreement among the raters. The Kappa provides an estimate of reliability or an index of agreement between two raters' observations or scores. Cohen's Kappa ranges between 0 and 1 and represents the

proportion of agreement corrected for chance (Morgan, Leech, Gloeckner, & Barrett, 2007).

The alternative method which used crowdsourcing prevented the measuring of inter-rater reliability with Cohen's Kappa. This was because raters were randomly and anonymously assigned to each work unit known as a Human Intelligence Test (HIT). Instead the techniques of multiple worker assignments per HIT (plurality), minimum work time per HIT, gold standard data, and advice of auditing were used to ensure the reliability of raters. These techniques are detailed in a later section entitled *Building the Prototype HITs for Amazon Mechanical Turk.*

Sources of the Variables

The objective of the research was to determine if the presence of negative-positive type feedback comments by buyers is a indicator (predictor) that a seller is behaving fraudulently (observed behavior). The correlational research design provides for discovering relationships between variables, measuring the degree and direction of relationships, and from the discovered relationships predictions can be made.

In terms of the correlational research design mechanics, the focus of the research study was determining if there is a relationship [hereafter called the primary relationship] between "negative-positive type feedback comments by buyers" and "a seller behaving fraudulently." If the primary relationship existed, then the next step was measuring the degree and direction of the primary relationship (if possible). The remaining step was to determine if predictions could be made based on the primary relationship.

In the research study, both of the variables in the primary relationship are abstracted from the collected public eBay data. Presence or absence of a negative-positive type

feedback was derived by coders reviewing the *Buyer Feedback Type* field (which must be positive) and *Buyer Feedback Comment* field's text format. The output from each of the three coders was placed into separate fields - *Negative-Positive 1, Negative-Positive 2,* and *Negative-Positive 3*. The *Negative-Positive* fields are categorical containing one of the following values – Y (Yes) or N (No). Based on majority rule, a final inter-coder agreed value was assigned to the *Negative-Positive Consensus* field.

The seller's behavior – honest or fraudulent – was derived by the evaluators' judgments of the seller's behavior based on all the available eBay public data and using a predefined criterion for what is fraudulent behavior. Each of the evaluators was asked to answer the following key question for each seller – Is the seller exhibiting fraudulent type behavior? The answer was either "No" the seller is not acting fraudulently (i.e. honest behavior) or "Yes" the seller is acting fraudulently (i.e. fraudulent behavior).

After a seller has been reviewed by all evaluators, the answer from each of the three evaluators was placed into separate fields - *Fraudulent-Type Behavior 1, Fraudulent-Type Behavior 2,* and *Fraudulent-Type Behavior 3*. The *Fraudulent-Type Behavior* fields are categorical containing one of the following values – N (No) or Y (Yes). Based on majority rule (2 out of 3) of the *Fraudulent-Type Behavior* fields' ratings, a final inter-evaluator agreed value of N or Y was assigned to the *Fraudulent-Type Behavior Consensus* field.

Independent Variable

The independent variable (predictor) is typically the variable being manipulated or changed and the dependent variable is the observed result of the independent variable being manipulated. For this research study, the independent variable was indicated by the

presence or absence of negative-positive type feedback in the *Buyer Feedback Comment* field. The *Negative-Positive Consensus* field was the independent variable.

Dependent Variable

The dependent variable is the event studied and expected to change whenever the independent variable is altered. The observed phenomenon was the type of seller behavior – honest or fraudulent - experienced by the buyer. The *Fraudulent-Type Behavior Consensus* field was the dependent variable.

Data Record Layout

In order for the data collection agent to perform its function of parsing and extracting data from the eBay web pages, exactly what data needed to be collected had to be clearly defined. One of the steps in the methodology required evaluators to make a judgment of classifying each seller's behavior as honest or fraudulent. Naturally, the evaluators wanted to review all the available data about a seller before forming an opinion. Even if not used as part of the data analysis, one advantage of collecting the additional data was that it might prove valuable in future research studies. Alternatively, unexpected events or relationships could be uncovered when using the additional data.

See Appendix C for details on the data record layout. Detailed for each data field are name, description, type, size, format, and comments. The eBay webpage source for each data field can be found detailed in the crawler design (see Appendix B). For ease of performing statistical analysis, only a single flat data file was created and seller data fields were duplicated in every record (i.e. a sales transaction with buyer feedback).

Data Obfuscation

Although the data collected was in the public domain, maintaining anonymity was still a requirement. The first potential issue was preventing the coders from being effected by any personal knowledge that they might have of an eBay seller's or buyer's identity via their eBay userid. For example – Do not want a coder saying, "Hey that's my sister's eBay userid!" The second potential issue was to prevent any bias by the coders based on any other extraneous data. For example - an eBay userid that is political (HEILHITLER) or derogatory (SLUTTYGIRL). Every eBay auction has a unique *Item Number* to identify the item being offered for sale. Each collected record detailed a single purchase by a buyer from a qualified seller with a corresponding *Item Number* uniquely identifying the auction. In order to prevent the possibility of a coder looking up information about an item using the *Item Number*, it was masked with a system generated autonumber field named *Feedback Number*. As coders only had access to the content of two fields [*Feedback Number, Buyer Feedback Comment*], this isolated the coders and ensured that no extraneous data effected how they performed their task.

The situation was reversed with evaluators as no data obfuscation needed to be taken. Evaluators needed to make a judgment in classifying each seller's behavior as honest or fraudulent. The evaluators wanted to review all the available data about a seller before forming an opinion. An evaluator was required to indicate any personal knowledge of a seller or buyer in the *Other Comments* section of the *Evaluator Worksheet* (see Appendix D). As no evaluator indicated any personal knowledge of a seller or buyer, it was not necessary for the researcher to review and determine what corrective action needed to be taken in the case of personal knowledge by an evaluator.

Identifying Fraudulent Sellers

There are only two sources with authority that can equitably state an eBay member is a fraudster – eBay and criminal court rulings. Due to confidentiality, eBay will not provide any details to third-parties on complaints against a member or indicate why a member's account was suspended or disabled. Therefore, an explicit confirmation that a specific online auction member was an opportunistic seller from the primary source – eBay - was not available. Observing the public actions of eBay –like suspending a member's account – did provide a secondary source from which some inferences could be drawn.

The probability that a person who commits a fraudulent act will be caught and prosecuted is very low. The execution of a fraudulent act often leaves the victim unaware it has taken place or too embarrassed to report it. The covert nature of fraud makes collecting sufficient evidence for prosecution and conviction time consuming and difficult. Nonviolent crime like online auction fraud has a lower priority with law enforcement agencies than violent crime against people or damage to property. Even when a fraudster is caught and prosecuted, the person often receives a light sentence or no sentence in return for restitution to the victims (C. Chua & Wareham, 2004). The result is the criminal court record containing formal prosecutions for online auction fraud are very limited in number. In addition, the court records could be sealed preventing public access to the details or not current enough to extract data from eBay as it can take years for a final legal verdict to be reached.

Studying online deception is problematic as with other deviant behavior the successful perpetrators work hard to avoid detection (Kauffman & Wood, 2000; Nikitkov & Stone,

2006). An opportunistic seller will employ deception tactics in order to mask his/her behavior and illicit activities. What can be done is quantifying the *perception* by others that a specific online auction member exhibits the behaviors and actions characteristic of a fraudulent seller. Based on the quantified perceptions, an inference can be drawn that a specific online auction member is behaving fraudulently. The technique of using inferences from secondary sources to indicate an individual's probability of being an opportunistic seller was done in prior research by Chua and Wareham (2008), Chua, Wareham and Robey (2007), and Pandit, Chau, Wang and Faloutsos (2007).

All secondary sources can only make inferences or statements without being definitive that an eBay member is behaving fraudulently. The relative measure of weight for an inference or statement varies based on the secondary source. For example - A single complaint message posted about a seller by one buyer on the eBay discussion board would have a lower weight than an investigative news reporter's article on an eBay member's potentially fraudulent acts. A single buyer's posting must be considered an opinion. Whereas an investigative reporter would be held to a higher standard with the expectation of being objective, confirming any facts presented, and responsibility as the reporter (or the publisher) could be taken to court for liable. However, the relative measure of weights can be variable for any given secondary source. Imagine the situation where multiple buyers instead of a single buyer posted complaint messages about a seller on the eBay discussion board. With a number of buyers making a complaint against a single seller, it raises the probability that the seller is engaged in fraudulent behavior (Surowiecki, 2004). The relative measure of weight for each secondary source was not a primary factor in this research. The constraint that must be remembered is

secondary sources are not definitive and any findings must be held with that limitation in mind. An example of mistakenly treating secondary sources as authoritative and definitive can be found in the study of Pandit et al. (2007). In their study a statement was made - “Through manual investigation (Website browsing, newspaper reports, etc) we located 10 users who were guaranteed fraudsters” (Pandit, et al., 2007, p. 207). Using secondary sources, a judgment based on the available evidence can be made with a degree of confidence that a specific online auction member as a seller is behaving in an honest or fraudulent manner. No secondary source can be used to definably state or label an eBay member as a guaranteed fraudster.

Coding – Identifying Seller Behavior as Honest or Fraudulent

Studying online deception is problematic when using conventional methods as with other deviant behaviors the successful perpetrators work hard to avoid detection. By developing explicit rules to distinguish between honest and fraudulent seller behavior, it was possible to appropriately and constantly categorize a seller’s behavior as honest or fraudulent.

In the initial plan, a minimum of three evaluators (who were unaware of the study’s purpose) would be recruited and would each review all the sellers. An evaluator would be required to make a judgment classifying each seller’s behavior as honest or fraudulent. Which raises the question – On what criteria will the evaluators base their judgment?

As human behavior is complex and sometimes inconsistent, attempting to find a single specific behavior pattern to signal fraudulent behavior is not realistic. Taking a clue from prior research into credit card fraud, online auction fraud detection is based on looking for red flags and behavior patterns (Bhargava, Zhong, & Lu, 2003). The mechanical

process of going through a long checklist of all the potential red flags and behavior patterns for even a single seller would be time consuming and any lapse by an evaluator could result in a misclassification. As the number of sellers that would need to be reviewed appeared to be in the hundreds, it would not be feasible to perform the task entirely manually. Nor is there an automated means for making the required judgment.

Fortunately, there was a publicly available software application that automatically searched for red flags and suspect behavior patterns in eBay auctions. The *Auction Inquisitor* software checks an eBay auction for over 200 common and not-so-common signs of fraud plus checks the seller's history, and finishes by presenting a report of the results with comments (Ford, 2010). Using *Auction Inquisitor* as a front end for the evaluation process provided the following advantages – greatly reduced the time required to review the red flags and suspect behavior patterns for a seller; enabled the review process to be performed consistently and without human error; and presented the results in a summarized and standardized format. It must be made clear that the *Auction Inquisitor* software did not make a judgment as to whether or not a seller's behavior was fraudulent. It only presented its findings in the form of a standardized summary report.

In the initial plan, each evaluator was to watch a training video on how to use the *Auction Inquisitor* software application. A copy of the *Evaluator Worksheet* would be provided to each evaluator (see Appendix D) and reviewed with the researcher. The *Evaluator Worksheet* summarized the rules for what behaviors are deemed as fraudulent for the research study (see prior section on *Definition of Fraud*). The procedure for performing the seller evaluation is detailed in Appendix E. Ten preselected sellers would be used for training to ensure that the evaluators experienced the full range of seller

behaviors and understand the criterion for fraudulent behavior. The evaluators would be physically separated in order to ensure that they worked independently. Each of the evaluators would be asked to answer the following question for each seller – Is the seller exhibiting fraudulent type behavior? The answer would be either "No" the seller is not acting fraudulently (i.e. honest behavior) or "Yes" the seller is acting fraudulently (i.e. fraudulent behavior). Upon successful completion of the training, the evaluators would start work on the actual experimental data. Presentation of the sellers to each evaluator would be random. After a seller was reviewed by all evaluators, the answer from each evaluator would be placed into separate fields - *Fraudulent-Type Behavior 1*, *Fraudulent-Type Behavior 2*, and *Fraudulent-Type Behavior 3*. The *Fraudulent-Type Behavior* fields are categorical containing one of the following values – N (No) or Y (Yes). Based on majority rule (2 out of 3) of the *Fraudulent-Type Behavior* fields' ratings, a final inter-evaluator agreed value of N or Y would be assigned to the *Fraudulent-Type Behavior Consensus* field.

In the initial plan, validity and reliability would be addressed by the following methods. The author of the research proposal would evaluate a random sample set of sellers independently and compare the results with those of the evaluators. This reliability method has been deemed as the most accurate by Kolbe and Burnett (1991) and has been used for textual analysis in prior research studies. Next Cohen's Kappa would be used to assess inter-evaluator reliability among the evaluators who were assessing fraudulent behavior among sellers. In each case one person who was observing the situation (assessing fraudulent behavior among sellers) was an indicator. The Kappa would provide an estimate of reliability or an index of agreement between two raters'

observations or scores. Cohen's Kappa ranges between 0 and 1 and represents the proportion of agreement corrected for chance (Morgan, et al., 2007). One Kappa would compare the fraudulent-type behavior between evaluator 1 and evaluator 2; one Kappa would compare evaluator 1 with evaluator 3; and one Kappa would compare evaluator 2 with evaluator 3. For inter-evaluator agreement, the majority ratings would be used (two out of three) to code *Fraudulent-Type Behavior Consensus* as N (No) or Y (Yes).

Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not

In a forensic case study of an opportunistic seller, it was found buyers sometimes embed negative comments in positive feedback as a means of avoiding retaliation from sellers and damage to their reputation. This category of positive feedback is described as "negative-positive" feedback (Nikitkov & Stone, 2006). An example of negative-positive feedback is "Good product, but slow shipping". The objective of the research was to determine if the presence of negative-positive type feedback comments by buyers is an indicator that a seller is acting fraudulently. In order to meet this objective, the *Buyer Feedback Comment* for every buyer needed to be reviewed and coded in order to identify all the negative-positive feedbacks. As negative-positive feedback requires that the *Buyer Feedback Type* be positive, any *Buyer Feedback Comment* that has *Buyer Feedback Type* other than positive was filtered out as it did not need to be evaluated by the coders.

In the initial plan, a minimum of three coders (who were unaware of the study's purpose) would be recruited and would each review all the buyer feedback comments. A coder would be required to make a judgment to classify a *Buyer Feedback Comment* as being in negative-positive format or not by assigning a value to the *Negative-Positive* field as Y (Yes) or N (No).

The criteria required for the coder's judgment would be minimal. eBay only provides for three types of feedback - negative, neutral, and positive. A subset of positive feedback would be flagged by the coders as negative-positive if it met one of the formats - "I was pleased with X, but unhappy about Y for the transaction" [+X, -Y] or "I was unhappy about Y, but was pleased with X for the transaction" [-X, +Y]. A diagram of the coder procedure can be found in Figure 10. As eBay *Feedback Type* is restricted to the value of negative, neutral or positive; invalid feedback types were not present. Seller auction sales without a feedback type do not appear in an eBay seller's transaction history and therefore were not collected or require review.

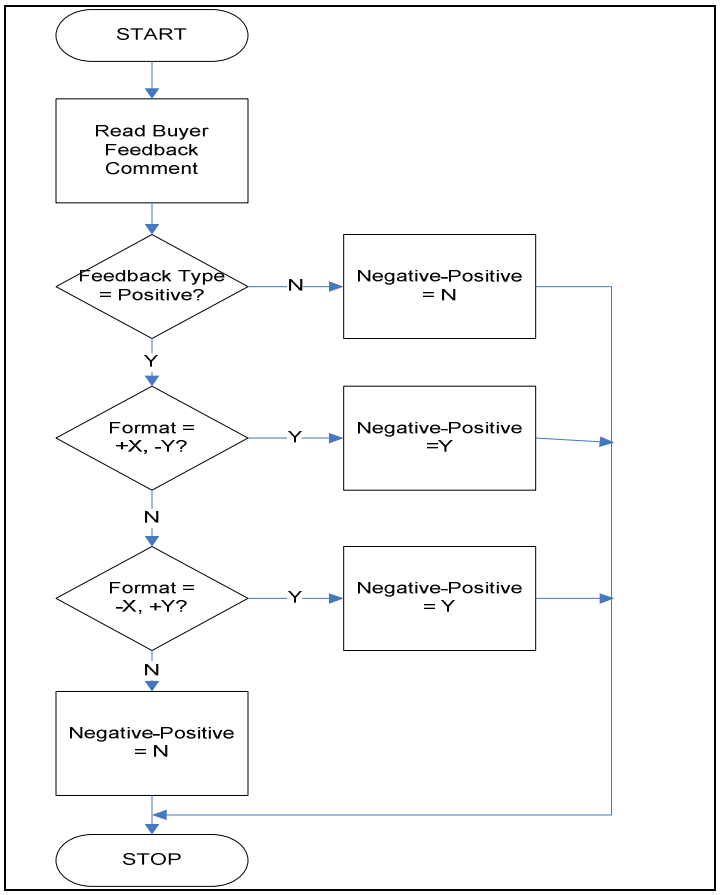


Figure 10. Flowchart of Coder Procedure

In the initial plan, each coder would receive and review with the researcher a copy of the *Coding: Indentifying Buyer Feedback Comment as Negative-Positive* document (see Appendix F). This document summarized the rules for classifying buyer feedback comments as negative-positive or not. A preselected sample of 100 buyer feedback comments would be used for training to ensure that coders experienced the full range of seller feedback comments and understood the criterion for classification as negative-positive type feedback or not. The coders would be physically separated in order to ensure that they worked independently. Each of the coders would be asked to answer the following question for each buyer feedback comment – Does the buyer feedback comment meet the criterion for negative-positive type feedback? The answer would be either "No" does not qualify as negative-positive type feedback or "Yes" does qualify as negative-positive type feedback. Upon successful completion of the training, the coders would start work on the actual experimental data. Presentation of the buyer feedback comments to each coder would be random. After a buyer feedback comment was reviewed by all coders, the answer from each coder would be placed into separate fields - *Negative-Positive 1, Negative-Positive 2,* and *Negative-Positive 3*. The *Negative-Positive* fields are categorical containing one of the following values – N (No) or Y (Yes). Based on majority rule (2 out of 3) of the *Negative-Positive* fields' ratings, a final inter-coder agreed value of N or Y would be assigned to the *Negative-Positive Consensus* field.

In the initial plan, validity and reliability would be addresses by the following methods. The author of the research proposal would evaluate a random sample set of buyer feedback comments independently and compare the results with those of the coders. This reliability method has been deemed as the most accurate by Kolbe and

Burnett (1991) and has been used for textual analysis in prior research studies. Next Cohen's Kappa would be used to assess inter-coder reliability among the coders who were reviewing the buyer feedback comments for negative-positive type comments. In each case one person who was observing the situation (coding the negative-positive feedback comments among buyers) was an indicator. The Kappa would provide an estimate of reliability or an index of agreement between two raters' observations or scores. Cohen's Kappa ranges between 0 and 1 and represents the proportion of agreement corrected for chance (Morgan, et al., 2007). One Kappa would compare the negative-positive feedback between coder 1 and coder 2; one Kappa would compare coder 1 with coder 3; and one Kappa would compare coder 2 with coder 3. For coder agreement, the majority ratings would be used (two out of three) to code *Negative-Positive Consensus* as Y (Yes) or N (No).

Population Size

In the initial plan, the population size needed to be estimated to determine the feasibility of the traditional methodology of using dedicated raters. Using a prototype of the proposed web crawler program, a full data extract from eBay was performed for the previously identified target - *Computers and Networking: Laptop* category. A full data extract included all sellers as it did not filter out sellers based on their feedback score. The full data extract procedure was repeated once a week for three weeks with the results summarized in figure 11. Where *Total Auction Items* was the number of individual items listed in the category for sale. Where *Total Unique Sellers* was the number of unique sellers (based on eBay userid) in the category. Elimination of duplicate sellers was a necessary step as a single seller can list several items for sale. Where *Total Feedback*

Comments was the composite of all feedback comments found in each unique seller's eBay member profile.

Week	Total Auction Items	Total Unique Sellers	Total Feedback Comments
1	15,823	438	361,040
2	15,282	406	355,469
3	16,431	446	365,056

Figure 11. Data Extracts for Category – Computers & Networking: PC Laptops & Notebooks

The findings of the three full data extractions showed a relatively small population of unique buyers ranging from 406 to 446. A small number of unique buyers could adversely effect the research's data analysis as the number of fraudulent sellers within the eBay member population is reported to be very small. Exactly how small the fraudulent seller population is could be seen by the 0.01 percent officially reported by eBay (B. Cox, 2003; Konrad, 2005). Based on this rate and a unique seller population of 446, the number of fraudulent sellers would be estimated at 0.0446 which effectively was zero. The number only rose to 0.20 percent based on a research study of eBay fraud by Gregg and Scott (2008). Using this calculation and a unique seller population of 446, the number of fraudulent sellers would be estimated at 0.892 which rounded up to one. Per prior cited research studies, the distribution of fraudulent sellers appears to be skewed and focused on specific categories like the *Computers and Networking: PC Laptops & Notebooks* category. Even with the skewing effect should the number of eBay sellers designated by the evaluators as exhibiting fraudulent type behavior had proven insufficient, two options were available:

1. Select another skewed category with a larger unique seller population.
2. Combine multiple skewed categories to create a larger unique seller population.

Whether or not either of these options would need to be implemented could only be determined after the evaluators reviewed the unique sellers and determined the number of sellers exhibiting fraudulent type behavior in the *Computers and Networking: PC Laptops and Notebooks* category. Therefore the most prudent course of action was for evaluators to complete their review of the unique sellers before the coders began work on the buyer feedback comments. Two other conclusions were draw from the small number of unique sellers that were found:

1. As the unique seller population needed to be maximized filtering the seller population size based on feedback score was not required.
2. Sampling method and size for sellers was a moot point as finding a small number of unique sellers required inclusion of the entire population.

Using the data extraction from week 3, an analysis was performed on the buyer feedback comments population (see Figure 12).

		Number	Percent
Total Feedback Comments		365,056	100.00%
	Minus Negative	3,273	(0.90%)
	Minus Neutral	3,030	(0.83%)
	Minus Blank	2,167	(0.59%)
Remaining Positive		356,586	97.68%
	Minus As Buyer	13,224	(3.62%)
	Minus Non-English	7,566	(2.00%)
Qualified Buyer Feedback Comments		**335,796**	**92.00%**

Figure 12. Analysis of Data Extract for Week 3

The *Total Feedback Comments* found was 365,056. All unqualified records were deleted from the *Total Feedback Comments* population:

Minus Negative – Any feedback comment with a feedback type of *Negative* was removed as negative-positive feedback requires a feedback type of positive.

Minus Neutral – Any feedback comment with a feedback type of *Neutral* was removed as negative-positive feedback requires a feedback type of positive.

Minus Blank – Any feedback comment with a feedback type of *blank* was removed as negative-positive feedback requires a feedback type of positive. eBay will set a feedback type to blank for partially deleted or censured comments.

The *Remaining Positive* number of 356,586 contained only feedback comments that had a feedback type of *Positive*.

Most eBay members switch between the roles of seller and buyer. Each qualified seller's eBay member profile can contain feedback for both roles. Therefore all feedback comments in which the seller was acting as a buyer needed to be eliminated as designated by *Minus As Buyer*.

A data set member which is different in some way from the general pattern is called an outlier. An unexpected set of outliers were found during the analysis of the data extract. Although eBay has websites hosted in over 30 countries, the ebay.com website located in the United States is the largest and is used by eBay members living in other countries. As a result, some of the buyer feedback comments from the international eBay members were not in English. Non-English buyer feedback comments were found written in French, German, Italian, Spanish, and other languages. Inclusion of non-English buyer feedback comments would result in ambiguity due to translation plus the additional expense of hiring translators. *The assumption was made that buyer feedback comments are consistent regardless of the language in which they are composed.* That is to say a buyer's compliment or complaint about a seller in the form of a feedback comment was independent of the spoken/written language used by the buyer. Therefore non-English

buyer feedback comments which constitute less than 2% of the total population were treated as outliers and excluded from the data to be analyzed. This exclusion was indicated by *Minus Non-English*.

The analysis of the pilot data extract for week 3 provided quantitative measurements for the magnitude of the proposed analysis work. As designated by *Total Unique Sellers* - the total number of sellers that would need to be reviewed is 446. As designated by *Qualified Buyer Feedback Comments* - the total number of buyer feedback comments that would need to be reviewed is 335,796. When the actual production data extraction was eventually performed for the dissertation report the resulting numbers did vary, but the magnitude remained the same. This consistent order of magnitude made it possible to estimate in advance the time and labor required (workload) to complete the analysis of sellers and buyer feedback comments.

Analysis of the Seller Workload Using Traditional Dedicated Raters

The research study required analysis of two components – buyer feedback comments and sellers. As previously stated the interpretation of the natural language contained in the buyer feedback comments must be done by a human as automated options do not provide the required accuracy. The analysis of the sellers was complex requiring a judgment to determine whether each seller is exhibiting fraudulent type behavior or not. As previously stated this judgment must be done by a human as an automated option does not exist.

Having established that both components would require human analysis, a framework for performing each analysis was specified. The seller analysis framework was described in the prior section entitled *Coding – Identifying Seller Behavior as Honest or*

Fraudulent. The mechanics to implement the framework are described in step-by-step detail for the evaluators per *Appendix E – Coding: Identifying Seller Behavior as Honest or Fraudulent*. The buyer feedback comment analysis framework was described in the prior section entitled *Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not*. The mechanics to implement the framework are described in step-by-step detail for the coders per *Appendix F – Coding: Indentifying Buyer Feedback Comment as Negative-Positive*.

Using the data extraction from week 3 and following the section entitled *Coding – Identifying Seller Behavior as Honest or Fraudulent*, a time-trial test was run using three individuals each assigned the role of evaluator. The researcher preselected a sample of 10 sellers to ensure that the evaluators experience the full range of seller behaviors. As the objective of the test was to determine the average time required to review a seller, inter-evaluator reliability was not measured. The average time to evaluate a single seller was 20 minutes. This was calculated based on elapsed time for each evaluator to complete the test divided by 10 sellers give the average time for the evaluator to review a single seller. The average time for each of the three evaluators was summed together and divided by three giving the overall average of 20 minutes. From this information, an estimated time to complete the analysis and the cost of the analysis was extrapolated using three dedicated raters as evaluators and a minimum wage rate of $8 per hour (Figure 13).

446 Sellers X 0.33 Hours/Seller = 148 Hours [18.5 workdays]

148 Hours X $8.00/Hour X 3 Evaluators = $3,552

Figure 13. Analysis of Sellers – Estimated Time and Cost

Analysis of the Feedback Workload Using Traditional Dedicated Raters

Using the data extraction from week 3 and following the section entitled *Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not,* a time-trial test was run using three individuals each assigned the role of coder. The researcher preselected a sample of 100 buyer feedback comments to ensure that the coders experienced the full range of buyer feedback comments. As the objective of the test was to determine the average time required to review a single buyer feedback comment, inter-coder reliability was not measured. The average time to evaluate a single buyer feedback comment was 15 seconds. This was calculated based on elapsed time for each coder to complete the test divided by 100 buyer feedback comments give the average time for the coder to review a single buyer feedback comment. The average time for each of the three coders was summed together and divided by three giving the overall average of 15 seconds. From this information, an estimated time to complete the analysis and the cost of the analysis was extrapolated using three dedicated raters as coders and a minimum wage rate of $8 per hour (Figure 14).

335,796 Buyer Feedback Comments X 15 Seconds/Buyer Feedback Comments X 1 Hour/3600 Seconds = 1400 Hours [175 workdays]

1400 Hours X $8.00/Hour X 3 Evaluators = $33,600

Figure 14. Analysis of Buyer Feedback Comments – Estimated Time and Cost

Analysis Summary of the Workload Using Traditional Dedicated Raters

Three factors needed to be considered for the successful implementation and completion of the research study – feasibility, time, and resources. A major cause of failure was found in the lack of financial resources as the total estimated cost was

$37,152 (Seller $3,552 + Feedback $33,600). The next issue was time with a minimum requirement of 175 workdays for the coders to complete their work which was not satisfactory. Lastly, measuring inter-rater reliability requires that all three raters for each analysis complete all the work. Hiring a new rater would mean scrubbing any work completed by the old rater and redoing all the work. The probability of one of the raters quitting the project before completing all the work was high. Although manageable – hiring a new evaluator would result in an additional 18.5 workday delay. Having to hire a new coder would result in an additional 175 workday delay which would not be viable. Thus the feasibility of using dedicated full-time raters was low.

One alternative to reduce the cost for analysis would be to minimize the number of buyer feedback comments that are reviewed. Random sampling would normally be the method used to achieve this goal. In research question 3 (RQ3), it was stated - For each seller will negative-positive type feedback comments from buyers fall into a pattern? As the size of the buyer feedback comment population in negative-positive format was an unknown at the time, inclusion of all negative-positive feedback comments was a prerequisite to analyzing the presence or absence of any pattern. The conclusion drawn was that the population would need to be analyzed in toto. In summary, the initial plan using the traditional method of dedicated raters was not viable and an alternative methodology for performing the two analyses was needed.

Introduction to Amazon Mechanical

One of the components of the Amazon Web Services suite is Amazon Mechanical Turk ("Amazon Web Services," 2010). Launched in 2005 as a commercial offering, Amazon Mechanical Turk (AMT) was initially used by Amazon for internal projects

("Amazon Mechanical Turk," 2010). Its purpose was to fulfill the demand for using human intelligence rather than a computer to perform a task. This type of task was called a Human Intelligence Task (HIT). A HIT is defined as a problem that humans find simple, but computers are unable to do or find extremely difficult to do. For example a HIT related to a photograph could be - “What animal is in this photograph?”

AMT is a commercial implementation of crowdsourcing. The concept of crowdsourcing was first described in a *Wired* magazine article as outsourcing tasks to a large group of people (Howe, 2006). Unlike user-generated content or social networks, participants in a crowdsourcing have no contact with one another. One AMT worker cannot see the results of another’s work. A problem is broken down into discrete tasks. Each task is self-contained. As the tasks are self-contained, it is possible for each task to be assigned to a different individual (or multiple individuals) and worked on simultaneously. The resulting architecture is a massively parallel human work force. The potential processing capacity of crowdsourcing architecture can be more fully appreciated based on an observation by von Ahn et al. (2004) where they calculated that a crowd of 5,000 people playing an appropriately designed computer game 24 hours a day could label all 425,000,000 images on the Google website in just 31days.

Within AMT users can function in two roles - requester and worker. Requesters post work to be done using units called Human Intelligence Tasks or HITs (See Figure 15).

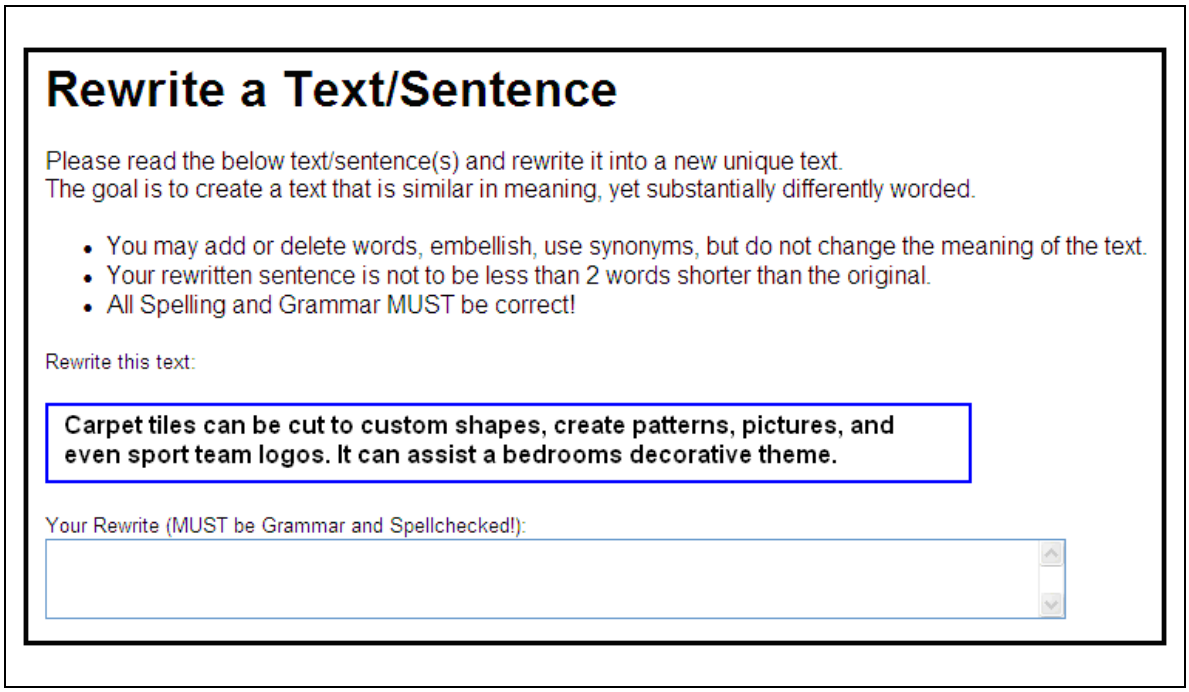


Figure 15. Sample Human Intelligence Task (HIT)

Each HIT has a value in the form of a micro-payment which can be as little as $0.01. Every HIT can be completed by one or multiple workers before it is removed from the list of available HITs. The requester sets the number of workers based on assignments set per HIT. An assignment is the maximum number of workers who can perform the task. A HIT can optionally have one or more qualifications. A qualification can be a system qualification provided by AMT like *Worker HIT Acceptance Rate*. Another type of qualification is the user-defined qualification. A user-defined qualification is a test built by a requester. For example requiring a worker to take a *Spanish Comprehension Test* and pass with a minimum grade before being allowed to work on HITs translating sentences from English to Spanish. A requester can specify a time limit within which workers must complete work on a HIT. The requester pays the workers for completed HITs, but has the ability to review and reject without payment any HIT deemed invalid.

The requester can block workers based on their AMT userid from working on specific HITs.

A person who signs up to perform work on AMT is described as a worker. AMT workers commonly refer to themselves as “Turkers” in online discussion forums and blogs (Snow, O'Connor, Jurafsky, & Ng, 2008). Workers are only paid upon completion of work on a HIT and approval of the requester. Tasks are randomly assigned to a worker within a HIT. Should a HIT have multiple assignments, a worker can only work on a given task within a HIT once. Before choosing to work on a HIT - a worker can see sample HITs, payment information, the time limit for working on a HIT, and any qualification requirements. Workers discover HITs based on a keyword search interface that provides HIT previews. It is the worker’s discretion to determine which HITs and the number of HITs that will be worked on. Payments for completed tasks can be redeemed by workers on Amazon via gift certificate or be later transferred to a worker's bank account.

A hypothetical example to illustrate the mechanics for AMT - Imagine you own a store that sells toys. Your store has a website on which customers can review your inventory of toys and make purchases. The website displays your entire store inventory of 2,000 toys. A picture and description for each toy to be displayed on the website are stored in a database. You recently received complaints from multiple customers that some of the toys’ pictures and descriptions do not match on the website.

The problem is “Does the toy’s picture correctly match its description?” In order to solve this problem you would manually need to compare every toy’s picture against its description. This is a time consuming task and prone to error due to its repetitive nature.

Alternatively you can use AMT. Acting as an AMT requester you need to create 2,000 HITs – one HIT for each toy found in the database. It is not necessary to manually create each of the 2,000 HITs. Using a HIT template (see Figure 16) and importing the contents of the database, the 2,000 HITs can be automatically created.

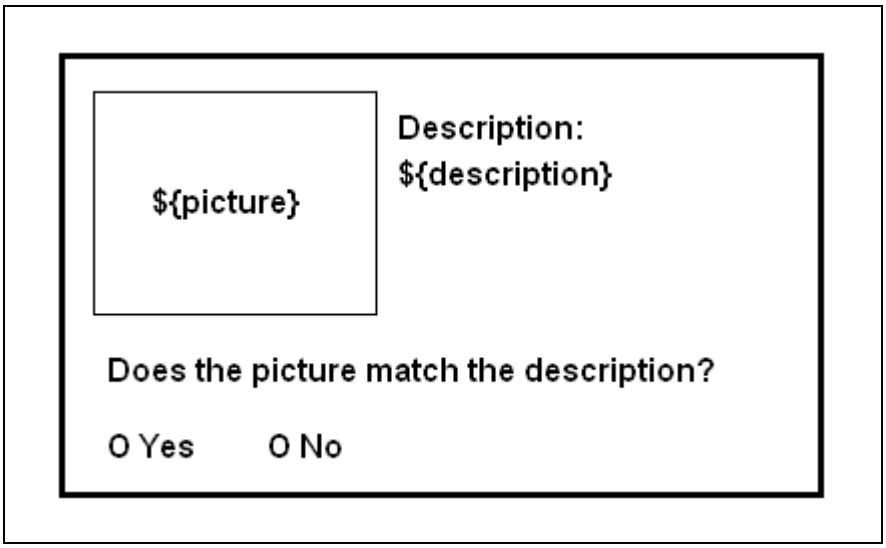


Figure 16. Sample HIT Template

First you create the HIT template. Next using the newly created HIT template and importing the contents of the website database, the 2,000 HITs are automatically created. As requester you need to "Publish" the HITs to make them available to workers. An example of the how a published HIT would look to a worker can be seen in Figure 17.

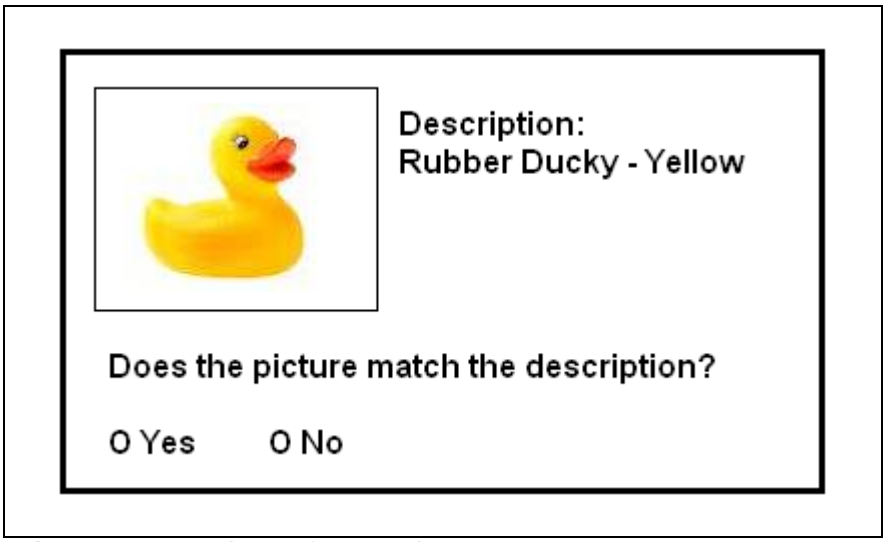


Figure 17. Sample HIT for a Worker

Almost immediately after being published the HITs will be discovered by workers in the Amazon Mechanical Turk's *List of Available HITs*. Multiple workers will simultaneously work on completing the HITs by clicking on the appropriate answer of YES or NO. As graphic image matching HITs are popular with workers, this number of HITs would typically be completed in less than an hour at a cost of $20 (2,000 HITS X $0.01/HIT).

AMT provides tools for a requester to monitor the HITs completion progress and review a worker's answer for each HIT. The requester pays the workers for completed HITs, but has the ability to review and reject without payment any HIT deemed invalid. The requester can block workers based on their AMT userid from working on specific HITs. The results are exported in the format of CSV data file. The results can then be analyzed to identify where a toy's picture and description did not match (See Figure 18).

HITId	Title	Reward	WorkerId	Picture	Description	Matched	
14DPDUP7AXX22	Toy Store Cleanup	$0.01	A1UJQS3FG7V1	http://mytoystore.com/image0001.jpg	Rubber ducky - Yellow	Y	
14DPDUP7AXX23	Toy Store Cleanup	$0.01	A1GLVY0SQ743	http://mytoystore.com/image0002.jpg	Ball - Red	Y	
14DPDUP7AXX24	Toy Store Cleanup	$0.01	A2E6QLLJKLWF	http://mytoystore.com/image0003.jpg	Pail and Shovel	N	◄ Mismatch
14DPDUP7AXX25	Toy Store Cleanup	$0.01	A1GLVY0SQ743	http://mytoystore.com/image0004.jpg	Jump Rope	Y	

Figure 18. Simplified and Annotated Example of HIT Results

Building the Prototype HITs for Amazon Mechanical Turk

In order to estimate the time and labor required (workload) for using AMT, a prototype Seller HIT and prototype Buyer Feedback HIT was constructed.

Within the AMT both requesters and workers are anonymous with everyone provided a unique system generated userid and identifiable information redacted. The two obvious concerns in using AMT arise when asking unseen, remote, and random strangers to perform a task. The first question was - How do you know that the workers will have the prerequisite skills or knowledge to perform correctly the task? The second question was - How do you know that the workers will actually make an effort to perform the task rather than just randomly click on responses?

The question of a worker having prerequisite skills or knowledge was addressed through the use of qualifications ("Amazon mechanical turk requester best practices guide," 2010). A HIT can optionally have one or more qualifications. A qualification can be a system qualification provided by AMT like *Worker HIT Acceptance Rate*. Another type of qualification is the user-defined qualification. A user-defined qualification is a test built by a requester. For example requiring a worker to take a *Spanish Comprehension Test* and pass with a minimum grade before being allowed to work on HITs translating sentences from English to Spanish.

Two qualifications came directly from the sections entitled *Coding – Identifying Seller Behavior as Honest or Fraudulent* and *Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not*. The two qualifications were common for both prototype HITs – the worker must be 18 years or older AND the worker must be a native English speaker. AMT has a mandated age requirement of 18 year or older for any worker. The age qualification must be satisfied before AMT will issue an AMT userid to the worker. A user-defined qualification named *Research Qualification Native English Speaker* was created to qualify a worker as native English speaker (See Figure 19).

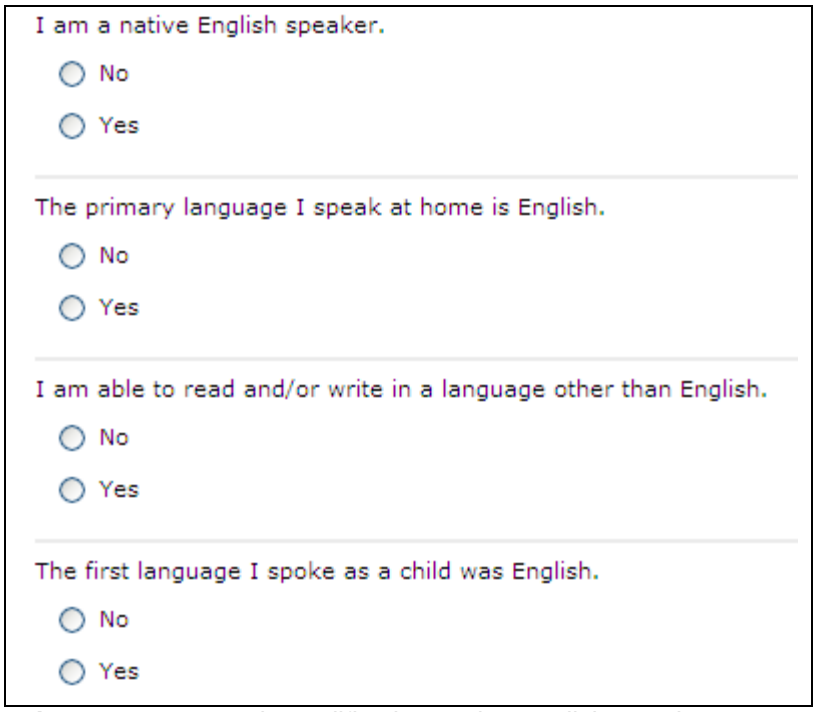


Figure 19. Research Qualification Native English Speaker

The definition of “native speaker” was taken in the content of "mother tongue" which is the first language a person heard/spoke as a child ("Merriam-Webster's collegiate dictionary," 2005). For the *Research Qualification Native English Speaker* qualification – the required answers to qualify/pass as a “Native English Speaker” were YES for “I am a

native English speaker" and YES for the "First language I spoke as a child was English." The other two questions were conspicuous distracters.

A third common qualification was based on a recommendation from the *Amazon Mechanical Turk Best Practices Guide* ("Amazon mechanical turk requester best practices guide," 2010). Per the guide, "To get the best selection of workers, we suggest using workers that have an approval rating of 95% or higher" which was designated by the system qualification named *Worker HIT Acceptance Rate*. This qualification was automatically managed by AMT and only needed to be included in the list of qualifications required for each of the prototype HITs.

For the prototype Seller HIT a user-defined qualification named *Research Qualification Seller Test* was created to test the worker's skills at performing the task of evaluating sellers for exhibiting fraudulent type behavior. The user-defined qualification was composed of a tutorial and a single seller which needed to be evaluated by the worker. Due to the extended time required by a worker to review a seller only a single seller was used in the qualification test. The qualification test was composed of 26 questions which were asked to assist and guide the worker in gathering the necessary data to base their final judgment. The 27th question was the final judgment question – "Did the seller exhibit fraudulent type behavior to buyers?" and asked for a NO or YES answer. As the qualification test seller clearly was exhibiting fraudulent type behavior, the answer required to qualify/pass the *Research Qualification Seller Test* was answering YES to the objective judgment question of "Did the seller exhibit fraudulent type behavior to buyers?" The *Research Qualification Seller Test* was set to manual which required the researcher to individually review and authorize each worker as qualified.

The manual option allowed the researcher to verify that the applicant worker actually completed the 27 questions for the qualification test and was not gaming by just randomly answering YES on the 27th question. The *Research Qualification Seller Test* can be found in *Appendix G - Research Qualification Seller Test.*

Passing of the qualifiers – *Research Qualification Native English Speaker* and *Research Qualification Seller Test* – permitted a worker to gain access to the prototype Seller HIT named *Research Prototype Seller*. For a test population - the researcher reused the same preselected sample of 10 sellers from the previously run evaluator time-trial test. Each of the *Research Prototype Seller* HITs was based on the same format as the *Research Qualification Seller Test*. It was composed of a single seller which needed to be evaluated by the worker. Twenty-six questions were asked to assist and guide the worker in gathering the necessary data to base their final judgment. The 27th question was the final judgment question – "Did the seller exhibit fraudulent type behavior to buyers?" and asked for a NO or YES answer. The last entry in the HIT form was an optional comment field to provide a means for feedback from workers. The *Research Prototype Seller* HIT was can be found in *Appendix H - Research Prototype Seller HIT*.

For the prototype Buyer Feedback Comment HIT a user-defined qualification named *Research Qualification Feedback Test* was created to test the worker's skills at performing the task of evaluating buyer feedback comments. The researcher preselected an additional sample of 50 buyer feedback comments to ensure that the AMT workers would experienced the full range of buyer feedback comments. The user-defined qualification was composed of a tutorial and 50 questions. Each question contained one buyer feedback comment which needed to be evaluated by the worker. For each question

a single buyer feedback comment was displayed, the question was asked “Is the following statement in negative-positive format?” and asked for a NO or YES answer. The worker’s responses were compared to the correct answers for each question. A grade of 90% or higher was required to qualify/pass the *Research Qualification Feedback Test*. The *Research Qualification Feedback Test* can be found in *Appendix I - Research Qualification Feedback Test.*

Passing of the qualifiers – *Research Qualification Native English Speaker* and *Research Qualification Feedback Test* – permitted a worker to gain access to the prototype Buyer Feedback HIT named *Research Prototype Feedback*. For a test population - the researcher reused the same preselected sample of 100 buyer feedback comments from the previously run coder time-trial test. Each of the *Research Prototype Feedback* HITs was similar in format to the *Research Qualification Feedback Test* but only contained instructions/tutorial and a single question. In order to reduce scrolling time, the instructions/tutorial were hidden by default, but could be toggled (display/hide) by clicking on the hyperlink. One buyer feedback comment was displayed, the question was asked “Is the following statement in negative-positive format?”, and the worker was asked for a NO or YES answer. The last entry in the HIT was an optional comment field which provided a means for feedback from workers. The *Research Prototype Feedback HIT* with the instructions hidden can be seen in *Appendix J - Research Prototype Feedback HIT with Instructions Hidden.* An example with the instructions displayed can be seen in *Appendix K - Research Prototype Feedback HIT with Instructions Displayed.*

The second issue was whether or not the AMT workers would do the HITs correctly. Even after qualifying/passing the pre-HIT qualifications, a worker could still give random

answers for a HIT. A perceived lack of accountability could motivate some AMT workers to complete as many tasks as possible by just arbitrarily clicking. A classic example of rational self interest where an individual attempts to maximize their [monetary] rewards while minimizing their effort and costs. This type of activity by AMT workers is known by the slang term of "gaming" (Downs, Holbrook, Sheng, & Cranor, 2010). In one of more recent developments, gaming has been taken to the next level by the use of autonomous software applications known as "bots" to simulate human activity (Dekel & Shamir, 2009).

AMT requires the requester to approve each HIT done by a worker. The requester pays the workers for completed HITs, but has the ability to review and reject without payment any HIT deemed invalid. The requester can block workers based on their AMT userid from working on specific HITs. As the requester is the ultimate authority on the disposition of any HIT, the question raised by the second issue was - What techniques can a requester employ to ensure or measure the quality of a HIT?

Multiple techniques were applied to ensure or measure the quality of the data provided by AMT workers. These selected techniques have been employed by prior researchers when they used AMT - multiple worker assignments per HIT (plurality), minimum work time per HIT, gold standard data, and advice of auditing.

Plurality (Multiple work assignments per HIT) is one of the three mechanisms built into AMT to help ensure quality. Snow et al. (2008) indicated for a large set of HITs, an aggregate of four to six workers matched the results of a single domain expert. The use of plurality has been tested and verified by Heilman and Smith (2010), Pinchak et al. (2009), and Heymann and Garcia-Molina (2008). When a simple majority of the workers

agree on the result, the result will be accepted as the “correct” answer. If no plurality emerges, this usually means that the HIT is ambiguous (Barr & Cabrera, 2006).

AMT automatically measures and records the elapsed time required for a work to complete a HIT. A requester has the ability to generate an ad hoc report while a HIT batch is being processed to list all HITs completed below a specified minimum work time. Extremely short HIT durations by a worker - especially if found for multiple HITs - is an indicator of suspect work (Kittur, Chi, & Suh, 2008).

Both the qualification HITs and the tutorial/instructions included in each data HIT clearly indicated that all workers would be audited. Signaling to potential workers that their answers would be critically analyzed for invalid or random responses has been proven to increase the quality and time spent on the HITs (Kittur, et al., 2008).

Gold standard data is a collection of preselected data that have a known set of answers. These answers are typically produced by one or more individuals who are trusted and a domain expert. Gold standard data was used to ensure the accuracy of the answers provided by the AMT workers. If answers provided by a worker significantly deviates from the gold standard, then there is a high degree of probability that the worker is poorly performing, not doing what was asked or is attempting to game the system. This technique has been used by Tang and Sanderson (2010), Sorokin and Forsyth (2008), and Callison-Burch and Dredze (2010). The mechanics for the technique was randomly inserting (also known as salting) gold standard data into HITs. A worker did not know if the data to be evaluated came from the new data or from the gold standard. Details on construct of the gold standard data sets can be found in the sections - *Creating Gold Standard Sellers* and *Creating Gold Standard Feedbacks*.

Analysis of the Seller Workload Using Amazon Mechanical Turk

A pilot run was done using the AMT *Research Prototype Seller* HIT. Parameters were set to match those of the previously completed time-trial run using traditional dedicated raters (evaluators). The HIT assignment was set to three to allow three workers to serve in the role of evaluator for each *Research Prototype Seller* HIT. For a test population - the researcher reused the same preselected sample of 10 sellers from the previously run evaluator time-trial test. AMT automatically calculated the average time for a worker to evaluate a single seller at 22 minutes.

Using AMT requires that all HITs be self-contained. The self-containment makes it possible for each HIT to be assigned to a different worker (or multiple workers) and processed simultaneously. The resulting architecture is a massively parallel human work force. The variability of the massive parallel architecture makes it difficult to calculate quantitatively the total time required to review all the feedbacks. Based on empirical evidence from prior research studies, the estimated total time required to process all the sellers would range from a few hours to a few days (Heilman & Smith, 2010; Su, Pavlov, Chow, & Baker, 2007). As AMT workers are paid piece-work per HIT, there was no cost for the time spent by workers.

The pilot run for the *Research Prototype Seller* HIT mimicked the time-trial test in having three evaluators (workers) reviewing each seller. The idea being that simple majority rule would be used to formulate the “final” answer for any question. Snow et al. (2008) indicated for a large set of HITs, an aggregate of four to six workers matched the results of a single domain expert. A majority of five workers was cited by Yan et al. (2010) as the best strategy in consistently achieving more than 95% accuracy. Heilman

and Smith (2010), Pinchark et al. (2009), and Heymann and Garcia-Molina (2008) also determined that five workers was the optimum number per HIT. Based on this evidence the number of workers assigned to a production Seller HIT was increased from three to five for the production runs.

Experiments by other researchers using AMT demonstrated that first response to five one-cent HITs is 50-60% faster than a single five-cent task (Yan, et al., 2010). A review of financial incentives showed that increasing the micro-payment of HITs resulted in an increase in the quantity of work done, but not the quality of the work (Mason & Watts, 2009). The conclusion - If the micro-payment is too high, financial resources are wasted and inefficient workers are attracted. Elasticity of HIT throughput appears to be more dependent on the number of available online workers rather than the size of the HIT's micro-payment. The best strategy for a requester to adopt is start the first HIT batch at a low micro-payment and only increase the micro-payment size in subsequent HIT batches in the event of low worker response.

The quality of the workers' data was a critical concern. Especially as the only data to be collected was the final judgment answer of YES or NO contained in the 27th question of "Did the seller exhibit fraudulent type behavior to buyers?" The answers for the other 26 questions were not collected or analyzed as their sole purpose was to assist and guide the worker in gathering the necessary data to base their final judgment. The population of production Seller HITs was salted with 10% Gold Standard Sellers. The 10% gold standard measure was within the suggested 5% to10% range ("Crowdflower - gold standard," 2010).

From the prototype test information, an estimated cost of the analysis was extrapolated using the proposed five evaluators (Figure 20).

446 Sellers X 1.10 Gold Standard Multiplier[1] X 1 HIT/Seller X $0.25/HIT X 5 Evaluators = $614

Note: 1 of every 10 Seller HITs will be a Gold Standard Seller.

Figure 20. Analysis of Sellers – Estimated Cost using AMT

Analysis of the Feedback Workload Using Amazon Mechanical Turk

A pilot run was done using the AMT *Research Prototype Feedback* HIT. Parameters were set to match those of the previously completed time-trial run using traditional dedicated raters (coders). The HIT assignment was set to three to allow three workers to serve in the role of coder for each *Research Prototype Feedback* HIT. For a test population - the researcher reused the same preselected sample of 100 feedbacks from the previously run coder time-trial test. AMT automatically calculated the average time for a worker to evaluate a single buyer feedback comment at 17 seconds.

Using AMT requires that all HITs be self-contained. The self-containment makes it possible for each HIT to be assigned to a different worker (or multiple workers) and processed simultaneously. The resulting architecture is a massively parallel human work force. The variability of the massive parallel architecture makes it difficult to calculate quantitatively the total time required to review all the feedbacks. Based on empirical evidence from prior research studies, the estimated total time required to process all the buyer feedback comments would range from a few hours to a few days (Heilman & Smith, 2010; Su, et al., 2007). As AMT workers are paid piece-work (per HIT), there was no cost for the time spent by workers.

The pilot run for the *Research Prototype Feedback* HIT mimicked the time-trial test in having three coders (workers) reviewing each feedback. The idea being that simple majority rule would be used to formulate the "final" answer for any question. Snow et al. (2008) indicated for a large set of HITs, an aggregate of four to six workers matched the results of a single domain expert. A majority of five workers was cited by Yan et al. (2010) as the best strategy in consistently achieving more than 95% accuracy. Heilman and Smith (2010), Pinchark et al. (2009), and Heymann and Garcia-Molina (2008) also determined that five workers was the optimum number per HIT. Based on this evidence the number of workers assigned to a production Seller HIT was increased from three to five for the production runs.

Experiments by other researchers using AMT demonstrated that first response to five one-cent HITs is 50-60% faster than a single five-cent task (Yan, et al., 2010). A review of financial incentives showed that increasing the micro-payment of HITs resulted in an increase in the quantity of work done, but not the quality of the work (Mason & Watts, 2009). The conclusion - If the micro-payment is too high, financial resources are wasted and inefficient workers are attracted. Elasticity of HIT throughput appears to be more dependent on the number of available online workers rather than the size of the HIT's micro-payment. The best strategy for a requester to adopt is start the first HIT batch at a low micro-payment and only increase the micro-payment size in subsequent HIT batches in the event of low worker response.

The number of questions (buyer feedback comments to be reviewed) in the production feedback HIT was raised from one as seen in the prototype Feedback HIT to ten. There were two compelling reasons to do this. The first was the need to cut costs as paying even

at the lowest possible rate of $0.01 when multiplied by hundreds of thousands of HITs results in a total cost of thousands of dollars. As the time and effort required answering a single question was minimal, pooling multiple questions together into a single HIT was a viable and common practice used by requestors (Feng, Besana, & Zajac, 2009; Finin, Murnane, Karandikar, Keller, & Martineau, 2010). Second, the quality of the workers' data was a critical concern. Multiple questions per HIT made it possible to salt each HIT with one or more Gold Standard Feedbacks (Finin, et al., 2010). Each production Feedback HIT was salted with one Gold Standard Feedback which resulted in a gold standard measure of 10%. The 10% gold standard measure was within the suggested 5% to 10% range ("Crowdflower - gold standard," 2010).

From the prototype test information, an estimated cost of the analysis was extrapolated using the proposed five coders (Figure 21).

335,796 Buyer Feedback Comments X 1.10 Gold Standard Multiplier[1]
X 0.10 HITs[2]/Buyer Feedback Comment X $0.01/HITs X 5 Coders = $1847

Note 1: 1 of the 10 feedback comments per HIT will be a gold standard question.
Note 2: 10 feedbacks/HIT is equal to 0.10 HIT/feedback.

Figure 21. Analysis of Buyer Feedback Comments – Estimated Cost Using AMT

Analysis Summary for the Workload Using Amazon Mechanical Turk

Three factors needed to be considered for the successful implementation and completion of the research study – feasibility, time, and resources. The required financial resources were viable as sufficient research funding was available to cover the total estimated cost of $2,461 (Seller $614 + Feedback $1,847). The next the issue was time – Based on empirical evidence from prior research studies, the estimated total time required to process all the seller and buyer feedback comments would range from a few hours to a

few days (Heilman & Smith, 2010; Su, et al., 2007). As AMT workers are paid piece-work per HIT, there was no cost for the time spent by workers. Since the maximum time required to process all the data was estimated at a few days, should it have proven necessary the process could have been repeated multiple times in the event of an unexpected glitch occurring or to process additional data that was collected. The only constraint would be securing additional funding. The feasibility of using AMT was proven based on the successful pilot runs of the prototype Seller HIT and the prototype Feedback HIT. As with prior researchers that have used AMT, the major concern was applying the appropriate techniques to ensure that quality data would be produced by the workers. For integrity, a new group of people served as raters in creating the gold standard data for the study. Selection and qualification of new raters followed the procedure previously defined in the sections - *Coding – Identifying Seller Behavior as Honest or Fraudulent* and *Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not*. In summary, the proposed alternative of using AMT to process the experimental data was a viable solution.

Creating Gold Standard Sellers

Gold standard data was used to ensure the accuracy of the answers provided by the AMT workers. If answers provided by a worker significantly deviated from the gold standard, then there was a high degree of probability that the worker was poorly performing, not doing what was asked or was attempting to game the system.

A quality control technique used by Tang and Sanderson (2010), Sorokin and Forsyth (2008), and Callison-Burch and Dredze (2010) was randomly inserting (also known as salting) gold standard data into HITs to identify poorly performing, malicious or gaming

workers. A worker did not know if the data to be evaluated came from the new data or from the gold standard. Workers that gave too many wrong answers to the gold standard were more likely to add noise to the overall results and needed to be filtered out. Noise is defined as the measure of deviation from the gold standard data (Hsueh, Melville, & Sindhwani, 2009).

Gold standard data is a collection of preselected data that have a known set of answers. These answers are typically produced by one or more individuals who are trusted and a domain expert. Snow et al. (2008) demonstrated using multiple non-experts averaged out the noise resulting in the same quality answer as an expert. This technique was then applied by Snow et al. (2008) to produce gold standard data used in training sets as no gold standard data existed. Similarly research by Callison-Burch (2009) on machine translation quality and by Nowak and Ruger (2010) on tagging of images supported the findings that when combined non-expert judgments were equal to or better than human experts. *As no gold standard data set existed for determining whether or not an eBay seller is exhibiting fraudulent type behavior, the technique of using multiple non-experts was used to create a Gold Standard Sellers.*

A gold standard with noise would only support cautious benchmarking as it requires performance of the workers be better than the baseline by more than that which can be attributed to the noise. As noise is defined as the measure of deviation from the gold standard data (Hsueh, et al., 2009), noise level is reduced as the inter-rater agreement for an answer is increased. Noise is totally eliminated when all the raters are in agreement for an answer. In order to produce gold standard data with no noise, only answers with a strict metric were included. Strict metric is defined as the raters having consensus for an

answer (Ku, Lo, & Chen, 2007). The use of strict metric (consensus) negated the need to measure inter-rater reliability using Cohen's Kappa.

When the production data extract was completed, a seller was randomly selected from the extracted population. The randomly selected seller was reviewed by five qualified and dedicated evaluators. The number of evaluators selected was based on the recommendations of Snow et al. (2008), Callison-Burch (2009), and Klebanov and Beigman (2009). The same five evaluators were used to review all the sellers. The evaluation process followed the procedure as specified in the section entitled *Coding – Identifying Seller Behavior as Honest or Fraudulent*. A seller was only added to the Gold Standard Sellers if all the evaluators had a consensus in their answer. Any seller that did not have evaluator consensus was discarded. The suggested quantity of gold standard data is from 5% to 10% of the total population ("Crowdflower - gold standard," 2010). Based on the unique seller population size of 502 (See Chapter 4 for details), the size of the Gold Standard Seller data set could range from 25 to 50. Sellers continued to be randomly selected by the researcher and evaluated by the evaluators until the Gold Standard Seller data set was populated with the minimum number of 25 required candidates. All sellers were unique within the Gold Standard Seller data set – no duplicates.

Creating Gold Standard Feedbacks

Gold standard data was used to ensure the accuracy of the answers provided by the AMT workers. If answers provided by a worker significantly deviated from the gold standard, then there was a high degree of probability that the worker was poorly performing, not doing what was asked or was attempting to game the system.

A quality control technique used by Tang and Sanderson (2010), Sorokin and Forsyth (2008), Callison-Burch and Dredze (2010), and other researchers was randomly inserting (also known as salting) gold standard data into HITs to identify poorly performing, malicious or gaming workers. A worker did not know if the data to be evaluated came from the new data or from the gold standard. Workers that gave too many wrong answers to the gold standard were more likely to add noise to the overall results and needed to be filtered out. Noise is defined as the measure of deviation from the gold standard data (Hsueh, et al., 2009).

Gold standard data is a collection of preselected data that have a known set of answers. These answers are typically produced by one or more individuals who are trusted and a domain expert. Snow et al. (2008) demonstrated using multiple non-experts averaged out the noise resulting in the same quality answer as an expert. This technique was then applied by Snow et al. (2008) to produce gold standard data used in training sets as no gold standard data existed. Similarly research by Callison-Burch (2009) on machine translation quality and by Nowak and Ruger (2010) on tagging of images supported the findings that when combined non-expert judgments were equal to or better than human experts. *As no gold standard data set existed for determining whether or not an eBay buyer feedback comment is in negative-positive format or not, the technique of using multiple non-experts was used to create a Gold Standard Feedbacks.*

A gold standard with noise would only support cautious benchmarking as it requires performance of the workers be better than the baseline by more than that which can be attributed to the noise. As noise is defined as the measure of deviation from the gold standard data (Hsueh, et al., 2009), noise level is reduced as the inter-rater agreement for

an answer is increased. Noise is totally eliminated when all the raters are in agreement for an answer. In order to produce gold standard data with no noise, only answers with a strict metric were included. Strict metric is defined as the raters having consensus for an answer (Ku, et al., 2007). The use of strict metric (consensus) negated the need to measure inter-rater reliability using Cohen's Kappa.

After the production data was extracted, it was filtered leaving only qualified data which was 382,768 buyer feedback comments (see Chapter 4 for details). An eBay buyer feedback comment was randomly selected from the filtered population. The randomly selected feedback comment was reviewed by five qualified and dedicated coders. The number of coders selected was based on the recommendations of Snow et al. (2008), Callison-Burch (2009), and Klebanov and Beigman (2009). The same five coders were used to review all the feedback comments. The evaluation process followed the procedure as specified in the section entitled *Coding – Indentifying Buyer Feedback Comment as Negative-Positive or Not*. A feedback comment was only added to the Gold Standard Feedbacks if all the coders had a consensus in their answer. The same five coders were used to review all the feedback comments. Any feedback comment that did not have coder consensus was discarded. The population of feedback comments to be evaluated was 382,768 (see Chapter 4 for details). Because of the immense amount of data to be processed, it was broken down into 50 batches (See section *Implementation of Production Feedback HIT for Amazon Mechanical Turk*). The calculated size of a batch was about 7,700 feedback comments. The suggested quantity of gold standard data is from 5% to 10% of the population ("Crowdflower - gold standard," 2010). Multiplying 5% times the 7,700 batch size yielded a result of 385. The size of the Gold Standard

Feedbacks data set could range from 385 to 770. Feedback comments continued to be randomly selected by the researcher and evaluated by the coders until the Gold Standard Feedbacks data set was populated with the minimum number of 385 required candidates. All feedback comments were unique within the Gold Standard Feedbacks data set – no duplicates.

Implementation of Production Seller HIT for Amazon Mechanical Turk

A production Seller HIT was created and named *Research Production Seller*. No changes were made to the production Seller HIT, therefore it had exactly the same format as the prototype Seller HIT (See *Appendix H - Research Prototype Seller HIT*). The production Seller HIT was used by AMT workers to answer YES or NO to the judgment question – "Did the seller exhibit fraudulent type behavior to buyers?" As in the pilot test, the following qualifications were placed on the production Seller HIT - *Worker HIT Acceptance Rate*, *Research Qualification Native English Speaker*, and *Research Qualification Seller Test*. Workers were only given permission to gain access to the production Seller HITs after qualifying/passing all the qualifications.

Snow et al. (2008) indicated for a large set of HITs, an aggregate of four to six workers matched the results of a single domain expert. A majority of five workers was cited by Yan et al. (2010) as the best strategy in consistently achieving more than 95% accuracy. Heilman and Smith (2010), Pinchark et al. (2009), and Heymann and Garcia-Molina (2008) also determined that five workers was the optimum number per HIT. Based on this evidence the number of workers assigned to the production Seller HIT was set to five.

The quality of the workers' data was a critical concern. Especially as the only data to be collected was the final judgment answer of YES or NO contained in the 27th question of "Did the seller exhibit fraudulent type behavior to buyers?" The answers for the other 26 questions was not collected or analyzed as their sole purpose was to assist and guide the worker in gathering the necessary data to base their final judgment. The population of production Seller HITs was salted with 5% Gold Standard Sellers which were generated in a prior section entitled *Creating Gold Standard Sellers*. The 5% gold standard measure was within the suggested 5% to10% range ("Crowdflower - gold standard," 2010).

Experiments by other researchers using AMT demonstrated that first response to five one-cent HITs is 50-60% faster than a single five-cent task (Yan, et al., 2010). A review of financial incentives showed that increasing the micro-payment of HITs resulted in an increase in the quantity of work done, but not the quality of the work (Mason & Watts, 2009). The conclusion - If the micro-payment is too high, financial resources are wasted and inefficient workers are attracted. Elasticity of HIT throughput appears to be more dependent on the number of available online workers rather than the size of the HIT's micro-payment. The best strategy for a requester to adopt is start the first HIT batch at a low micro-payment and only increase the micro-payment size in subsequent HIT batches in the event of low worker response.

The total seller population of 502 was broken up into 10 batches for processing on AMT. This was done for three reasons. First, per the "best strategy for a requester" multiple batches provided a mechanism to adjust micro-payments (if necessary) while completing the data processing at the lowest possible cost. Second, small batches made it easier to monitor and block any mass attempt at gaming by comparing worker answers to

the Gold Standard Sellers. Third, it provided time to review the HIT's comment field for feedback from workers. Small batches made it possible to incorporate valid suggestions or make corrections without having to reprocess all the seller data. No suggestions were incorporated and no corrections were required for the production run. The comment field at the bottom of the HIT allowing for worker feedback replicated the technique used by Kosara and Ziemkiewicz (2010), Nowak and Ruger (2010), and Sorokin and Forsyth (2008).

Implementation of Production Feedback HIT for Amazon Mechanical Turk

A production Feedback HIT was created and named *Research Production Feedback* (See *Appendix L – Research Production Feedback HIT*). The number of questions (buyer feedback comments to be reviewed) in the production Feedback HIT was raised to ten compared to the one as seen in the prototype Feedback HIT. There were two compelling reasons to do this. The first was the need to cut costs as paying even at the lowest possible rate of $0.01 when multiplied by hundreds of thousands of HITs results in tens thousands of dollars for a total cost. As the time and effort required answering a single question was minimal, pooling multiple questions together into a single HIT was a viable and common practice used by requestors (Feng, et al., 2009; Finin, et al., 2010; Wenzel, 2008). Second, the quality of the workers' data was a critical concern. Multiple questions per HIT made it possible to salt each HIT with one or more Gold Standard Feedbacks (Finin, et al., 2010).

As in the pilot test, the following qualifications were placed on the production Feedback HIT - *Worker HIT Acceptance Rate, Research Qualification Native English Speaker*, and *Research Qualification Feedback Test*. Workers were only given

permission to gain access to the production Feedback HITs after qualifying/passing all the qualifications.

Snow et al. (2008) indicated for a large set of HITs, an aggregate of four to six workers matched the results of a single domain expert. A majority of five workers was cited by Yan et al. (2010) as the best strategy in consistently achieving more than 95% accuracy. Heilman and Smith (2010), Pinchark et al. (2009), and Heymann and Garcia-Molina (2008) also determined that five workers was the optimum number per HIT. Based on this evidence the number of workers assigned to the production Feedback HIT was set to five.

The quality of the workers' data was a critical concern. Each production Feedback HIT was salted with one Gold Standard Feedback which resulted in a gold standard measure of 10%. The 10% gold standard measure was within the suggested 5% to10% range ("Crowdflower - gold standard," 2010). The Gold Standard Feedbacks were generated in a prior section entitled *Creating Gold Standard Feedbacks*.

Experiments by other researchers using AMT demonstrated that first response to five one-cent HITs is 50-60% faster than a single five-cent task (Yan, et al., 2010). A review of financial incentives showed that increasing the micro-payment of HITs resulted in an increase in the quantity of work done, but not the quality of the work (Mason & Watts, 2009). The conclusion - If the micro-payment is too high, financial resources are wasted and inefficient workers are attracted. Elasticity of HIT throughput appears to be more dependent on the number of available online workers rather than the size of the HIT's micro-payment. The best strategy for a requester to adopt is start the first HIT batch at a

low micro-payment and only increase the micro-payment size in subsequent HIT batches in the event of low worker response.

The total feedback population of 382,768 was broken up into 50 batches for processing on AMT. This was done for four reasons. First, the massive size of the total feedback population was easier to handle when broken down into small batches. Second, per the "best strategy for a requester" multiple batches provided a mechanism to adjust micro-payments (if necessary) while completing the data processing at the lowest possible cost. Third, small batches made it easier to monitor and block any attempt at gaming by comparing worker answers to the Gold Standard Feedbacks. Fourth, it provided time to review the *Production Feedback* HIT's comment field for feedback from workers. Small batches made it possible to incorporate valid suggestions or make corrections without having to reprocess all the buyer feedback comment data. No suggestions were incorporated and no corrections were required for the production run. The comment field at the bottom of the HIT allowing for worker feedback replicated the technique used by Kosara and Ziemkiewicz (2010), Nowak and Ruger (2010), and Sorokin and Forsyth (2008).

Data Analysis

Data was entered into IBM SPSS Statistics 18 for Windows software application for analysis. Descriptive statistics were used to describe the sample and included frequency and percentages for nominal and categorical data. Means and standard deviations were applied to interval or ratio data. Per Bartlett, Kotrlik, and Higgins (2001) the following standards were used - for categorical data a 5% margin of error is acceptable; for continuous data a 3% margin of error is acceptable; for a dichotomous variable like

Fraudulent-Type Behavior Consensus a 5% margin of error is acceptable; and a 95% confidence level with $p = 0.5$ is acceptable for most basic research studies. For a dichotomous (divided or dividing into two sharply distinguished parts or classifications) variable, a 5% margin of error is acceptable (Bartlett, et al., 2001). A 95% confidence level and $p = 0.5$ were assumed for the research study as this is acceptable for most basic research studies (Bartlett, et al., 2001).

For each research question, a null hypothesis and alternative hypothesis are stated. Details are provided indicating the variables that would be used and statistical calculations that would be performed. Based on the principle of falsifiability (Gavin, 2008), statistical calculations were performed to test the null hypothesis for rejection. If the null hypothesis was rejected, then the alternative hypothesis would be examined to determine if that could be accepted. The result for each of the research questions is detailed in Chapter 4.

Research Question 1 (RQ1)

Does negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Null Hypothesis (H1o): Negative-positive type feedback comments from buyers do not predict evaluators' consensus of seller fraudulent-type behavior.

Alternative Hypothesis (H1a): Negative-positive type feedback comments from buyers do predict evaluators' consensus of seller fraudulent-type behavior.

A logistic regression was conducted to assess whether or not negative-positive type feedback comments from buyers predicted evaluators' consensus of seller fraudulent behavior. For this analysis, the independent (predictor) variable was *Negative-Positive*

Consensus field and the dependent (criterion) variable was seller behavior. Seller behavior was represented by the *Fraudulent-Type Behavior Consensus* field.

Logistic regression (also known as the logistic model or logit model) was the appropriate statistic to analyze the data as the research question is to examine how an independent variable predicts a mutually exclusive dichotomous (divided or dividing into two sharply distinguished parts or classifications) criterion variable.

The Chi-square significance test was used to test the null hypothesis of no association between the independent variable (Negative-Positive Consensus) and the dependent variable (Fraudulent-Type Behavior Consensus).

Research Question 2 (RQ2)

Does the number of negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Null Hypothesis (H2o): The number of negative-positive type feedback comments does not predict evaluators' consensus of seller fraudulent-type behavior.

Alternative Hypothesis (H2a): The number of negative-positive type feedback comments predicts evaluators' consensus of seller fraudulent-type behavior.

A logistic regression was conducted to assess whether or not the number of negative-positive type feedback comments predicted evaluators' consensus of seller fraudulent behavior. Logistic regression was the appropriate way to analyze the data as research question 2 was to examine how an independent variable predicts a mutually exclusive dichotomous (divided or dividing into two sharply distinguished parts or classifications) criterion variable. In this case, the independent variable was obtained by counting the number of negative-positive comments to achieve a continuous value. The dependent or

criterion variable was consensus of seller fraudulent-type behavior which was dichotomized (1 = Y, 0 = N).

Research Question 3 (RQ3)

For each seller will negative-positive type feedback comments from buyers fall into a cluster?

Null Hypothesis (H3o): For each seller will negative-positive feedback comments do not fall into a cluster.

Alternative Hypothesis (H3a): For each seller negative-positive feedback comments fall into a cluster.

For the testing of whether or not negative-positive type feedback comments fell into a cluster, a Chi-square test of Independence was used. A cluster was determined when negative-positive type comments were found grouped around traditional comments in the sellers' feedback transaction history. For example, when negative-positive type comments were separated by two traditional comments, and then followed by another occurrence of a negative-positive comment, a cluster was identified. In a cluster, the negative-positive type comments could be separated by as many as two traditional comments. For this analysis, the feedback either fell into the cluster (Yes) or not (No).

Summary

The objective of the research study was to determine if the presence of negative-positive type feedback comments by buyers is a predictor that a seller is behaving fraudulently. The correlational research design provided for discovering relationships between variables, measuring the degree and direction of relationships, and from the discovered relationships predictions could be made. The correlational research design

(see Appendix A) was implemented using an automated data collection agent in order to efficiently sift through the massive quantities of data on eBay and locate the qualified sellers. The methodology was constructed with the goal of reducing the subjectivity and increasing the reliability of categorizing seller behavior as honest or fraudulent and buyer feedback comments as negative-positive or not.

Chapter 4

Results

Introduction

This chapter provides a presentation of the research findings and analysis of the data that was collected. It includes a review of the objective of the research study; the data collection procedure; the three research questions (with null and alternative hypothesis for each); data analysis for the research questions; and a summary of results.

Objective of the Study

In a forensic case study of an opportunistic seller by Nikitov and Stone (2006), it was found buyers sometimes embedded negative comments in positive feedback as a means of avoiding retaliation from sellers and damage to their reputation. This category of positive feedback is described as "negative-positive" feedback. An example of negative-positive feedback is "Good product, but slow shipping." The objective of this study was investigating the concept of using negative-positive feedback as a signature to identify potential opportunistic sellers in an online auction population.

Data Collection

The issue of obtaining a sufficient population of sellers that exhibited fraudulent type behavior was previously discussed in the section entitled *Population Size* in Chapter 3 Methodology. Each of the three full data extractions from the pilot study found relatively small populations of unique buyers - 406, 438, and 446 (see Figure 11). A small number of unique buyers could adversely effect the research's data analysis as the number of fraudulent sellers within the eBay member population is reported to be very small. Per

prior cited research studies, the distribution of fraudulent sellers appears to be skewed and focused on specific categories. Based on this information, the category of *Computers and Networking: PC Laptops and Notebooks* was selected for its potential in containing multiple fraudulent sellers.

The web crawler (see Appendix B) used was custom designed for the eBay website to retrieve the raw data. The web crawler retrieved web pages, parsed the webpages to find the required data, determined if the found data met the selection criteria, and stored the qualified data for later analysis in a Comma Separated Variables (CSV) ASCII file as specified in Appendix C. The search space used by the web crawler was bounded by all sellers in the category of *Computers and Networking: PC Laptops and Notebooks*. The result of the production full data extraction was a data set composed of 467,071 buyer feedback comments created by 502 unique eBay sellers.

The evaluators reviewed the unique eBay sellers and identified based on majority rule (3 of 5) the sellers exhibiting fraudulent behavior in the *Computers and Networking: PC Laptops and Notebooks* category. Out of a total of 502 unique eBay userids, the number of sellers identified as exhibiting fraudulent behavior was 19. This translated to 3.78% (19/502) of the total sellers were exhibiting fraudulent behavior. This number was sufficiently large enough to eliminate the need to rerun the web crawler using a new category or multiple categories in order to locate more eBay sellers exhibiting fraudulent type behavior.

Descriptive Statistics of the Collected Data

A summary of the collected data from the web crawler run can be seen in Figure 22.

		Number	Percent
Total Feedback Comments		467,071	100.00%
	Minus Negative	2,422	(0.52%)
	Minus Neutral	2,757	(0.59%)
	Minus Blank	1,048	(0.22%)
Remaining Positive		460,844	98.67%
	Minus As Buyer	74,865	(16.03%)
	Minus Non-English	3,211	(0.69%)
Qualified Buyer Feedback Comments		382,768	81.95%

Figure 22. Analysis of Extracted Production Data

The *Total Feedback Comments* found was 467,071. All unqualified records were deleted from the *Total Feedback Comments* population by the researcher:

Minus Negative – Any feedback comment with a feedback type of *Negative* was removed as negative-positive feedback requires a feedback type of positive.

Minus Neutral – Any feedback comment with a feedback type of *Neutral* was removed as negative-positive feedback requires a feedback type of positive.

Minus Blank – Any feedback comment with a feedback type of *blank* was removed as negative-positive feedback requires a feedback type of positive. eBay will set a feedback type to blank for partially deleted or censured comments.

The *Remaining Positive* number of 460,844 contained only feedback comments that had a feedback type of *Positive*.

Most eBay members switch between the roles of seller and buyer. Each qualified seller's eBay member profile can contain feedback for both roles. Therefore all feedback comments in which the seller was acting as a buyer were eliminated by the researcher as designated by *Minus As Buyer*.

Inclusion of non-English buyer feedback comments could result in ambiguity due to translation plus the additional expense of hiring translators. *The assumption was made that buyer feedback comments are consistent regardless of the language in which they are composed.* That is to say a buyer's compliment or complaint about a seller in the form of a feedback comment is independent of the spoken/written language used by the buyer. Therefore non-English buyer feedback comments which constitute less than 1% of the total population were treated as outliers and excluded from the data to be analyzed. This exclusion made by the researcher was indicated by *Minus Non-English*.

As designated by *Qualified Buyer Feedback Comments* - the total number of buyer feedback comments that needed to be reviewed by the coders was 382,768.

Amazon Mechanical Turk Processing – Sellers

From the total population of the 502 unique eBay sellers, one seller at a time was randomly pulled and evaluated by dedicated raters (evaluators) until 25 sellers were found having consensus of all evaluators (5 of 5). The seller was then added to the Gold Standard Sellers data set. As a matter of record, all the Gold Standard Sellers were classified as honest. This left 477 eBay sellers which needed to be processed. For quality control purposes, the 25 Gold Standard Sellers were added back into pool – resulting in 502 unique eBay sellers to be reviewed by Amazon Mechanical Turk (AMT) evaluators. The sellers were randomly divided among the ten batches for processing on AMT.

A total of 19 sellers were designated by the AMT evaluators as fraudulently behaving sellers based on majority rule (3 of 5). An additional 18 sellers were tagged by AMT evaluators as *potentially* fraudulent sellers, but each of these sellers only received one or two votes which were insufficient to make a majority and be classified as fraudulent

sellers. Time for processing was approximately two days. The micro-payment was $0.30 per Human Intelligence Test (HIT) with five assignments.

Amazon Mechanical Turk Processing – Buyer Feedback Comments

As designated by *Qualified Buyer Feedback Comments* (see Figure 22) - the total number of buyer feedback comments that needed to be reviewed by the coders was 382,768. As 50 batches would be used, the estimated size per batch was 7,700. The gold standard was set to 5% of the batch size which was 385 feedback comments (7,700 X 0.05). Feedback comments were randomly pulled and evaluated by dedicated raters (coders) until 385 feedback comments were found having consensus of all coders (5 of 5). The feedback comment was then added to the Gold Standard Feedbacks data set. This left 382,383 buyer feedback comments remaining to be evaluated (382,768 – 385).

The remaining feedback comments were randomly divided among 50 batches for processing on AMT. Each HIT was composed of ten feedback comments. Nine feedback comments for the HIT came from the batch. For quality control purposes, the tenth feedback comment in each HIT was randomly salted with one of the 385 Gold Standard Feedbacks. Repetitive use of Gold Standard Feedbacks in the batches was not an issue as many feedback comments like “Good seller!” were commonly used by multiple buyers.

Out of 382,768 feedback comments, 2,247 were identified by coders as negative-positive feedback comments based on majority rule (3 of 5). Thus negative-positive feedback comments constituted only 0.59% of the total qualified positive buyer feedback comments (2,247/382,768). Time for processing was approximately five days. The micro-payment was $0.01 per HIT with five assignments.

Amazon Mechanical Turk – Quality Control

The techniques of qualification tests, multiple worker assignments per HIT (plurality), minimum work time per HIT, gold standard data, and advice of auditing were used to ensure the reliability of raters. The *Research Qualification Seller Test* was set to manual which required the researcher to individually review and authorize each worker as qualified. Seven AMT workers were rejected for the *Research Qualification Seller Test*. No seller production HITs were rejected. For the feedback production HITS, the work (in entirety) done by three AMT workers was rejected. One AMT worker was obvious gaming as only N (No) was entered as an answer to every question. The other two AMT workers failed to correctly answer multiple Gold Standard Feedbacks, it was concluded that they were either gaming by randomly answering or had poor performance. When HITs were rejected and released for processing by other workers, a reject message was sent to the effected AMT worker explaining that the required level of quality was not met. The rejected AMT worker was then blocked from working on any more HITs.

Analysis Delimitations

Data was entered into IBM SPSS Statistics 18 for Windows software application for analysis. Descriptive statistics were used to describe the sample and included frequency and percentages for nominal and categorical data. Means and standard deviations were applied to interval or ratio data. Per Bartlett, Kotrlik and Higgins (2001) the following standards were used - for categorical data a 5% margin of error is acceptable; for continuous data a 3% margin of error is acceptable; for a dichotomous variable like *Fraudulent-Type Behavior Consensus* a 5% margin of error is acceptable; and a 95% confidence level with $p = 0.5$ is acceptable for most basic research studies. For a

dichotomous (divided or dividing into two sharply distinguished parts or classifications) variable, a 5% margin of error is acceptable (Bartlett, et al., 2001). A 95% confidence level and $p = 0.5$ were assumed for the research study as this is acceptable for most basic research studies (Bartlett, et al., 2001).

For each research question, a null hypothesis and alternative hypothesis are stated. Details are provided indicating the variables that were used and statistical calculations that were performed. Based on the principle of falsifiability (Gavin, 2008), statistical calculations were performed to test the null hypothesis for rejection.

Research Question 1 (RQ1)

Does negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Null Hypothesis (H1o): Negative-positive type feedback comments from buyers do not predict evaluators' consensus of seller fraudulent-type behavior.

Alternative Hypothesis (H1a): Negative-positive type feedback comments from buyers do predict evaluators' consensus of seller fraudulent-type behavior.

A logistic regression was conducted to assess whether or not negative-positive type feedback comments from buyers predicted evaluators' consensus of seller fraudulent behavior. For this analysis, the independent (predictor) variable was *Negative-Positive Consensus* field and the dependent (criterion) variable was seller behavior. Seller behavior was represented by the *Fraudulent-Type Behavior Consensus* field.

Logistic regression was the appropriate statistic to analyze the data as the research question was to examine how an independent variable predicts a mutually exclusive dichotomous (divided or dividing into two sharply distinguished parts or classifications) criterion variable. Results of the logistic regression are displayed in Figure 23.

Independent Variable	*B*	*SE*	Wald	*p*	Exp(B)
Negative Positive Feedback	0.60	0.06	97.27	< 0.001	1.82
Model	Y = -2.40 + 0.06* Negative-Positive Feedback				

Figure 23. Negative-Positive Feedback Comments Predicting Evaluators' Consensus of Seller Fraudulent Behavior

Any *p* less than 0.05 are significant. As seen in Figure 23, the *p* for the logistic coefficient was < 0.001 which means the logistic coefficient was statistically significant.

The Chi-Square test calculation performed was represented by Chi^2 with one degree of freedom. The degree of freedom is equal to the number of standard normal deviates being summed – one. The resulting Chi-Square calculation was Chi^2 (1) = 84.40.

The *p* is the probability of observing a test statistic at least as extreme in a Chi-Square distribution. Any *p* less than 0.05 are significant. Using a *Table of* χ^2 *Value vs. P-Value* with Chi^2 (1) = 84.40, the resulting *p* was < 0.001 which was classified as statistically significant (Fisher, 1995).

A Chi-Square significance test was used to test the null hypothesis of no association between the independent variable (*Negative-Positive Consensus*) and the dependent variable (*Fraudulent-Type Behavior Consensus*). The Chi-Square test was significant with Chi^2 (1) = 84.40, $p < 0.001$. It clearly rejected the null hypothesis that no independent variable (*Negative-Positive Consensus*) was correlated to the dependent variable (*Fraudulent-Type Behavior Consensus*). With the Chi-Square test as significant and the logistical regression's *p* as significant, it suggested that negative-positive type feedback comments from buyers predicts evaluators' consensus of seller fraudulent behavior.

Two descriptive measures of goodness-of-fit are Cox and Snell (1989) and Nagelkerke (1991). In linear regression, R^2 has a clearly defined definition as the proportion of the variation in the dependent variable that can be explained by the predictor(s) in the linear model. Several attempts have been made to devise an equivalent of R^2 for the logistic model. None currently render the meaning of the variance (Menard, 2000). None correspond to predictive efficiency. For these two reasons, the two R^2 indices were not included in the evaluation of the logistic model.

Wald statistics (Harrell, 2001) is for testing the significance of the explanatory (independent) variables in the logistics model. As only a single independent variable *Negative-Positive Consensus* was used, it rendered this statistic moot.

In Figure 23, B represents the regression coefficient for the predictor which is *Negative-Positive Consensus*. A positive regression coefficient means that the explanatory (independent) variable increases the probability of explanatory variable decreases the probability of the outcome. A large regression coefficient means that the explanatory variable strongly influences the probability of the outcome. A near-zero regression coefficient means the explanatory variable has little influence on the probability of the outcome. The value of B was 0.6 which showed an increase in probability of the outcome, but with a less than one multiplier the explanatory variable influence was moderated.

The exponent of B in the logistic regression yields the odds ratio. Odds ratios whose confidence limits are greater or less than one are statistically significant. For SPSS the odds ratio is labeled as Exp(B). The logit b = 0.6 in the B column in Figure 23 resulted in a corresponding odds ratio [Exp(B)] of 1.82. The results of the logistic regression

suggested that as buyers tended to have negative-positive feedback; sellers were 1.82 times more likely to be fraudulent.

		Negative-Positive Consensus		
Fraudulent Type Behavior		**Yes**	**No**	**Total**
	Yes	15	4	19
	No	101	382	483
			Total Sellers	502

Figure 24. Comparing Coders Negative-Positive Feedback Consensus to Seller Fraudulent Behavior by Evaluators

The actual counts of sellers based on coders negative-positive feedback consensus compared to sellers exhibiting fraudulent type behavior as found by the evaluators is summarized in Figure 24. The off-diagonal cells in the table containing the values of four and 101 showed the lack of buyer negative-positive feedback comments when sellers were not exhibiting fraudulent type behavior. Conversely, the other off-diagonal cell in the table containing the values of 15 and 382 showed the presence of buyer negative-positive feedback comments and when sellers exhibited fraudulent type behavior.

The null hypothesis H1o was rejected for RQ1.

Research Question 2 (RQ2)

Does the number of negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Null Hypothesis (H2o): The number of negative-positive type feedback comments does not predict evaluators' consensus of seller fraudulent-type behavior.

Alternative Hypothesis (H2a): The number of negative-positive type feedback comments predicts evaluators' consensus of seller fraudulent-type behavior.

A logistic regression was conducted to assess whether or not the number of negative-positive type feedback comments predicted evaluators' consensus of seller fraudulent

behavior. The independent (predictor) variable was the *Number of Negative-Positive Feedbacks Comments* and the dependent (criterion) variable was seller behavior. Seller behavior was represented by *Fraudulent-Type Behavior Consensus* (1 = Y, 0 = N). In this case, the independent variable was obtained by counting the number of negative-positive feedback comments for each seller to achieve a continuous value. The total seller population was 502 sellers. A total of 19 sellers were previously identified by evaluators as exhibiting fraudulent behavior. The remaining 483 sellers were previously identified as honest by the evaluators. The total number of buyer feedback comments previously categorized by coders as negative-positive type was 2,247.

Logistic regression was the appropriate statistic to analyze the data as the research question was to examine how an independent variable predicts a mutually exclusive dichotomous (divided or dividing into two sharply distinguished parts or classifications) criterion variable. Results of the logistic regression are displayed in Figure 24.

Independent Variable	***B***	***SE***	**Wald**	***p***	**Exp(B)**
Number of Negative-Positive Feedback Comments	0.04	0.01	13.02	< 0.001	1.04

Model Y = -3.51 + 0.04* Number of Negative-Positive Feedback Comments

Figure 25. Number of Negative-Positive Feedback Comments Predicting Evaluators' Consensus of Seller Fraudulent Behavior

Any *p* less than 0.05 are significant. As seen in Figure 25, the *p* for the logistic coefficient was < 0.001 which means the logistic coefficient was statistically significant.

The Chi-Square test calculation performed was represented by Chi^2 with one degree of freedom. The degree of freedom is equal to the number of standard normal deviates being summed – one. The resulting Chi-Square calculation was Chi^2 (1) = 10.92.

The *p* is the probability of observing a test statistic at least as extreme in a Chi-Square distribution. Any *p* less than 0.05 are significant. Using a *Table of χ^2 Value vs. P-Value* with Chi^2 (1) = 10.92, the resulting *p* was < 0.001 which was classified as statistically significant (Fisher, 1995).

A Chi-Square significance test was used to test the null hypothesis of no association between the independent variable (*Number of Negative-Positive Feedback Comments*) and the dependent variable (*Fraudulent-Type Behavior Consensus*). The Chi-Square test was significance with Chi^2 (1) = 10.92, $p < 0.001$. It clearly rejected the null hypothesis that no independent variable (*Number of Negative-Positive Feedback Comments*) was linearly correlated to the log odds of the dependent variable (*Fraudulent-Type Behavior Consensus*). With the Chi-Square test as significant and the logistical regression's *p* as significant, it suggested that the number of negative-positive feedback comments from buyers predicts evaluators' consensus of seller fraudulent behavior.

Two descriptive measures of goodness-of-fit are Cox and Snell (1989) and Nagelkerke (1991). In linear regression, R^2 has a clearly defined definition as the proportion of the variation in the dependent variable that can be explained by the predictor(s) in the linear model. Several attempts have been made to devise an equivalent of R^2 for the logistic model. None currently render the meaning of the variance (Menard, 2000). None correspond to predictive efficiency. For these two reasons, the two R^2 indices were not included in the evaluation of the logistic model.

Wald statistics (Harrell, 2001) is for testing the significance of the explanatory (independent) variables in the logistics model. As only a single independent variable *Negative-Positive Consensus* was used, it rendered this statistic moot.

In Figure 24, B represents the regression coefficient for the predictor which is *Negative-Positive Consensus*. A positive regression coefficient means that the explanatory (independent) variable increases the probability of explanatory variable decreases the probability of the outcome. A large regression coefficient means that the explanatory variable strongly influences the probability of the outcome. A near-zero regression coefficient means the explanatory variable has little influence on the probability of the outcome. The value of B was 0.04 which showed an increase in probability of the outcome, but with a less than one multiplier the explanatory variable influence was highly moderated.

The exponent of B in the logistic regression yields the odds ratio. Odds ratios whose confidence limits are greater or less than one are statistically significant. For SPSS the odds ratio is labeled as Exp(B). The logit b = 0.04 in the B column in Figure 23 resulted in a corresponding odds ratio [Exp(B)] of 1.04. The results of the logistic regression suggested that for every one unit increase in the number of negative-positive feedback comments, sellers were 1.04 times more likely to be fraudulent.

The null hypothesis H2o was rejected for RQ2.

Research Question 3 (RQ3)

For each seller will negative-positive type feedback comments from buyers fall into a cluster?

Null Hypothesis (H3o): For each seller will negative-positive feedback comments do not fall into a cluster.

Alternative Hypothesis (H3a): For each seller negative-positive feedback comments fall into a cluster.

A Chi-Square test was conducted to assess whether or not negative-positive type feedback comments fall into a cluster. A cluster was determined when negative-positive type comments were found grouped around traditional [not negative-positive] comments in the sellers' feedback comments history. For example, when negative-positive type comments were separated by two traditional comments, and then followed by another occurrence of a negative-positive comment, a cluster was identified. In a cluster, the negative-positive type comments could be separated by as many as two traditional comments. For this analysis, the feedback either fell into the cluster (Yes) or not (No). Results of the Chi-Square test are displayed in Figure 26.

Chi^2	*df*	*p*	No Cluster	Cluster	Expected
426.18	1	<.001	694	109	401.5

Figure 26. Chi-Square on Negative-Positive Feedback Comments Falling into a Cluster

The *p* is the probability of observing a test statistic at least as extreme in a Chi-Square distribution. Any *p* less than 0.05 are significant. Using a *Table of χ^2 Value vs. P-Value* with Chi^2 (1) = 426.18, the resulting *p* was < 0.001 which was classified as statistically significant (Fisher, 1995).

The Chi-Square test calculation performed was represented by Chi^2 with one degree of freedom. The degree of freedom is equal to the number of standard normal deviates being summed – one. The resulting Chi-Square calculation was Chi^2 (1) = 426.18. The results suggested that negative-positive type feedback comments did not fall into a cluster, therefore the null hypothesis was accepted. No clustering of negative-positive type feedback was revealed for 694 sellers and 109 sellers did reveal clustering of negative-positive type comments. The expected count for each cell was 401.5 [(694+109)/2)] suggesting that fewer sellers than expected had negative-positive type comments that was clustered.

The null hypothesis H3o was accepted for RQ3.

Summary of Results

The research was divided into four parts – collecting the data using a web crawler, manually scrubbing the collected data, coding the data using crowdsourcing, and performing data analysis on the three research questions using SPSS.

The web crawler searched the category of *Computers and Networking: PC Laptops and Notebooks* extracting raw data consisting of 467,071 eBay buyer feedback comments. After scrubbing the data to only include qualified buyer feedback comments and eliminating outliers consisting of non-English comments, the remaining dataset to be processed contained 382,768 buyer feedback comments. From the scrubbed dataset, a total of 502 unique eBay sellers were identified.

Using traditional dedicated raters to process the collected data was not viable due to extensive time required and high monetary cost. An alternative solution of crowdsourcing was used with service provided by Amazon Mechanical Turk.

Crowdsourcing proved viable as all the work was processed in less than seven days with a considerable cost savings compared to traditional dedicated raters. Multiple techniques were used to ensure data quality - qualification tests and data quality techniques of multiple worker assignments per HIT (plurality), minimum work time per HIT, gold standard data, and advice of auditing.

Evaluators identified 19 out of the 502 unique eBay sellers as exhibiting fraudulent behavior. This translated into 3.78% of the sellers classified as behaving fraudulently. The remaining 483 sellers were classified as honest.

Coders categorized 2,247 out of 382,768 buyer feedback comments as negative-positive type. This translated into 0.59% of the total buyer feedback comments were negative-positive type.

The research study focused on the determining if negative-positive type feedback comments by buyers are a predictor that a seller is behaving fraudulently. Three research questions were used in framing an answer for this primary question.

For research question 1 - *Does negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?* The null hypothesis of *Negative-positive type feedback comments from buyers do not predict evaluators' consensus of seller fraudulent-type behavior* was rejected based on the results of the logistic regression and Chi-Square test. The results of the logistic regression suggested that as buyers tended to have negative-positive feedback; sellers were 1.82 times more likely to be fraudulent.

For research question 2 - *Does the number of negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?*

The null hypothesis of *The number of negative-positive type feedback comments does not predict evaluators' consensus of seller fraudulent-type behavior* was rejected based on the results of the logistic regression and Chi-Square test. The results of the logistic regression suggested that for every one unit increase in the number of negative-positive feedback comments, sellers were 1.04 times more likely to be fraudulent.

For research question 3 - *For each seller will negative-positive type feedback comments from buyers fall into a cluster?* The null hypothesis of *For each seller will negative-positive feedback comments do not fall into a cluster* was accepted based on the results of the Chi-Square test.

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Conclusions

The research had a good outcome as an exploratory study. It identified a variable that appears to be a new indicator for identifying potentially fraudulent sellers in eBay. The research study focused on the determining if negative-positive type feedback comments by buyers are a predictor that a seller is behaving fraudulently.

The findings of Zhang (2006) showed that eBay buyers provided 99% positive comments, 0.7% negative comments, and 0.3% neutral comments. In January 2008, eBay made a fundamental change to the feedback system where sellers could leave only positive or neutral ratings for buyers. That meant buyers were free to leave negative feedback without fear of feedback-retaliation. Logically, buyers should have responded by providing negative feedback when appropriate. Gregg and Scott (2008) reported that although the new policy has been in effect for a year, the status quo remained with eBay still reporting less than 1% negative feedback; most members had a 99% or higher feedback rating; and the percentage of fraudulent transactions continued to rise.

This research study was conducted almost two years after the change in the eBay feedback system was implemented. It found almost exactly the same conditions previously reported by Gregg and Scott (2008) - eBay still reporting less than 1% negative feedback and most members had a 99% or higher feedback rating. The research study found out of 467,071 buyer feedback comments – 98.67% were positive comments, 0.52% were negative comments, 0.59% were neutral comments, and 0.22% were blank

comments (see Chapter 3 - Figure 22). Thus the premise for the research study – buyer reluctance to report negative feedback – was confirmed.

Three research questions were used in framing an answer for the research objective – Is the presence of negative-positive type feedback comments by buyers a predictor that a seller is behaving fraudulently? Each research question is presented and its findings from Chapter 4 analyzed.

Research Question 1 (RQ1)

Does negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

The null hypothesis of *Negative-positive type feedback comments from buyers do not predict evaluators' consensus of seller fraudulent-type behavior* was rejected based on the results from the logistic regression. The Chi-Square test was significance with Chi^2 (1) = 84.40, $p < 0.001$. As the null hypothesis was rejected, the alternative hypothesis H1a must be accepted.

The results of the logistic regression suggested that as buyers tended to have negative-positive feedback; sellers were 1.82 times more likely to be fraudulent. This was evidence that the presence of even a single negative-positive feedback type comment had a strong correlation with a seller exhibiting fraudulent behavior.

Prior studies by Goes, Tu, and Tung (2009), Gregg and Scott (2008), and Pandit, Chau, Wang and Faloutsos (2007) used only negative feedback ratings and comments to identify sellers as fraudulent. As a signature, negative feedback ratings composed only 0.7% per Zhang (2006) and 0.52% per the research study of the total feedback population. Negative-positive feedback comments found in the research study composed

0.48% (2,247/467,071) of the total feedback population. Like the other signatures - negative and neutral feedback ratings - negative-positive type feedback composed only a small percentage of the total feedback population.

It would be more appropriate to measure negative-positive feedback comments within only the total positive feedback population. The prerequisite for a negative-positive feedback is the requirement of the feedback type being positive. Within this smaller population, negative-positive feedback composed 0.59% (2,247/382,768) of the total positive feedback population. This was a slightly higher percent than the signature indicator of a negative rating at 0.52%. The larger presence of negative-positive type feedback would be consistent based on two buyer perceptions held by Nikitkov and Stone (2006). First - a positive feedback rating from the buyer even with a negative comment would not invite feedback retaliation from the seller. Second – the negative statement is concealed within the feedback comment, making it a more socially acceptable and a less drastic action than a blatant negative rating with negative comment.

Research Question 2 (RQ2)

Does the number of negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

The null hypothesis of *The number of negative-positive type feedback comments does not predict evaluators' consensus of seller fraudulent-type behavior* was rejected based on the results from the logistic regression. The Chi-Square test was significance with Chi^2 (1) = 10.92, $p < 0.001$. As the null hypothesis was rejected, the alternative hypothesis H2a must be accepted.

The results of the logistic regression suggested that for every one unit increase in the number of negative-positive feedback comments, sellers were 1.04 times (4%) more likely to be fraudulent.

The finding of the correlation between an increasing number of negative-positive feedback comments and an increasing probability that a seller was acting fraudulently was expected. A similar relationship was found in prior studies by Goes, Tu, and Tung (2009), Gregg and Scott (2008), and Pandit, Chau, Wang and Faloutsos (2007) with negative feedback. These researchers noted that a single incidence of a negative rating would normally not be sufficient to indicate that a seller was fraudulent. The presence of multiple negative ratings increased the probability that a seller was acting fraudulently.

Research Question 3 (RQ3)

For each seller will negative-positive type feedback comments from buyers fall into a cluster?

The Chi-Square test was significance with Chi^2 (1) = 426.18, $p < 0.001$. No clustering of negative-positive type feedback was revealed for 694 sellers and 109 sellers did reveal clustering of negative-positive type comments The expected count for each cell was 401.5 [(694+109)/2)] indicating that fewer sellers than expected had negative-positive type comments that was clustered. The results suggested that negative-positive type feedback comments did not fall into a cluster, therefore the null hypothesis of *For each seller will negative-positive feedback comments do not fall into a cluster* was accepted.

Limitations

The major limitations of the research study were tied to three issues – fraudulent sellers, data source, and the ability to generalize the results.

Locating eBay sellers that were behaving fraudulently was a difficult task. The primary source – eBay - refused to provide any type of information about why a seller was suspended or NARU (Not A Registered User). This complicated the study by requiring an extensive methodology to sift through and identify potentially fraudulent sellers. An identified fraudulent seller could not be stated definitively as a fraudster, but rather as having a high probability of exhibiting fraudulent behavior.

The second limitation was the data source. A custom web crawler was used for data collection. A web crawler had the obvious advantages of speed, ability to extract a massive quantity of data, and accuracy. When compared to more traditional approaches like surveys and experiments, it did not allow the researcher to establish controls that could have made the data a better fit for analysis. The raw public data was not as neat and clean when compared to a survey which is designed with analysis in mind. After data collection, considerable effort was required to convert the raw public data into a form that could actually be utilized for analysis. Mechanisms were required in the methodology to insure an unbiased data collection and conversion.

The third limitation was the ability to generalize the research results beyond eBay. Because it is the 800 pound gorilla in the online auction market and has been extensively studied by prior researchers, eBay was the logical choice. However, eBay has other characteristics that might effect outcomes. One factor that could effect applying the results to another online auction company is use of a different feedback mechanism. For eBay, once the buyer or seller posts feedback it is immediately available to the other party. Other online auctions have different feedback mechanisms. For example, a company could prevent viewing of feedback until either both parties post feedback or the

time allowed to post feedback has expired. This prevents either party from being influenced by the feedback from the other thus potentially negating the issue of tit-for-tat with negative feedback. Another limiting factor is the type of online market studied. As eBay is an online auction, attempting to apply the results to a fixed-priced online market like Amazon or Half.com would not be appropriate as other mitigating conditions could be present.

Causal Direction

Correlational research design will not identify the causes or reasons for the observed behavior. This is because a correlational relationship between variables could be the result of an outside source. Based on this possibility, it must be understood that the correlation does not necessarily explain cause and effect. Hence the maxim – Correlation does not equal causation (Aldrich, 1995).

Under certain conditions, it is possible to have a high degree of confidence that there is causality between two variables. Determining the direction of causality can be difficult or impossible to quantify. Casual direction can be hinted if information about time is available. This is because a cause must precede its effects under classic Newtonian physics and natural laws. The type of data used was time-stamped historical transaction logs which provided the ability to indicate the direction of causality.

The direction of causality was from seller to buyer. A buyer cannot provide feedback about the item purchased or the seller until after the item is physically received. Making the statement “Good packaging, but slow shipping” is not logical or grounded until the package is physically in the buyer’s possession. An explicit negative feedback or hidden

complaint using a negative-positive statement in positive feedback by a buyer would be in response to negative or fraudulent action by a seller.

Implications and Recommendations

The research study has both theoretical and practical implications. It presented a conceptual basis for the study of using negative-positive buyer feedback comments to identify fraudulently behaving sellers. Empirical evidence from the study proved that negative-positive type feedback comments do exist although they constitute 0.59% of the total positive feedback population. Statistical analysis supported the hypothesis that there was a correlation between negative-positive type feedback comments and a seller behaving fraudulently. In addition, it supported a correlation between the number of negative-positive feedback type comments and an increasing probability that a seller was behaving fraudulently.

The contributions to knowledge were twofold. First was identifying a potential new signature – *negative-positive type feedback comments* - for identifying fraudulently behaving sellers. Second was *demonstrating the use of crowdsourcing as an effective and cost efficient means to detect fraudulent sellers in online auctions*.

In January 2008, eBay made a fundamental change to the feedback system where sellers could leave only positive or neutral ratings for buyers. That meant buyers were free to leave negative feedback without fear of feedback-retaliation. One year after the eBay policy was implemented, Gregg and Scott (2008) found that buyers were still reluctant to provide negative feedback. As a contribution to the research literature, this research study extended the work of Gregg and Scott (2008) by finding that buyers were still reluctant even after two years to provide negative feedback.

As potential practical contribution – the new signature combined with crowdsourcing could be used by eBay's Security Department for detecting potentially fraudulent sellers. After a feedback comment is entered by an eBay buyer; it could be automatically processed. Only a positive rating with a feedback comment would need to be evaluated. Basic textual analysis could be performed on the feedback comment looking at syntax, structure, and content. Only a qualified positive feedback comments would need to be evaluated by placing an API call to AMT. The processing does not need to be real-time, but timely updating of an eBay seller's profile would reduce the window of opportunity for a fraudulent seller. Although hidden from public view, the negative-positive correlation factor when added to an eBay user's profile could be internally used by eBay as one more tool in identifying and monitoring potentially fraudulent sellers.

Textual analysis of the buyer feedback comments was gross and not granular. The text contained in a positive feedback comment was evaluated in entirety as a binary - negative-positive type feedback (Y) or not (N). Future research using a more detailed data mining of the feedback comment texts could provide scalar indicators or predictors for identifying fraudulent sellers. For example – "Good packing, slow shipping" would be a low level indicator as a seller could live in a remote location or typically uses a slow shipper. "Good packing, but product was not new" would be a higher level indicator of fraudulent behavior as the seller said the product was new, but sent used.

A question that could be asked is - Can *Buyer Complaint* category in negative-positive type feedback comments be used to fine tune indicators or predictors for identifying fraudulent sellers? For future research, textual analysis of the buyer feedback comment could be performed based on *Buyer Complaint* categories – product, shipping,

communication, and other (non-specific). Each negative-positive buyer feedback comment would be classified into one or more of the *Buyer Complaint* categories using a vector like Φ_n = [<product>, <shipping>, <communication>, <other>]. This text mining technique of feedback comment text has been used in prior research by Ghose, Ipeirotis, Sundararajan (2005).

Prior studies by Goes, Tu, and Tung (2009), Gregg and Scott (2008), and Pandit, Chau, Wang and Faloutsos (2007) used only negative feedback ratings and comments to identify sellers as fraudulent. An interesting future study would be comparing the two methods – negative feedback and negative-positive feedback. The proposed study would further validate each method and provide a comparative measure of their effectiveness in identifying fraudulent sellers. Locating fraudulent sellers for the proposed study would need to be done independently using a grounded method like a survey of eBay buyers, police reports, etc.

Summary

Online auction fraud represents a serious threat to e-commerce and undermines online trust. As fraud is pervasive, growing in use, and difficult to detect in online auctions; new techniques are needed for the early detection of opportunistic sellers. An opportunistic seller is someone who attempts to negate online auction safeguards and exploit buyers for monetary gain.

Understanding and identifying occurrences of online deception is critical for increasing participation in online auctions and other forms of e-commerce, as victims of fraud will leave the online auction market and potential new customers withhold participation based on fear of becoming a fraud victim. Identifying online deception is

important as deception in any form is the enemy of trust and some degree of trust is required for all business transactions. Opportunistic sellers use deception tactics to create an illusion of trustworthiness to the buyers' detriment.

Reputation systems are used in online communities as normally a member has no prior knowledge or experience interacting with another member. Unlike a traditional auction which relies on direct reciprocity as in "I trust you because you were trustworthy with me before." An online auction relies on indirect reciprocity as in "I trust you because you were trustworthy with others before." In both cases past trustworthiness is a prerequisite for future transactions. It is the information about reputation that enables trust by inducing a reciprocal response.

The eBay reputation system is based on feedback provided by buyers and sellers. For each transaction the buyers and sellers can opt to appraise the other party by leaving feedback. Feedback consists of a positive, negative or neutral rating with an optional short comment. Feedback score is a number used to measure a member's reputation on eBay. A high feedback score means that a member has received a high number of positive ratings from other members. Every member of eBay has a feedback score. Prior research studies and third-party reports of fraud show rates substantially higher than eBay reputation system's reported negative feedback rate of less than 1%. The conclusion was most buyers were withholding reports of negative feedback in fear of retaliation from the seller.

Nikitov and Stone (2006) found that buyers sometimes embedded negative comments in positive feedback as a means of avoiding retaliation from sellers and damage to their reputation. The researchers surmised that these "negative-positive" feedback postings

contained hidden signals to the buyer community about a problematic or fraudulent seller. The composition of negative-positive feedback included both positive and negative aspects of a transaction. Negative-positive complaints were usually in the formats of "I was pleased with X, but unhappy about Y for the transaction" or "I was unhappy about Y, but was pleased with X for the transaction." Typical examples are "Good product, but slow shipping" and "Took 7 days and 2 messages before replying to my email, but product was well packaged."

The concept of using negative-positive feedback as a signature to identify potential opportunistic sellers in an online auction population was never explored. This gap provided a narrowly scoped and tightly bounded area for research with a goal of the *early detection of online auction opportunistic sellers through the use of negative-positive feedback.*

The objective of the research was to determine *if the presence of negative-positive type feedback comments by buyers (independent variable) is a predictor that a seller is behaving fraudulently (dependent variable).* A correlational research design was selected as it provided for discovering relationships between variables, measuring the degree and direction of relationships, and from the discovered relationships predictions could be made.

The research study implemented a correlational research design using an automated data collection agent. The research study required the extraction and analyzing of data that met predefined qualifying conditions from the immense eBay data sets. Manually sifting through data sets of this magnitude was not practical due to the time and labor

required to extract the qualified data. Instead customized software in the form of web crawler was used to search, locate, and extract the qualified data from the eBay website.

Detection of negative-positive feedback by buyers required the examination, interpretation, and categorization of each buyer's feedback comment text. As natural language communications are variable in form, subject to contextual use, can be incomplete, and prone to errors in spelling and/or grammar; it was necessary to transpose the relevant written text into a formatted and coded structure. A coded structure provides data uniformity and enables automated analysis.

Contextual analysis is a systematic method for analyzing data in a standardized way. The term textual analysis is used when contextual analysis is applied to written communication. Using textual analysis provided a powerful and effective technique for coding and categorizing the buyer feedback comments. Codes were used to create categorical variables representing the original textual information. The resulting categorical variables were analyzed using standard statistical methods.

The textual communications found in feedback comments were in natural language format with complex overtones and subtle nuances which precluded any easy method for representation in a coded structure. As automated textual analysis software currently have limited capabilities and accuracy, it was necessary to use a human to make the appropriate judgment of whether or not a feedback comment was in negative-positive format.

The objective of the research was to determine if the presence of negative-positive type feedback comments by buyers (independent variable) is a predictor that a seller is behaving fraudulently (dependent variable). Identification of feedback comments by

buyers as negative-positive or not has been addressed. Next sellers needed to be identified as behaving honestly or fraudulently.

There are only two sources with authority that can equitably state an eBay member is a fraudster – eBay and criminal court rulings. Due to confidentiality, eBay will not provide any details to third-parties on complaints against a member or indicate why a member's account was suspended or disabled. Therefore, an explicit confirmation that a specific online auction member was an opportunistic seller from the primary source – eBay - was not available. As criminal court records containing formal prosecutions for online auction fraud are very limited in number, could be sealed preventing public access to the details or take years for a final legal verdict to be reached; they were not used. Therefore secondary sources were used to draw inferences on a seller's behavior as honest or fraudulent.

Human behavior is complex and sometimes inconsistent; attempting to find a single specific behavior pattern to signal fraudulent behavior would not be realistic. Taking a clue from prior research into credit card fraud, online auction fraud detection is based on looking for red flags and behavior patterns (Bhargava, et al., 2003). The mechanical process of going through a long checklist of all the potential red flags and behavior patterns for even a single seller would be time consuming and any lapse by an evaluator would result in a misclassification. An automated means for making the required judgment to classify a seller as behaving honestly or fraudulently was not available. In order to reduce the manual labor required a software application named *Auction Inquisitor* that automatically searched for red flags and suspect behavior patterns in eBay auctions was used. *Auction Inquisitor* as a front end for the evaluation process provided

the following advantages – greatly reduced the time required to review the red flags and suspect behavior patterns for a seller; enabled the review process to be performed consistently and without human error; and presented the results in a summarized and standardized format.

A pilot test was performed using dedicated raters to evaluate the sellers as honest or fraudulent and code buyer feedback comments as negative-positive type feedback or not. Based on the results of the pilot test, the time required to process the data and estimated labor costs were not feasible. An alternative method of crowdsourcing was tested. It proved feasible in terms of time required to process and estimated costs. The crowdsourcing service was provided using Amazon Mechanical Turk.

When using crowdsourcing, data quality control is a major issue as unseen, remote, and random strangers are being asked to perform your task. First - How do you know that the workers will have the prerequisite skills or knowledge to perform correctly the task? Second - How do you know that the workers will actually make an effort to perform the task rather than just randomly click on responses? These issues were addressed by using qualification tests and data quality techniques of multiple worker assignments per HIT (plurality), minimum work time per HIT, gold standard data, and advice of auditing.

The research study focused on the determining if negative-positive type feedback comments by buyers are a predictor that a seller is behaving fraudulently. Three research questions were used in framing an answer for this primary question. The result for each research question is summarized here.

Research Question 1 (RQ1): Does negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Null Hypothesis (H1o): Negative-positive type feedback comments from buyers do not predict evaluators' consensus of seller fraudulent-type behavior.

Alternative Hypothesis (H1a): Negative-positive type feedback comments from buyers do predict evaluators' consensus of seller fraudulent-type behavior.

The null hypothesis H1o was rejected based on the results from the logistic regression. The Chi-Square test was significance with Chi^2 (1) = 84.40, $p < 0.001$. As the null hypothesis was rejected, the alternative hypothesis H1a was accepted. Sellers were 1.82 times more likely to be fraudulent with the presence of even a single negative-positive feedback type comment.

Research Question 2 (RQ2): Does the number of negative-positive type feedback comments from buyers predict evaluators' consensus of seller fraudulent-type behavior?

Null Hypothesis (H2o): The number of negative-positive type feedback comments does not predict evaluators' consensus of seller fraudulent-type behavior.

Alternative Hypothesis (H2a): The number of negative-positive type feedback comments predicts evaluators' consensus of seller fraudulent-type behavior.

The null hypothesis H2o was rejected based on the results from the logistic regression. The Chi-Square test was significance with Chi^2 (1) = 10.92, $p < 0.001$. As the null hypothesis was rejected, the alternative hypothesis H2a was accepted. The results of the logistic regression indicated that for every one unit increase in the number of negative-positive feedback comments, sellers were 1.04 times (4%) more likely to be fraudulent.

Research Question 3 (RQ3): For each seller will negative-positive type feedback comments from buyers fall into a cluster?

Null Hypothesis (H3o): For each seller will negative-positive feedback comments do not fall into a cluster.

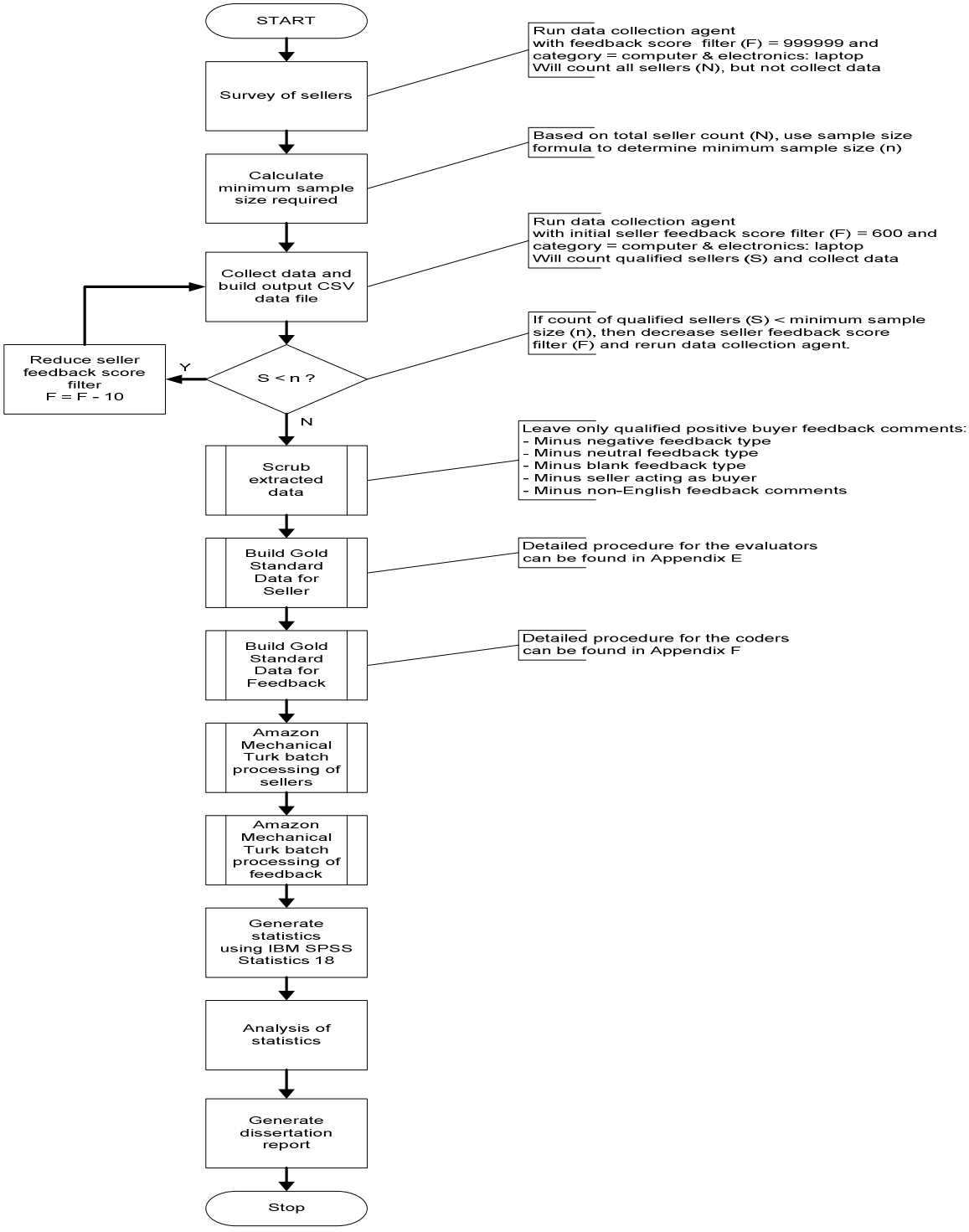
Alternative Hypothesis (H3a): For each seller negative-positive feedback comments fall into a cluster.

The Chi-Square test was significance with Chi^2 (1) = 426.18, $p < 0.001$. No clustering of negative-positive type feedback was revealed for 694 sellers and 109 sellers did reveal clustering of negative-positive type comments The expected count for each cell was 401.5 [(694+109)/2)] indicating that fewer sellers than expected had negative-positive type comments that was clustered. The results suggested that negative-positive type feedback comments did not fall into a cluster, therefore the null hypothesis H3o was accepted.

The research study focused on the determining if negative-positive type feedback comments by buyers were a predictor that a seller was behaving fraudulently. The research had a good outcome as an exploratory study. It confirmed the variable – *negative-positive type feedback comment* - as a new indicator for identifying potentially fraudulent sellers on eBay. Multiple occurrences of negative-positive type feedback comments by buyers increased the probability that a seller was behaving fraudulently.

Appendix A

Methodology Overview



Appendix B

CSV Data File Schema

For every feedback entry found in a qualified seller's transaction history, a record will be created. Each record will have the following format:

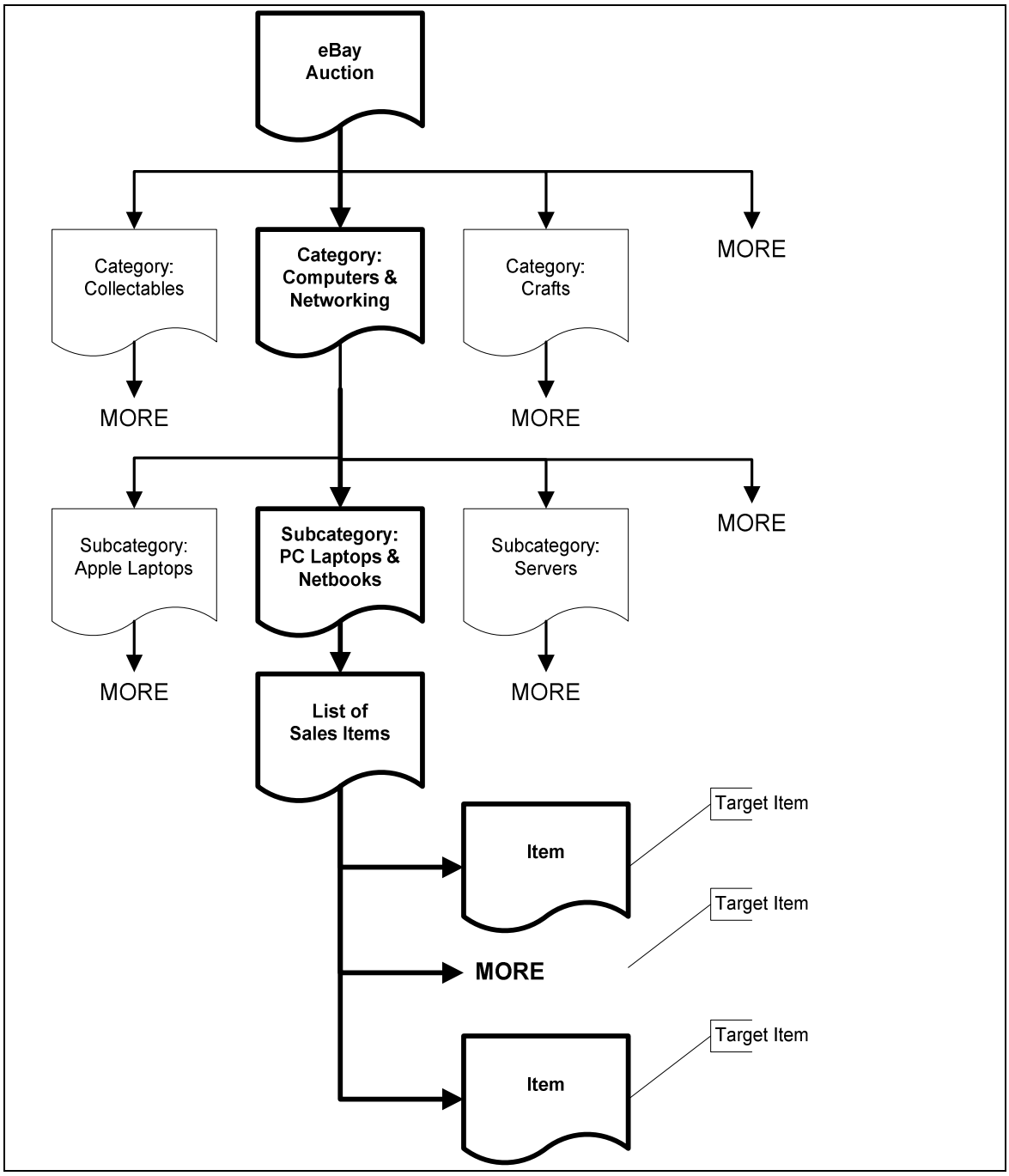
Data File Schema		
Field Name	**Field Format**	**Rules**
Feedback Number	6 digits	Autonumber (unique)
Item Number	15 digits	Extracted from Buyer Item Purchased
Seller eBay Userid	30 characters	Can not be blank.
Seller Feedback Score	6 digits	0 to 999999
Seller Positive Feedback Percent	5 digits	000.00 to 100.00
Seller Member Since	20 characters	Can not be blank.
Seller Status	25 characters	Can not be blank.
Buyer Feedback Type	8 characters	NEGATIVE POSITIVE NEUTRAL
Buyer Feedback Comment	80 characters	Optional Could be blank
Buyer eBay Userid	30 characters	Can not be blank.
Buyer Feedback Date	15 characters	Can not be blank.
Buyer Item Purchased	100 characters	Can not be blank. Includes item number.
Buyer Item Cost	15 characters	Can not be blank.
Seller Reply Info	80 characters	Optional Could be blank.
Seller Reply Text	80 characters	Optional Could be blank.
Buyer Follow-up Info	80 characters	Optional Could be blank.
Buyer Follow-up Text	80 characters	Optional Could be blank.
Record Layout Continues on Next Page		

Record Layout Continued from Prior Page		
↓ Data Analysis ↓		
Evaluator Userid 1	14 characters	Not blank
Fraudulent-Type Behavior 1	1 character	Y – Yes N - No
Evaluator Userid 2	14 characters	Not blank
Fraudulent-Type Behavior 2	1 character	Y – Yes N - No
Evaluator Userid 3	14 characters	Not blank
Fraudulent-Type Behavior 3	1 character	Y – Yes N - No
Evaluator Userid 4	14 characters	Not blank
Fraudulent-Type Behavior 4	1 character	Y – Yes N - No
Evaluator Userid 5	14 characters	Not blank
Fraudulent-Type Behavior 5	1 character	Y – Yes N - No
Fraudulent-Type Behavior Consensus	1 character	Y – Yes N - No
Coder Userid 1	14 characters	Not blank
Negative-Positive 1	1 character	Y – Yes N – No
Coder Userid 2	14 characters	Not blank
Negative-Positive 2	1 character	Y – Yes N – No
Coder Userid 3	14 characters	Not blank
Negative-Positive 3	1 character	Y – Yes N – No
Coder Userid 4	14 characters	Not blank
Negative-Positive 4	1 character	Y – Yes N – No
Coder Userid 5	14 characters	Not blank
Negative-Positive 5	1 character	Y – Yes N – No
Negative-Positive Consensus	1 character	Y – Yes N – No

The data schema is not normalized. Ergo - Seller eBay Userid, Seller Feedback Score, Seller Positive Feedback Percent, Seller Member Since, and Seller Status fields are duplicated in each record. This has been done to make the evaluation, coding, and statistical processing easier.

Appendix C

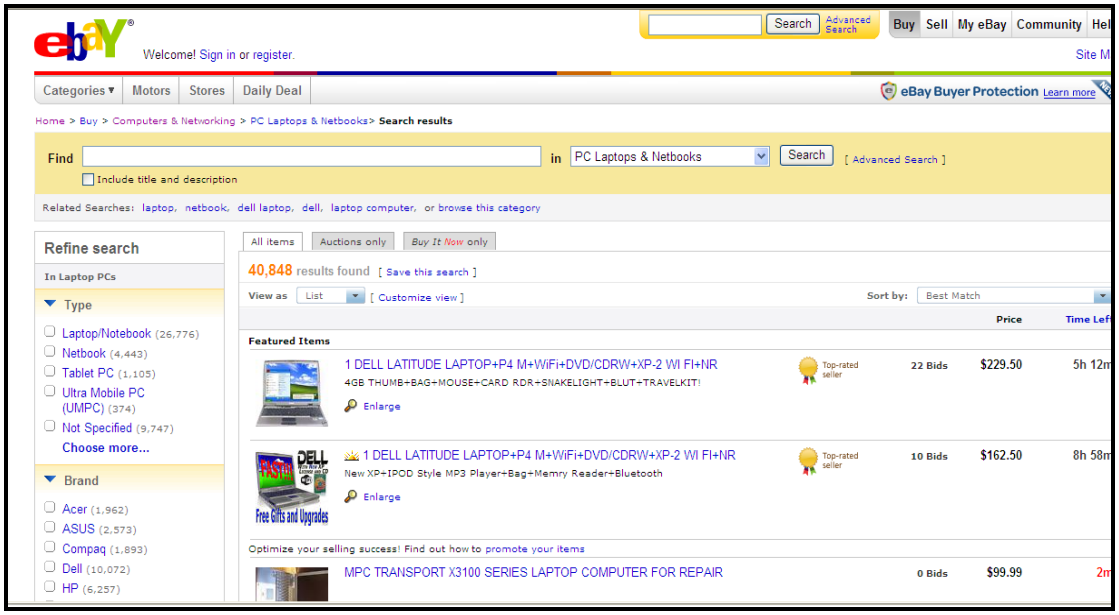
Data Collection Agent



Overview of the catalog structure for organizing eBay sales items on which the crawler will need to transverse.

Overview of the Crawling Mechanics

1. Start up the eBay crawling program.
2. When prompted with:
   Minimum feedback score to qualify seller?
3. Type in 600 and press the ENTER key
   NOTE: The program will maintain and eventually print out two numbers.
   - Total number of sellers found (crawled).
   - Total number of qualified sellers (>= minimum feedback score).
4. Based on the structure for organizing eBay sales items, it will be necessary to manually provide the starting point. Use the following address:
   http://computers.shop.ebay.com/PC-Laptops-Netbooks-/
5. The crawler will now be at the main list of sales items.



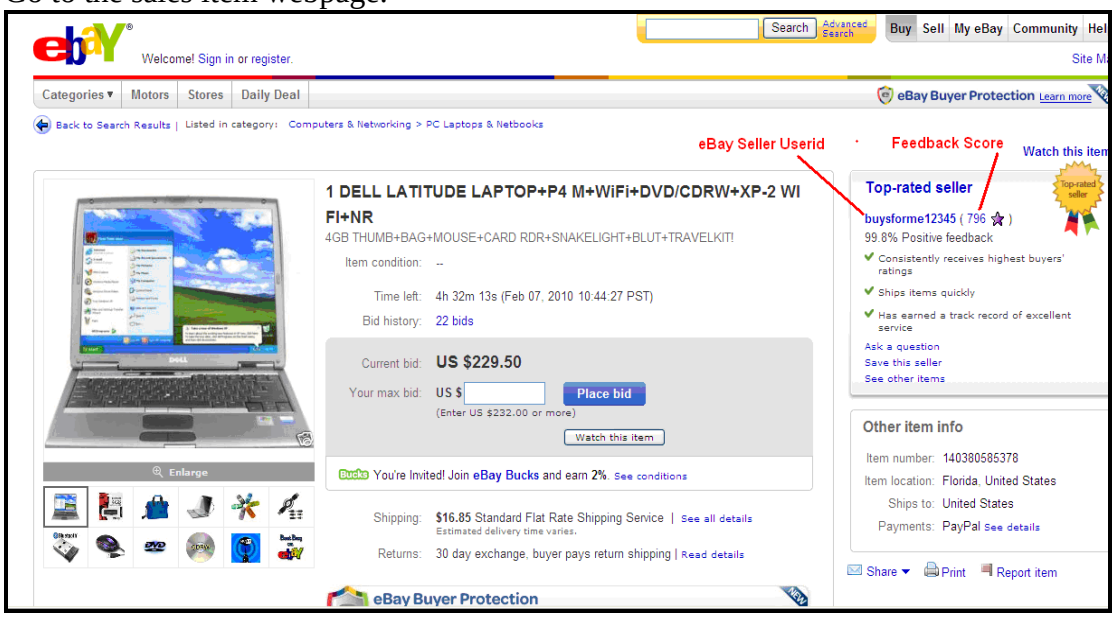
6. **Begin loop to process all sales items in the specified sales item list.**
   Find the next (unprocessed) sales item on the webpage.
   WARNING: We have two levels of complexity for this.
   Level 1 - a page can contain 1 or more sales items.
   Level 2 - there can be more than one page
7. If End-of-List then:
   Close the CVS ASCII data file.
   Display a message:
   ##### total number of sellers crawled
   ##### total number of qualified sellers
   Have a CLOSE button to close the crawler dialog window.
   Terminate the program.

   OR

   Next sales item was found – continue to next step.

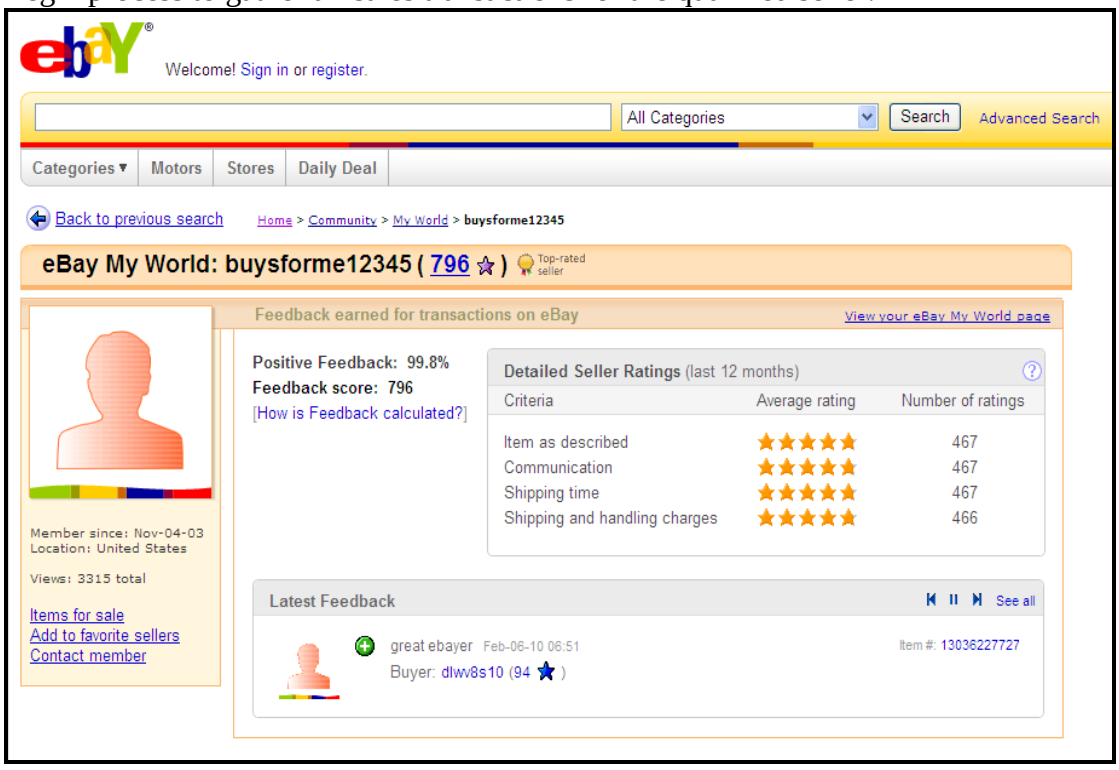
   COMMENT: Basically keep looping till all sales items are processed.

8. Go to the sales item webpage.

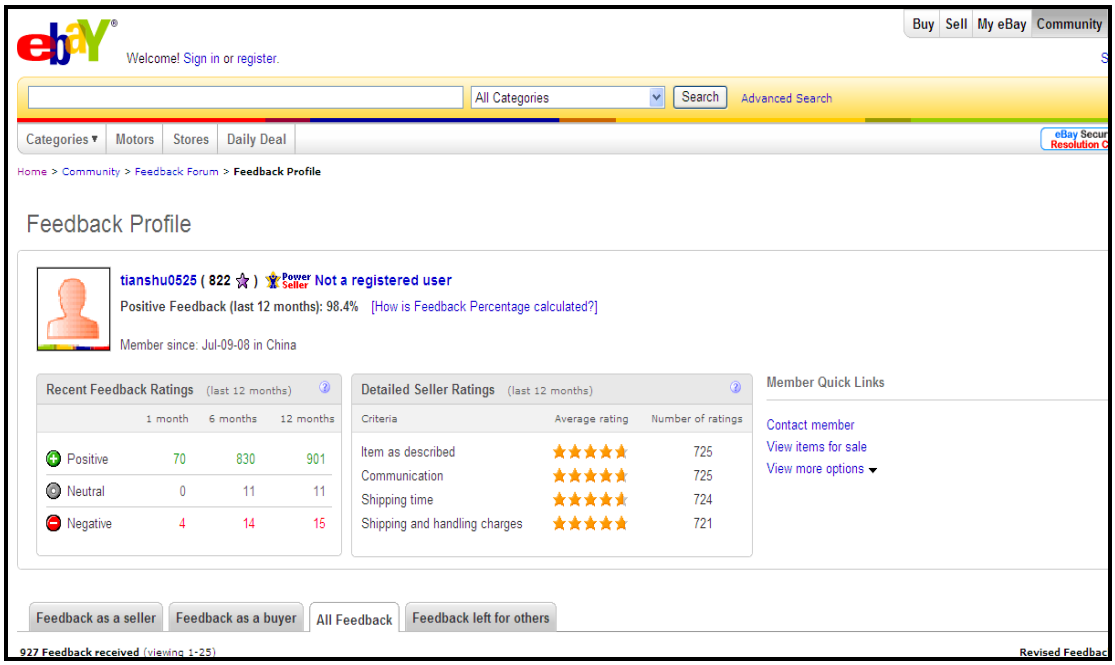


9. Locate **eBay Seller Userid**.
10. Check **Sellers Crawled List** – has this seller already been crawled?
    If **YES**, then do not continue – return to step 6.
    If **NO**, then continue to next step.
11. Increment **Total Number of Sellers Found** by 1.
12. Locate **Feedback Score** for seller.
13. Is the seller's feedback score >= Minimum feedback score?
    If **NO**, then do not continue – return to step 6.
    If **YES**, then continue to next step
14. Increment **Total Number of Qualified Sellers** by 1.
15. Add **eBay Seller Userid** to the **Sellers Crawled List**.
16. Click hyperlink for **eBay Seller Userid**

17. Begin process to gather all sales transactions for the qualified seller.

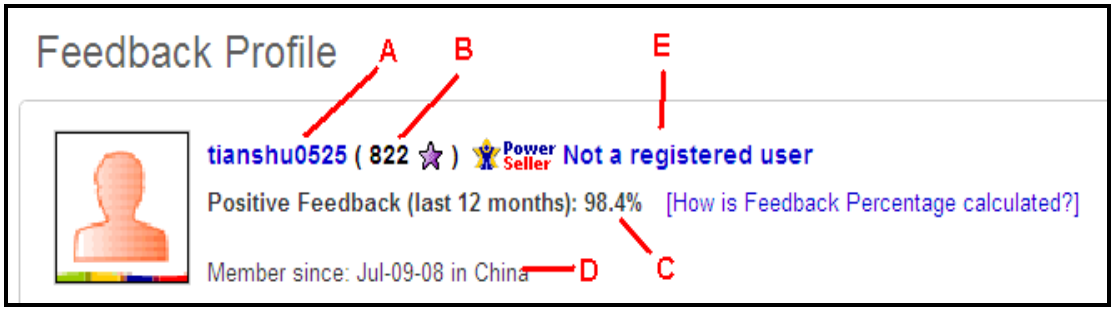


18. Locate and click on the hyperlink named *See All*
19. The crawler is now looking at Feedback Profile webpage for the specified eBay Userid.

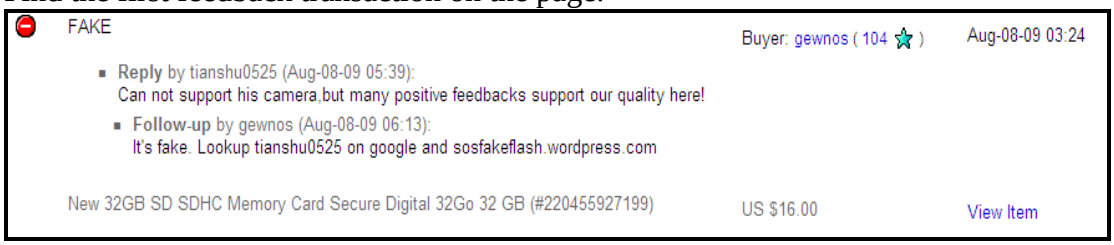


NOTE: To see the remainder of the webpage you would need to scroll down.
Now that crawler is at the appropriate screen, from this point on the program needs to scrap/extract all the required data.

20. First collect the eBay member's general data.
    This data will appear on every feedback record for the seller in the CSV ASCII data file.
    A Seller eBay Userid
    B Seller Feedback Score
    C Seller Positive Feedback Percent
    D Seller Member Since
    E Seller Status

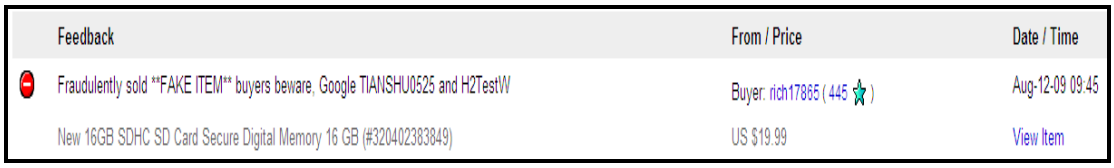


21. **Begin loop to scrap/extract all feedback records for the seller.**
    Find the first feedback transaction on the page.



22. Extract the data for the feedback transaction:
    A Buyer Feedback Type
    B Buyer Feedback Comment
    C Buyer eBay Userid
    D Buyer Feedback Date
    E Buyer Item Purchased
    F Buyer Item Cost
    G Seller Reply Info
    H Seller Reply Text
    I Buyer Follow-up Info
    J Buyer Follow-up Text

NOTE: This is a fully populated feedback with every optional data field being used.



	Feedback	From / Price	Date / Time
⛔	Fraudulently sold **FAKE ITEM** buyers beware, Google TIANSHU0525 and H2TestW	Buyer: rich17865 (445 ★)	Aug-12-09 09:45
	New 16GB SDHC SD Card Secure Digital Memory 16 GB (#320402383849)	US $19.99	View Item

NOTE: This is a typical feedback with only the required data field.

23. Assemble the scraped off data into to a CSV (tab delimitated) text file format.
    See **Appendix C** for detailed information on the record layout.
24. Display a message on the progress:
    ##### total number of sellers crawled
    ##### total number of qualified sellers
25. Attempt to find the NEXT feedback entry.
    WARNING: We have two levels of complexity for this.
    Level 1 - a page can contain 1 or more feedback entries.
    Level 2 - there can be more than one page
26. If End-of-List then:
        Display a message:
        ##### total number of sellers crawled
        ##### total number of qualified sellers
        Go back to step 6
    OR
        Next item was found – Go back to step 21
    COMMENT: Basically keep looping till all feedback entries are processed.

Appendix D

Evaluator Worksheet

Evaluator Worksheet				
Seller eBay Userid:				
Fraudulent-Type Behavior?	**No (Honest)**	☐	**Yes (Fraudulent)**	☐
Attributes of Fraud	**Comments**			
Condition (said new was used)				
Deficit attributes				
Failed to ship				
Incorrect color shipped				
Incorrect product				
Incorrect quantity shipped				
Missing or damaged parts				
Not genuine (copy)				
Poor or badly packaged				
Product not as described				
Shipped late				
Other: _____				
Secondary Sources Found	☐ NO	☐ YES		
Secondary Source Comments				
Other Comments				
Evaluator Userid				

Appendix E

Coding: Identifying Seller Behavior as Honest or Fraudulent

Objective
The objective of this assignment is for you to make a judgment if an eBay seller is exhibiting fraudulent type behavior when selling to eBay buyers.

Legal Disclaimer
Inclusion of an eBay userid in this study does NOT imply that said person (or organization) has in the past exhibited fraudulent type behavior. Nor does it imply that said person (or organization) is currently exhibiting fraudulent type behavior. All the eBay userids included in this research study were selected at random.

Confidentiality
Your judgment will remain confidential to ensure the integrity of the research study. For the research report - all evaluator identifying information will be redacted in order to protect the privacy of participating workers. For the research report - all eBay identifying information will be redacted in order to protect the privacy of the eBay members.

Overview of the Process
In order to make your judgment, you will need to complete the following steps:

- Understand what actions constitute fraudulent type behavior.
- Review the online profile of the eBay userid.
- Search using Google for references of the eBay userid on the Internet.
- Review an analytical report on the eBay userid.
- Using the above data answer the question:
  Is the seller exhibiting fraudulent type behavior?

Details for these steps will be provided below.

Estimated Time to Complete the Assignment
Making an informed judgment is a complex process and takes time.
There is no time limit – Take all the time you want to gather all the data necessary and make your final judgments.

What is Fraudulent Behavior?

For this research study - fraudulent type behavior will be defined as follows:

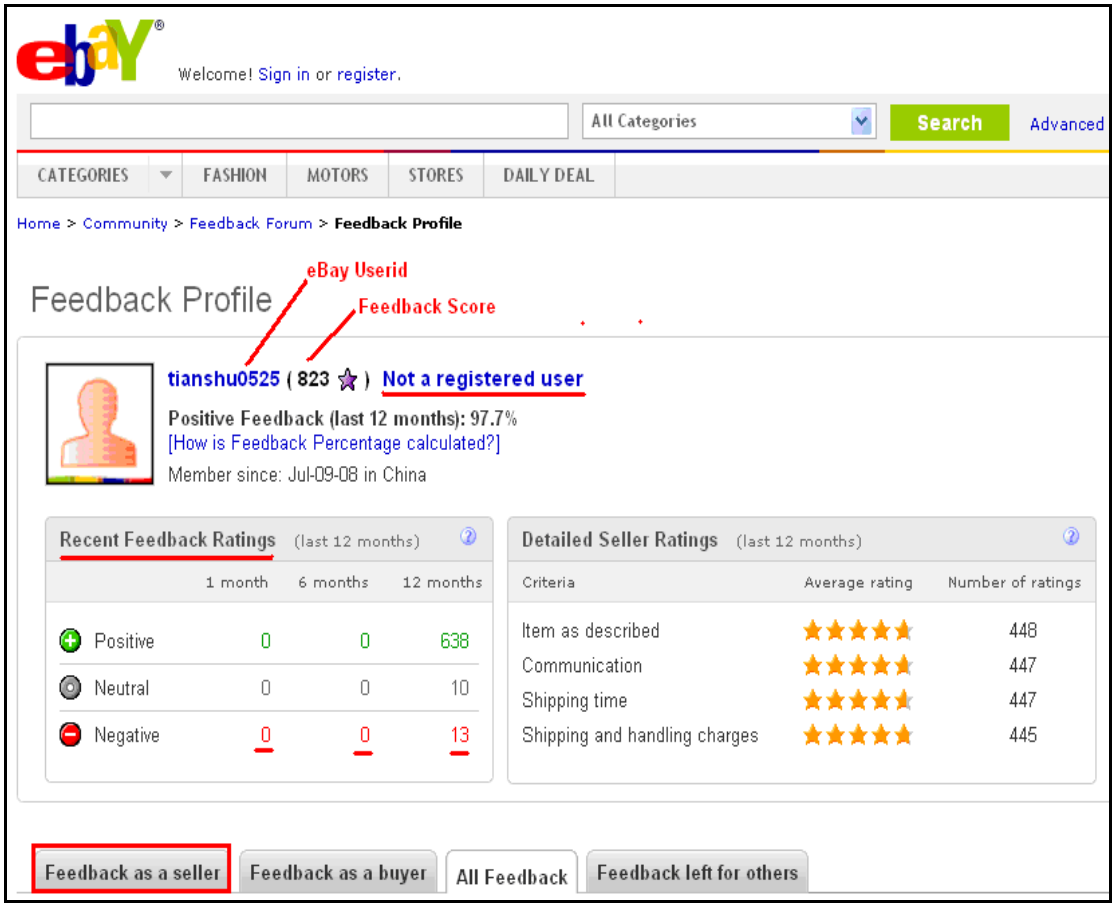
- If the seller ships an item later than agreed upon without reimbursing the buyer for the delay, late shipping constitutes fraudulent type behavior.
- If the product differs from the item's auction description in make, model or condition (i.e. used vs. new), constitutes fraudulent type behavior.
- If the seller does not explicitly state that the item is not genuine (i.e. a copy), constitutes fraudulent type behavior.
- If any deficit attributes of the product are not explicitly stated (i.e. headphones with a six-inch cord rather than the standard three to six foot cord), constitutes fraudulent type behavior.
- If the product is damaged in shipment due to poor packaging, constitutes fraudulent type behavior.
- If the seller collected the buyer's money and failed to ship the item, constitutes fraudulent type behavior.

Instructions

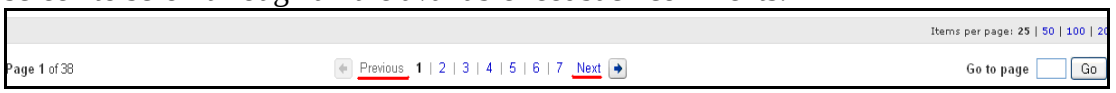
1. Log on the PC with your provided evaluator userid and password.
2. Start up the **Auction Inquisitor** program.
3. Pull a form from the pile of sellers which you are to review.
   NOTE: Forms are prefilled with **Seller's eBay Userid** and **Buyer Item Purchased**.
4. If all the sellers have been reviewed, then logoff the computer and stop evaluating.
5. Open up a web browser using Internet Explorer or Firefox.
6. Go to the following address:
   http://pages.ebay.com/services/forum/feedback-login.html
7. You should now be at the Feedback Forum: Find Member page.

8. In the white box located below **eBay Users ID** type in the *Seller's eBay Userid*
9. Click on the **FIND MEMBER** button
10. You will now be at the **Feedback Profile** for the Seller's eBay Userid.
    Click on the **FEEDBACK AS SELLER** tab located at the bottom of the **Feedback**

Profile.



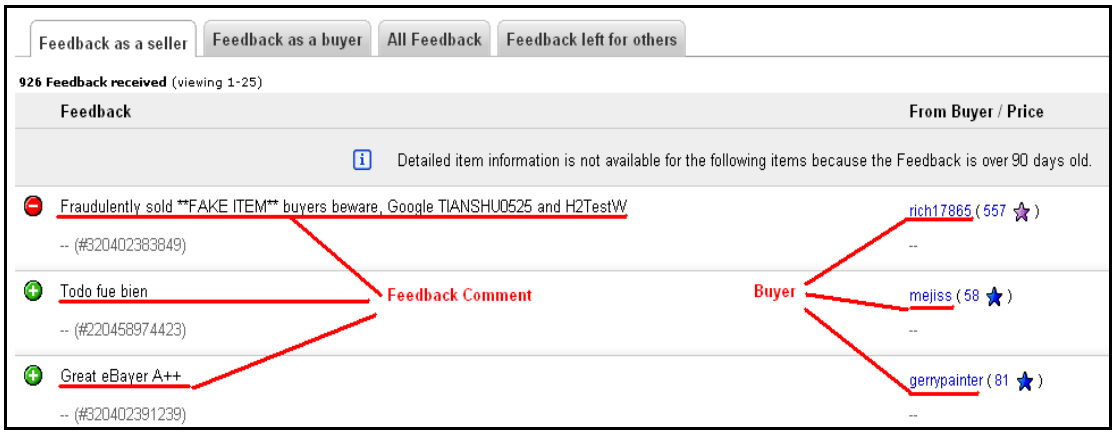
11. You will now only see feedback comments made by buyers that purchased a product from eBay userid Seller's eBay Userid.
    Use the PREVIOUS and NEXT options on the bottom of the **Feedback Profile** screen to scroll through all the available feedback comments.



12. The **What is Fraudulent Behavior?** paragraph at the beginning of this document indicated what actions constitute fraudulent type behavior for this research study.

    The questions below focus these actions to assist you in the review process.
    You have the OPTION of using the *Evaluator Worksheet* to write notes or comments regarding the seller being reviewd.

    You can see in the below picture where the find the **Feedback Comment** and the eBay userid of the **Buyer** who wrote the comment.



Did the seller ship an item later than agreed upon without reimbursing the buyer for the delay?

Did the product differ from the item's auction description in make, model or condition (i.e. used vs. new, wrong color, marked/damaged)?

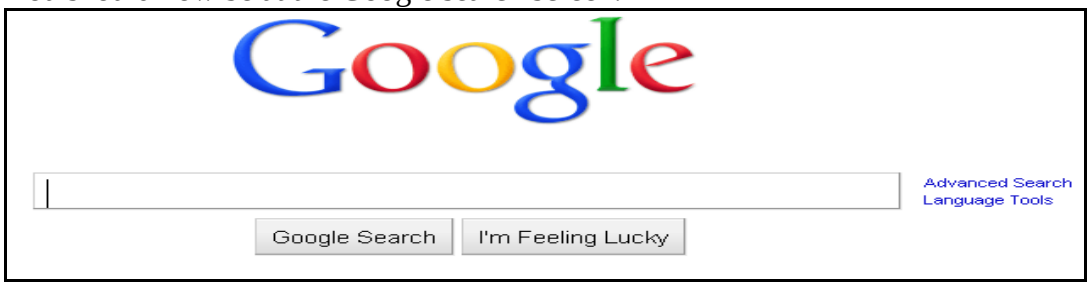
Did the seller not explicitly state that the item was not genuine (i.e. a copy) and shipped a fake or facsimile?

Any deficit attributes of the product that were are not explicitly stated by the seller (i.e. headphones with a six-inch cord rather than the standard three to six foot cord)?

Was the product is damaged in shipment due to poor packaging by the seller?

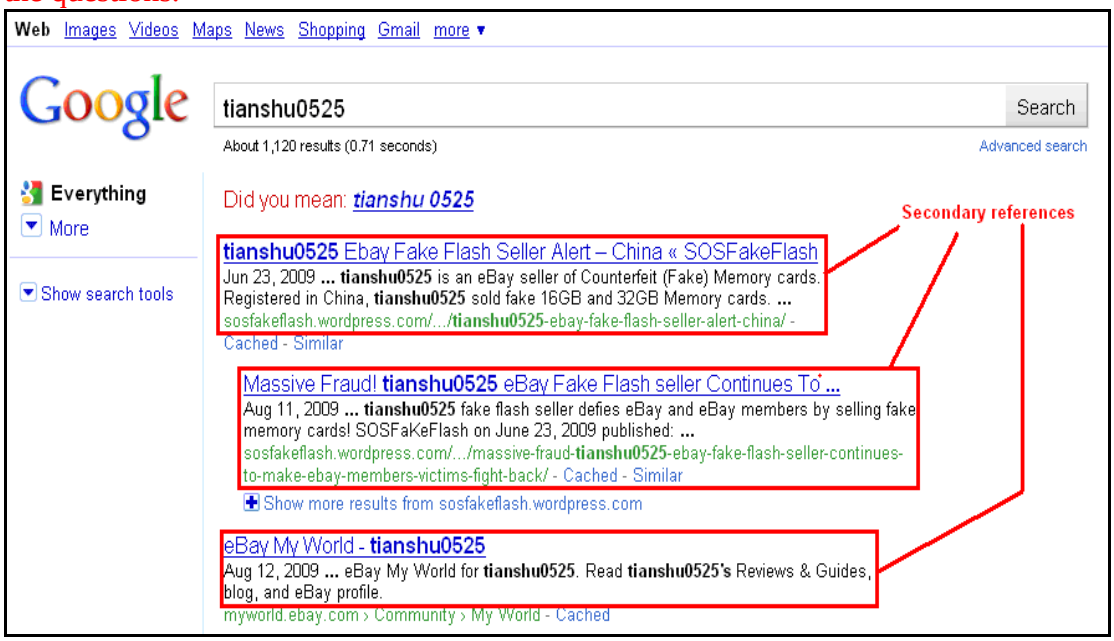
Do you find any other feedback from the buyer that would indicated potentially fraudulent type behavior by the seller?

13. **Secondary Reference**
    Open up another web browser using Internet Explorer or Firefox.
14. Go to the following address:
    http://www.google.com
15. You should now be at the Google search screen:



16. Type in the eBay userid Seller's eBay Userid
17. Click on the **GOOGLE SEARCH** button.
18. You are looking for secondary sources on the Internet that reference the eBay userid. These references (if any) need to be used in making your judgment on whether or not

the eBay userid is exhibiting fraudulent type behavior.
WARNING: The below image is only an example and should not be used to answer the questions.



Did you find one or more secondary references using Google search?

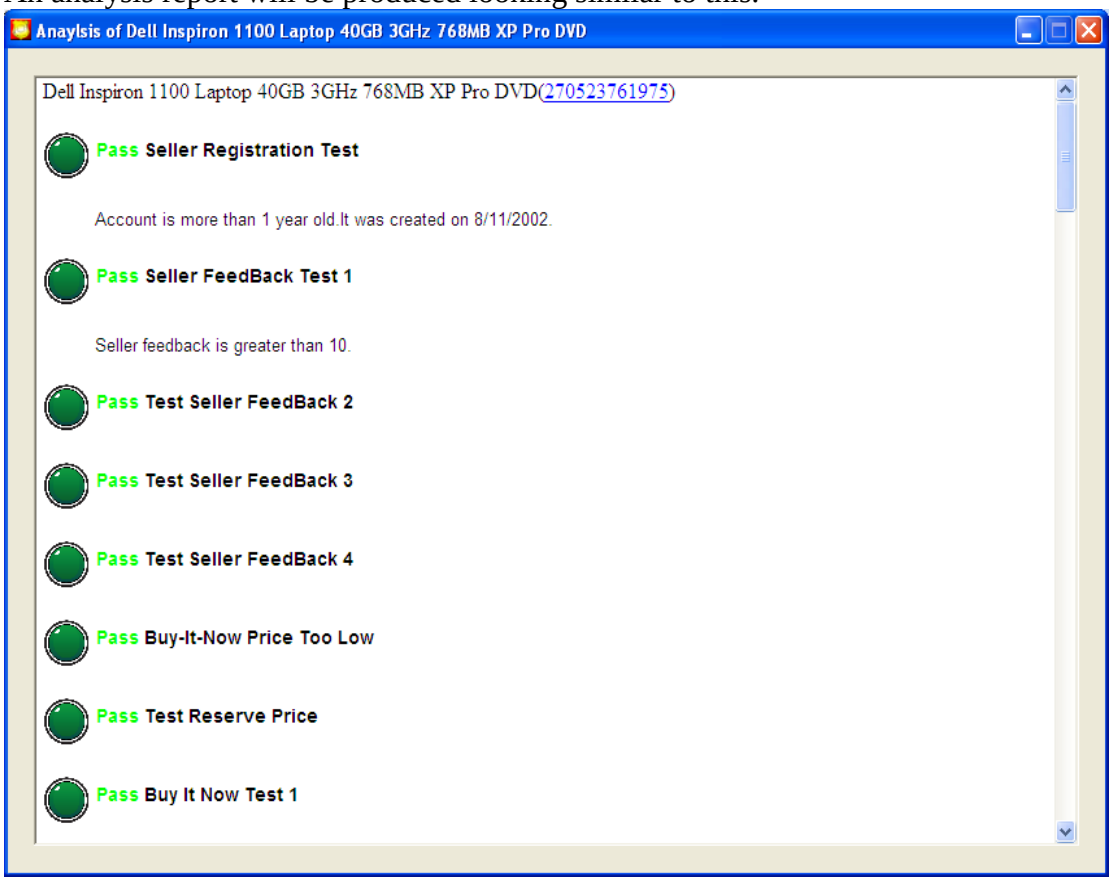
Do one or more of the secondary references found using Google search provide evidence that the eBay userid was exhibiting fraudulent type behavior?

19. **Analytical Report**
    A review of the analytical report must be included in making your judgment on whether or not the eBay seller is exhibiting fraudulent type behavior.
    Next to the **Seller's eBay Userid** in the form, find the **Buyer Item Purchased**.
    The **Item Number** should look similar to the following format (#270523761975).
20. Switch to the **Auction Inquisitor** program.

21. Type in the **Item Number** into the white box above the words:
Enter Auction Number or Auction URL here

22. Click on the **Analyze Auction** button.

23. An analysis report will be produced looking similar to this.



24. Using the **Evaluator Worksheet** as a guide, review the analysis report.

25. **Framing Your Judgment**
    Would you buy on eBay a product from this eBay seller?

    Would you recommend this eBay userid as a seller to a friend?

    Would you recommend this eBay userid as a seller to a family member?

26. **Final Judgment**
    Now you need the answer the final question of:
    Is the eBay seller exhibiting fraudulent type behavior?
    NO - the seller is not acting fraudulently (i.e. honest behavior)
    YES - the seller is acting fraudulently (i.e. fraudulent behavior).
27. Mark your judgment on the form next to **Fraudulent Type Behavior?**
    by placing an **X** in the **NO (HONEST)** or **YES (FRAUDULENT)** check box.
28. Place the completed form in the “done” pile.
29. Close the **Auction Inquisitor** analysis window.
30. Close the web browser window.
31. Repeat the review process on the next seller - Go to step 3.

Appendix F

Coding: Indentifying Buyer Feedback Comment as Negative-Positive

Tutorial

You will be presented with a statement to categorize.
The provided statement was made by a BUYER in response to a purchase from a SELLER.
Your task will be to determine if the provided statement is in NEGATIVE-POSITIVE format or not.

Key Concept

A statement in NEGATIVE-POSITIVE format contains a MINIMUM of one negative declaration AND one positive declaration.

Constructs

Details of a simple statement's construct:
I was pleased with X, but unhappy about Y for the transaction.
Positive declaration => I was pleased with X
Negative declaration => unhappy about Y

Negative-positive statements are usually in a simple format like:
"I was happy about X, but unhappy about Y for the transaction."
"I was unhappy about X, but was pleased with Y for the transaction."

Examples (Positive then negative):
Good product, but slow shipment.
Great quality, but poor packaging.

Examples (Negative then positive):
Not exactly what I expected, but well packaged.
Slow delivery, but great quality.

Alternative complex NEGATIVE-POSITIVE formats use conjunctions [and, but], prepositions [with], multiple sentences or in combination.

Examples (complex formats):
Good product and slow shipment.
Not exactly what I expected and well packaged.
Good product with slow shipment.
Poor service and good quality.
Not exactly what I expected. Well packaged.

Examples (Complex formats with multiple negative and/or positives):
Good product. Well packaged. Slow shipment.

Good product. Well packaged, but slow shipment.
Good product. Well packaged with slow shipment.
Good product. Poorly packaged. Fast shipment.
Took 7 days to reply to my email. Slow shipment, but well packaged.

Examples that are NOT in negative-positive format:
**** (non-informational)
The weather today was beautiful. (not relevant)
Great seller! (one positive)
Shipped the wrong color! (one negative)
Good product and good shipment. (two positives)
Good product. Good packaging. (two positives)
Took 7 days to reply to my email and poorly packaged. (two negatives)

Example Question #1
Is the following statement in negative-positive format?
(123456) Good packing, but slow delivery.
O NO
O YES
Answer: YES - the statement is in negative-positive format.
Meets the MINIMUM of one positive declarative AND one negative declarative.

Example Question #2
Is the following statement in negative-positive format?
(123457) Good product. Fast shipping. Securely packaged.
O NO
O YES
Answer: NO - the statement is NOT in negative-positive format.
Has 3 positive declarative AND 0 (zero) negative declarative.
Does NOT meet the MINIMUM of one positive declarative AND one negative declarative.

Additional Notes
- The provided statements were made by a BUYER in response to a purchase from a SELLER.
- The provided statements have NOT been edited.
- Natural language communications are variable in form, subject to contextual use, can be incomplete, missing punctuation, can have errors in spelling, and/or can have errors in grammar.
- Your task is first to interpret the provided statement as best as possible.
- Next you are to render your best judgment on whether or not the provided statement is in negative-positive format.

Abbreviated Example of a Coder Worksheet

Coder Worksheet		
Coder Userid	**123**	
Feedback Number	**Statement**	**Answer**
123450	Slow delivery, but great quality.	
123451	Not exactly what I expected, but well packaged.	
123452	Shipped the wrong color!	

Instructions to Coder

1. You will be given a Coder Worksheet.
   Verify that your **Coder Userid** matches that found on the worksheet.
   In the above example – 123 is the Coder Userid.

2. Each line in the worksheet contains a **statement** that you will need to evaluate.
   Find the first line in the worksheet that has NOT been evaluated.
   If all lines have been evaluated, then STOP evaluating.

3. Read the **statement** in the line.
   In the above example – the first line's **statement** is:
   Slow delivery, but great quality.

4. Is the **statement** in *negative-positive* format?
   If **YES**, then write **Y** under **Answer** and go to step 2.
   If **NO**, then write **N** under **Answer** and go to step 2.

   NOTE
   You must provide a Y or N for the Answer in every line.
   Do not leave any Answer blank.

Appendix G

Research Qualifications Seller Test

Background
The results from this HIT assignment will be used in an academic research project.
The emphasis is on **QUALITY** and NOT quantity.
Failure to provide quality answers will result in rejection of payment for all work done.

Quality will be insured using these two (and other) techniques:
1. A gold standard statement [previously evaluated/answered question] has been randomly inserted into each HIT assignment.
   Failure to correctly answer the gold standard statement is grounds for rejection of payment.
2. Multiple workers (3-5) will be used to evaluate each HIT assignment.
   Multiple mismatches with other workers' answers for a given HIT assignment are grounds for rejection of payment

Objective
The objective of this assignment is for you to make a judgment if an eBay seller is exhibiting fraudulent type behavior when selling to eBay buyers.

Legal Disclaimer
Inclusion of an eBay userid in this study does NOT imply that said person (or organization) has in the past exhibited fraudulent type behavior. Nor does it imply that said person (or organization) is currently exhibiting fraudulent type behavior. All the eBay userids included in this research study were selected at random.

Confidentiality
Your judgment will remain confidential to ensure the integrity of the research study. For the research report - all Amazon Mechanical Turk identifying information will be redacted in order to protect the privacy of participating workers. For the research report - all eBay identifying information will be redacted in order to protect the privacy of the eBay members.

Overview of the Process
In order to make your judgment, you will need to complete the following steps:
- Understand what actions constitute fraudulent type behavior.
- Review the online profile of the eBay userid.
- Search using Google for references of the eBay userid on the Internet.
- Review an analytical report on the eBay userid.
- Use the above data to answer the question - Is the seller exhibiting fraudulent type behavior?

Details for these steps will be provided below.

Estimated Time to Complete the Assignment
Making an informed judgment is a complex process and takes time.
Depending on the quantity of data to review for an eBay userid, time required will vary from 20-40 minutes to gather all the data and make your final judgment.
Maximum time allocated to complete this test is 60 minutes – after which AMT will automatically issue a "fail" grade.

What is Fraudulent Type Behavior?
For this research study - fraudulent type behavior will be defined as follows:

- **If the seller ships an item later than agreed upon without reimbursing the buyer for the delay, late shipping constitutes fraudulent type behavior.**
- **If the product differs from the item's auction description in make, model or condition (i.e. used vs. new), constitutes fraudulent type behavior.**
- **If the seller does not explicitly state that the item is not genuine (i.e. a copy), constitutes fraudulent type behavior.**
- **If any deficit attributes of the product are not explicitly stated (i.e. headphones with a six-inch cord rather than the standard three to six foot cord), constitutes fraudulent type behavior.**
- **If the product is damaged in shipment due to poor packaging, constitutes fraudulent type behavior.**
- **If the seller collected the buyer's money and failed to ship the item, constitutes fraudulent type behavior.**

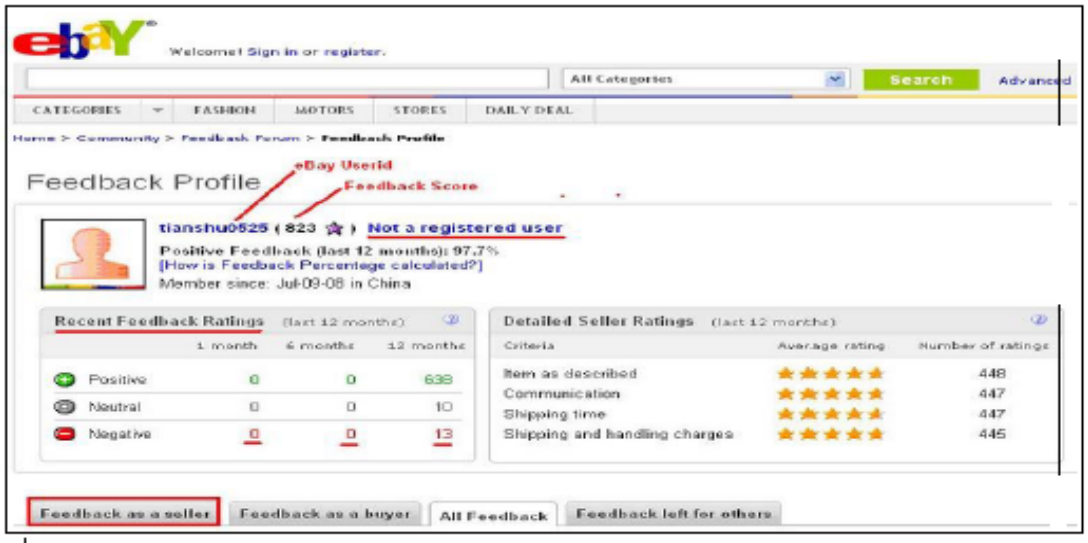
Seller's eBay Userid to Review
SMILENTANGO

☐ I have read and understand the rules

Instructions

1. Open up a web browser using Internet Explorer or Firefox.
2. Go to the following address:
   http://pages.ebay.com/services/forum/feedback-login.html
3. You should now be at the **Feedback Forum: Find Member** page

4. In the white box located below **eBay Users ID** type in SMILENTANGO
5. Click on the **FIND MEMBER** button
6. You will now be at the **Feedback Profile** for the eBay userid SMILENTANGO
WARNING: The below image is only an example and should not be used to answer the questions.



7. **You will need to look at the Feedback Profile for the eBay userid SMILENTANGO in your web browser in order to answer the following group of questions.**

Question # 1

Are the words **"Not a registered user"** displayed in the Feedback Profile?

- ○ No
- ○ Yes

Question # 2
Using the numbers under RECENT FEED BACK RATINGS
Does the seller have any negative feedback under 1 month?

- No
- Yes

Question # 3
Using the numbers under RECENT FEED BACK RATINGS
Does the seller have any negative feedback under 6 months?

- No
- Yes

Question # 4
Using the numbers under RECENT FEED BACK RATINGS
Does the seller have any negative feedback under 12 months?

- No
- Yes

---

8. Click on the **FEEDBACK AS SELLER** tab located at the bottom of the **Feedback Profile**.
9. You will now only see feedback comments made by buyers that purchased a product from eBay userid SMILENTANGO.
Use the **PREVIOUS** and **NEXT** options on the bottom of the **Feedback Profile** screen to scroll through all the available feedback comments.

10. **You will need to scroll through the feedback comments in order to answer the following group of questions.**
WARNING: The below image is only an example and should not be used to answer the questions.

Question # 5
Did the seller ship an item later than agreed upon without reimbursing the buyer for the delay?

- No
- Yes

Question # 5
Did the seller ship an item later than agreed upon without reimbursing the buyer for the delay?

- ○ No
- ○ Yes

Question # 6
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

Buyer

Question # 7
Did the product differ from the item's auction description in make, model or condition (i.e. used vs. new, wrong color, marked/damaged)?

- ○ No
- ○ Yes

Question # 8
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feed back

Buyer

Question # 9
Did the seller not explicitly state that the item was not genuine (i.e. a copy) and shipped a fake or facsimile?

- ○ No
- ○ Yes

Question # 10
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feed back

Buyer

Question # 11
Any deficit attributes of the product that were are not explicitly stated by the seller (i.e. headphones with a six-inch cord rather than the standard three to six foot cord)?

- No
- Yes

Question # 12
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

Buyer

Question # 13
Was the product is damaged in shipment due to poor packaging by the seller?

- No
- Yes

Question # 14
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

Buyer

Question # 15
Did the seller collect the buyer's money and fail/refuse to ship the item?

- No
- Yes

Question # 16
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feed back

Buyer

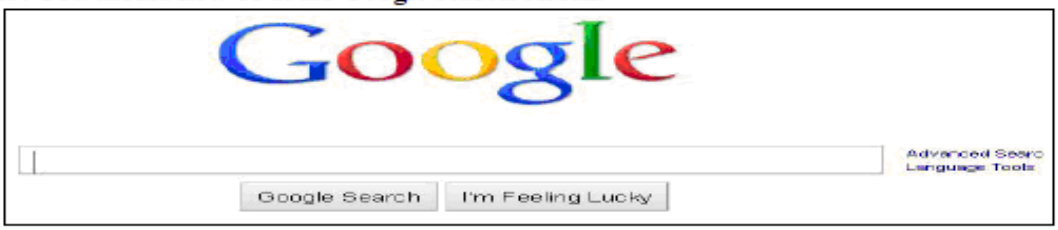
Question # 17
Do you find any other feedback from the buyer that would indicated potentially fraudulent type behavior by the seller?

- No
- Yes

Question # 18
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

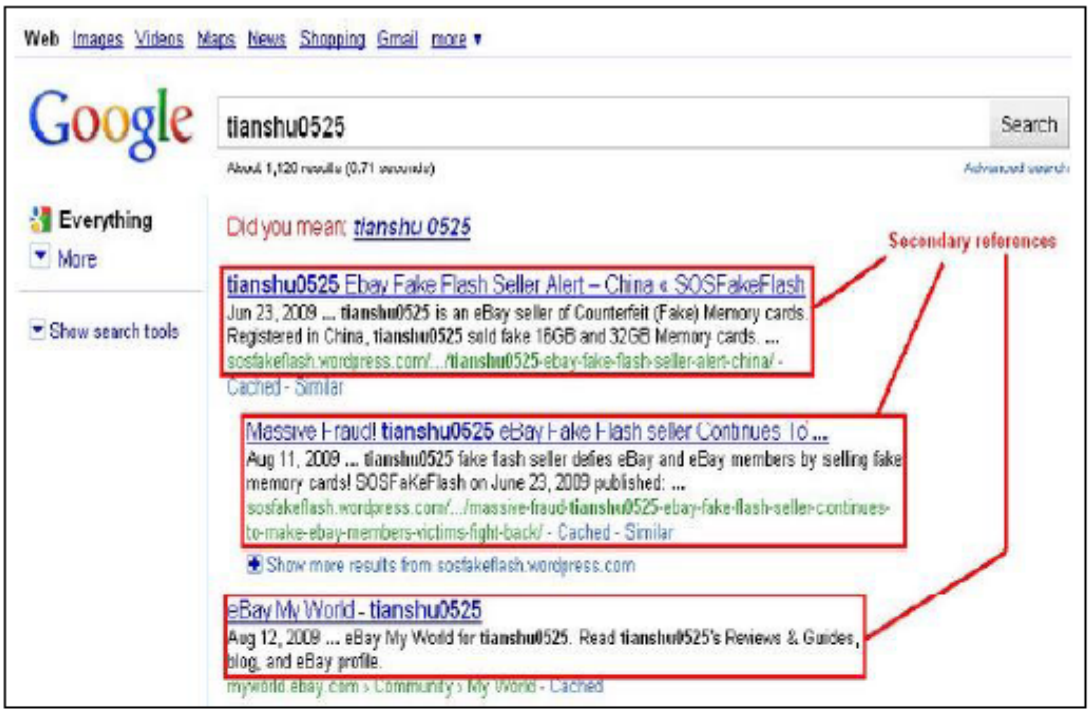
Buyer

11. Open up another web browser using Internet Explorer or Firefox.
12. Go to the following address:
    http://www.google.com
13. You should now be at the **Google search** screen:



14. Type in the eBay userid SMILENTANGO
15. Click on the **GOOGLE SEARCH** button.
16. You are looking for secondary sources on the Internet that reference the eBay userid. These references (if any) need to be used in making your judgment on whether or not the eBay userid is exhibiting fraudulent type behavior.

WARNING: The below image is only an example and should not be used to answer the questions.



17. **You will need to look at your web browser screen of the Google search for the eBay userid in order to answer the following group of questions:**

Question # 19

Did you find one or more secondary references using Google search?

- ○ No
- ○ Yes

Question # 20

Do one or more of the secondary references found using Google search provide evidence that the eBay userid was exhibiting fraudulent type behavior?

- ○ No
- ○ Yes

Question # 21

If Yes, then cut-and-paste one secondary reference to support this.

17. **Please review the below analytical report. It needs to be included in making your judgment on whether or not the eBay userid is exhibiting fraudulent type behavior.**

512MB PC133 SDRAM SODIMM 133MHz
LAPTOP NOTEBOOK MEM RAM(310176545249)

Pass **Seller Registration Test**

Account is more than 1 year old.It was created on 5/20/2004.

Pass **Seller FeedBack Test 1**

Seller feedback is greater than 10.

Pass **Test Seller FeedBack 2**

Pass **Test Seller FeedBack 3**

Pass **Test Seller FeedBack 4**

Pass **Buy-It-Now Price Too Low**

... removed for brevity ...

Pass **Shipping Check**

Verify that the shipping costs are reasonable and that a true value for shipping has been posted. Never purchase from a seller who does not disclose shipping costs until after auction close. That is a common scam to increase the price.

Fail **Shipping :** Sellers shipping details

Seller Has not included shipping Details. Sellers should always include shipping costs in the auction. Beware of sellers who claim they will reveal shipping after the auction. If the auction price is low, they may inflate shipping costs excessively. Honest sellers always put the dollar amount of shipping in the auction.

Payment Details

TEST SUMMARY

Passed 10 Tests
Unclear 4Tests
FAILED 6 Tests

feedback. If a buyer does not pay, then negative feedback should be expected, but sellers should never threaten negative feedback in the auction itself. Dont do business with people who threaten you before you have even placed a bid.

Question # 22
Does anything in the analytical report provide evidence that the eBay userid was exhibiting fraudulent type behavior?

- ○ No
- ○ Yes

Question # 23
If Yes, then cut-and-paste one text item from the analytical report to support this.

Framing Your Judgment

Question # 24
Would you buy on eBay a product from this eBay userid?

- ○ No
- ○ Yes

Question # 25
Would you recommend this eBay userid as a seller to a friend?

- ○ No
- ○ Yes

Question # 26
Would you recommend this eBay userid as a seller to a family member?

- ○ No
- ○ Yes

Question # 27
FINAL JUDGMENT
The objective of this assignment is for you to make a judgment if the eBay seller SMILENTANGO exhibited fraudulent type behavior when selling to eBay buyers.

Before making your judgment:
*** Review the paragraph on What is Fraudulent Type Behavior?**
*** Review the data you collected above**
*** Review your answers to previous questions.**

Has the seller exhibited fraudulent type behavior?

- ○ No
- ○ Yes

Feedback or Suggestion (OPTIONAL)

You have the option below to provide feedback or make a suggestion to improve this study.

Submit

NOTE
The *Test Method* is manual rather than automatic. Approval (i.e. pass/fail) of the worker's test requires a manual approval by the requester.

Appendix H

Research Prototype Seller HIT

Background
The results from the HIT tasks will be used in an academic research project.
Ergo the emphasis is on **QUALITY** and NOT quantity.
Failure to provide quality answers will result in rejection of payment for all work done.

Quality will be insured using these two (and other) techniques:
1. A gold standard statement [previously evaluated/answered question] has been randomly inserted into each HIT assignment.
   Failure to correctly answer the gold standard statement is grounds for rejection of payment.
2. Multiple workers (3-5) will be used to evaluate each HIT assignment.
   Multiple mismatches with other workers' answers for a given HIT assignment are grounds for rejection of payment.

BONUS PAYMENTS
Sorry, as funds are limited - bonus payments will be limited to a few workers who produced the highest quality.

Objective
The objective of this assignment is for you to make a judgment if an eBay seller is exhibiting fraudulent type behavior when selling to eBay buyers.

Legal Disclaimer
Inclusion of an eBay userid in this study does NOT imply that said person (or organization) has in the past exhibited fraudulent type behavior. Nor does it imply that said person (or organization) is currently exhibiting fraudulent type behavior. All the eBay userids included in this research study were selected at random.

Confidentiality
our judgment will remain confidential to ensure the integrity of the research study. For the research report - all Amazon Mechanical Turk identifying information will be redacted in order to protect the privacy of participating workers. For the research report - all eBay identifying information will be redacted in order to protect the privacy of the eBay members.

Overview of the Process
In order to make your judgment, you will need to complete the following steps:
- Understand what actions constitute fraudulent type behavior.
- Review the online profile of the eBay userid.
- Search using Google for references of the eBay userid on the Internet.
- Review an analytical report on the eBay userid.
- Use the above data to answer the question - Is the seller exhibiting fraudulent type behavior?

Details for these steps will be provided below.

Estimated Time to Complete the Assignment
Making an informed judgment is a complex process and takes time.
Depending on the quantity of data to review for an eBay userid, time required will vary from 20-40 minutes to gather all the data and make your final judgment.
Maximum time allocated to complete this test is 60 minutes – after which AMT will automatically issue a "fail" grade.

What is Fraudulent Type Behavior?
For this research study - fraudulent type behavior will be defined as follows:

- **If the seller ships an item later than agreed upon without reimbursing the buyer for the delay, late shipping constitutes fraudulent type behavior.**
- **If the product differs from the item's auction description in make, model or condition (i.e. used vs. new), constitutes fraudulent type behavior.**
- **If the seller does not explicitly state that the item is not genuine (i.e. a copy), constitutes fraudulent type behavior.**
- **If any deficit attributes of the product are not explicitly stated (i.e. headphones with a six-inch cord rather than the standard three to six foot cord), constitutes fraudulent type behavior.**
- **If the product is damaged in shipment due to poor packaging, constitutes fraudulent type behavior.**
- **If the seller collected the buyer's money and failed to ship the item, constitutes fraudulent type behavior.**

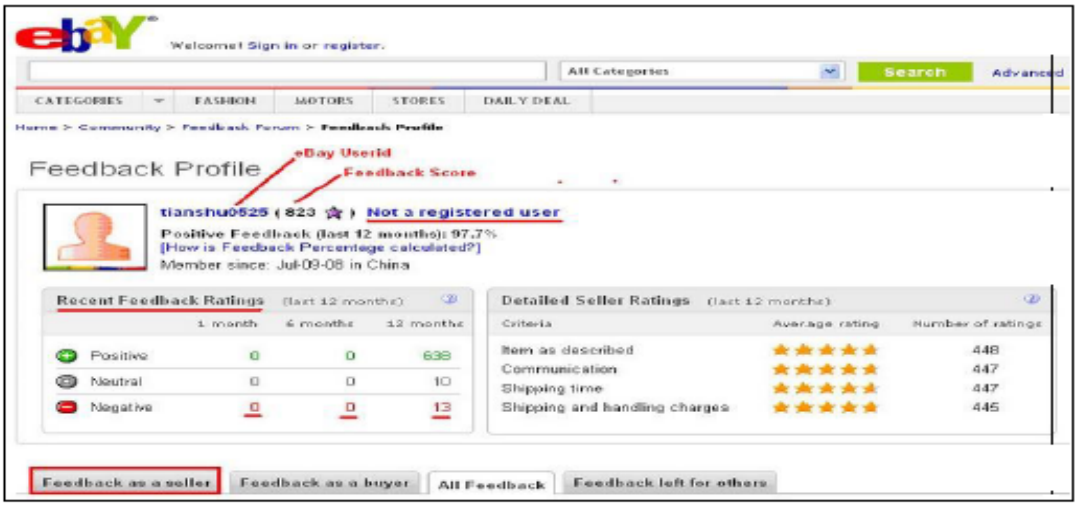
Seller's eBay Userid to Review
SMILENTANGO

☐ I have read and understand the rules

Instructions

1. Open up a web browser using Internet Explorer or Firefox.
2. Go to the following address:
   http://pages.ebay.com/services/forum/feedback-login.html
3. You should now be at the **Feedback Forum: Find Member** page

4. In the white box located below **eBay Users ID** type in SMILENTANGO
5. Click on the **FIND MEMBER** button
6. You will now be at the **Feedback Profile** for the eBay userid SMILENTANGO
   WARNING: The below image is only an example and should not be used to answer the questions.



7. **You will need to look at the Feedback Profile for the eBay userid SMILENTANGO in your web browser in order to answer the following group of questions.**

Question # 1
Are the words "**Not a registered user**" displayed in the Feedback Profile?

- ○ No
- ○ Yes

Question # 2
Using the numbers under RECENT FEED BACK RATINGS
Does the seller have any negative feedback under 1 month?

- ○ No
- ○ Yes

Question # 3
Using the numbers under RECENT FEED BACK RATINGS
Does the seller have any negative feedback under 6 months?

- ○ No
- ○ Yes

Question # 4
Using the numbers under RECENT FEED BACK RATINGS
Does the seller have any negative feedback under 12 months?

- ○ No
- ○ Yes

---

8. Click on the **FEEDBACK AS SELLER** tab located at the bottom of the **Feedback Profile**.
9. You will now only see feedback comments made by buyers that purchased a product from eBay userid SMILENTANGO.

Use the **PREVIOUS** and **NEXT** options on the bottom of the **Feedback Profile** screen to scroll through all the available feedback comments.

10. **You will need to scroll through the feedback comments in order to answer the following group of questions.**

WARNING: The below image is only an example and should not be used to answer the questions.

Question # 5
Did the seller ship an item later than agreed upon without reimbursing the buyer for the delay?

- ○ No
- ○ Yes

Question # 6
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

Buyer

Question # 7
Did the product differ from the item's auction description in make, model or condition (i.e. used vs. new, wrong color, marked/damaged)?

- ○ No
- ○ Yes

Question # 8
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feed back

Buyer

Question # 9
Did the seller not explicitly state that the item was not genuine (i.e. a copy) and shipped a fake or facsimile?

- ○ No
- ○ Yes

Question # 10
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feed back

Buyer

Question # 11
Any deficit attributes of the product that were are not explicitly stated by the seller (i.e. headphones with a six-inch cord rather than the standard three to six foot cord)?

- ○ No
- ○ Yes

Question # 12
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

Buyer

Question # 13
Was the product is damaged in shipment due to poor packaging by the seller?

- ○ No
- ○ Yes

Question # 14
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

Buyer

Question # 15
Did the seller collect the buyer's money and fail/refuse to ship the item?

- ○ No
- ○ Yes

Question # 16
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feed back

Buyer

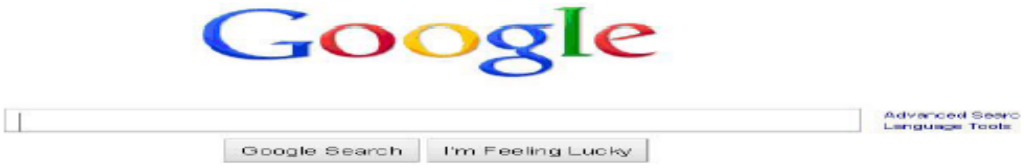
Question # 17
Do you find any other feedback from the buyer that would indicated potentially fraudulent type behavior by the seller?

- ○ No
- ○ Yes

Question # 18
If Yes, then cut-and-paste one feedback comment and buyer to support this.
Feedback

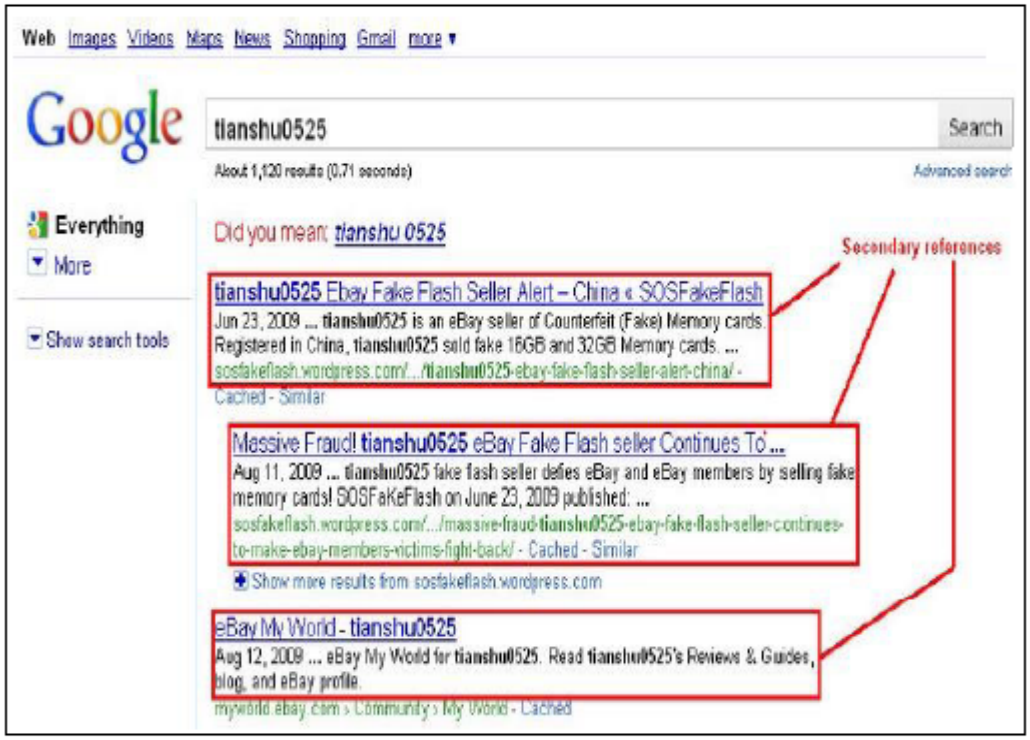
Buyer

11. Open up another web browser using Internet Explorer or Firefox.
12. Go to the following address:
    http://www.google.com
13. You should now be at the **Google search** screen:



14. Type in the eBay userid SMILENTANGO
15. Click on the **GOOGLE SEARCH** button.
16. You are looking for secondary sources on the Internet that reference the eBay userid. These references (if any) need to be used in making your judgment on whether or not the eBay userid is exhibiting fraudulent type behavior.

WARNING: The below image is only an example and should not be used to answer the questions.



17. **You will need to look at your web browser screen of the Google search for the eBay userid in order to answer the following group of questions:**

Question # 19

Did you find one or more secondary references using Google search?

- ○ No
- ○ Yes

Question # 20

Do one or more of the secondary references found using Google search provide evidence that the eBay userid was exhibiting fraudulent type behavior?

- ○ No
- ○ Yes

Question # 21

If Yes, then cut-and-paste one secondary reference to support this.

17. **Please review the below analytical report. It needs to be included in making your judgment on whether or not the eBay userid is exhibiting fraudulent type behavior.**

512MB PC133 SDRAM SODIMM 133MHz
LAPTOP NOTEBOOK MEM RAM(310176545249)

Pass **Seller Registration Test**

Account is more than 1 year old.It was created on 5/20/2004.

Pass **Seller FeedBack Test 1**

Seller feedback is greater than 10.

Pass **Test Seller FeedBack 2**

Pass **Test Seller FeedBack 3**

Pass **Test Seller FeedBack 4**

Pass **Buy-It-Now Price Too Low**

... removed for brevity ...

Pass **Shipping Check**

Verify that the shipping costs are reasonable and that a true value for shipping has been posted. Never purchase from a seller who does not disclose shipping costs until after auction close. That is a common scam to increase the price.

Fail **Shipping :** Sellers shipping details

Seller Has not included shipping Details. Sellers should always include shipping costs in the auction. Beware of sellers who claim they will reveal shipping after the auction. If the auction price is low, they may inflate shipping costs excessively. Honest sellers always put the dollar amount of shipping in the auction.

Payment Details

TEST SUMMARY

Passed **18 Tests**
Unclear **4Tests**
FAILED **6 Tests**

feedback. If a buyer does not pay, then negative feedback should be expected, but sellers should never threaten negative feedback in the auction itself. Dont do business with people who threaten you before you have even placed a bid.

Question # 22
Does anything in the analytical report provide evidence that the eBay userid was exhibiting fraudulent type behavior?

- ○ No
- ○ Yes

Question # 23
If Yes, then cut-and-paste one text item from the analytical report to support this.

Framing Your Judgment

Question # 24
Would you buy on eBay a product from this eBay userid?

- No
- Yes

Question # 25
Would you recommend this eBay userid as a seller to a friend?

- No
- Yes

Question # 26
Would you recommend this eBay userid as a seller to a family member?

- No
- Yes

Question # 27
FINAL JUDGMENT
The objective of this assignment is for you to make a judgment if the eBay seller SMILENTANGO exhibited fraudulent type behavior when selling to eBay buyers.

Before making your judgment:
*** Review the paragraph on What is Fraudulent Type Behavior?**
*** Review the data you collected above**
*** Review your answers to previous questions.**

Has the seller exhibited fraudulent type behavior?

- No
- Yes

Feedback or Suggestion (OPTIONAL)

You have the option below to provide feedback or make a suggestion to improve this study.

Submit

Appendix I

Research Qualifications Feedback Test

BACKGROUND
This test is qualifying you to work on HIT assignments that will be used in an academic research project. The test mimics what you will experience in the HIT assignments.

The emphasis is on **QUALITY** and NOT quantity.
Failure to provide quality answers will result in rejection of payment for all work done

Quality will be insured using these two (and other) techniques:
1. At least one gold standard statement [previously evaluated statement] has been randomly inserted into each HIT.
   Failure to correctly categorize the gold standard statements is grounds for rejection of payment.
2. Multiple workers (5-10) will be used to evaluate each HIT.
   Multiple mismatches with other workers' responses in a given HIT are grounds for rejection of payment.

Estimated Time to Complete the Assignment
Maximum time allocated to complete this test is 60 minutes – after which AMT will automatically issue a "fail" grade.

☐ I have read and understand the rules

INSTRUCTIONS
You will be presented with a statement to categorize.
The provided statement was made by a **BUYER** in response to a purchase from a **SELLER**.
Your task will be to determine if the provided statement is in **NEGATIVE-POSITIVE** format or not.

Key Concept
A statement in NEGATIVE-POSITIVE format contains a MINIMUM of one negative declaration AND one positive declaration.

Details of a simple statement's construct:
I was pleased with X, but unhappy about Y for the transaction.
Positive declaration => I was pleased with X
Negative declaration => unhappy about Y

Negative-positive statements are usually in a simple format like:
"I was happy about X, but unhappy about Y for the transaction."
"I was unhappy about X, but was pleased with Y for the transaction."

Examples (Positive then negative):
Good product, but slow shipment.
Great quality, but poor packaging.

Examples (Negative then positive):
Not exactly what I expected, but well packaged.
Slow delivery, but great quality.

Alternative complex NEGATIVE-POSITIVE formats use conjunctions [and,but], prepositions [with], multiple sentences or in combination.

Examples (complex formats):
Good product and slow shipment.
Not exactly what I expected and well packaged.
Good product with slow shipment.
Poor service and good quality.
Not exactly what I expected. Well packaged.

Examples (Complex formats with multiple negative and/or positives):
Good product. Well packaged. Slow shipment.
Good product. Well packaged, but slow shipment.
Good product. Well packaged with slow shipment.
Good product and fast shipment. However, poorly packaged.
Took 7 days to reply to my email. Slow shipment, but well packaged.

Examples that are NOT in negative-positive format:
**** (non-informational)
The weather today was beautiful. (not relevant)
Great seller! (one positive)
Shipped the wrong color! (one negative)
Good product and good shipment. (two positives)
Good product. Good packaging. (two positives)
Took 7 days to reply to my email and poorly packaged. (two negatives)

Example Question #1
Is the following statement in negative-positive format?
(123456) Good packing, but slow delivery.
O NO
O YES
Answer: YES - the statement is in negative-positive format.
Meets the MINIMUM of one positive declarative AND one negative declarative.

Example Question #2
Is the following statement in negative-positive format?
(123457) Good product. Fast shipping. Securely packaged.
O NO
O YES
Answer: NO - the statement is NOT in negative-positive format.
Has 3 positive declarative AND 0 (zero) negative declarative.
Does NOT meet the MINIMUM of one positive declarative AND one negative declarative.

Additional Notes
- The provided statements were made by a BUYER in response to a purchase from a SELLER.
- The provided statements have NOT been edited.
- Natural language communications are variable in form, subject to contextual use, can be incomplete, missing punctuation, can have errors in spelling, and/or can have errors in grammar.
- Your task is first to interpret the provided statement as best as possible.
- Next you are to render your best judgment on whether or not the provided statement is in negative-positive format.
- At the bottom of each HIT is a text field where you have the OPTION to leave a comment or feedback.

☐ I have read the tutorial

Is the following statement in negative-positive format?

Quick shipment, but wrong color.

○ No
○ Yes

Is the following statement in negative-positive format?

Great product! Would buy from again!

○ No
○ Yes

Is the following statement in negative-positive format?

Slow shipping. Good packaging.

○ No
○ Yes

....The additional 47 questions were deleted for brevity...

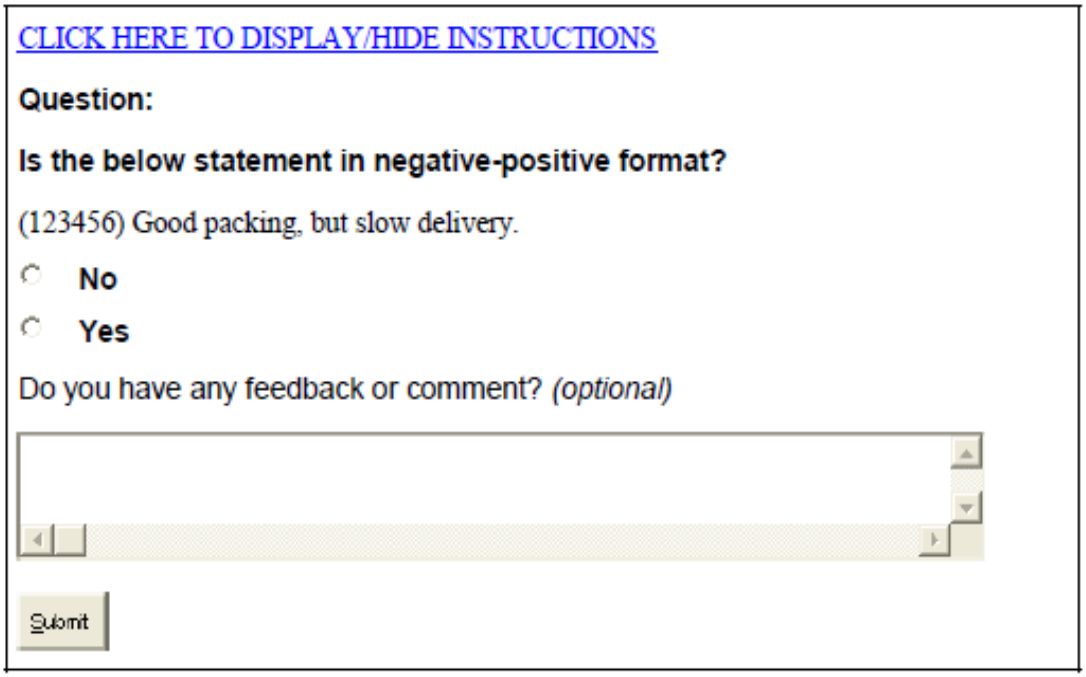
Do you have any feedback or comment? *(optional)*

Submit

Appendix J

Research Prototype Feedback HIT With Instructions Hidden

Template: Research - coder - q1 - v1



Appendix K

Research Prototype Feedback HIT With Instructions Displayed

Template: Research - coder - q1 - v1

INSTRUCTIONS

You will be presented with a statement to categorize.
The provided statement was made by a BUYER in response to a purchase from a SELLER.
Your task will be to determine if the provided statement is in NEGATIVE-POSITIVE format or not.

Key Concept
A statement in NEGATIVE-POSITIVE format contains a MINIMUM of one negative declaration AND one positive declaration.

Details of a simple statement's construct:
I was pleased with X, but unhappy about Y for the transaction.
Positive declaration => I was pleased with X
Negative declaration => unhappy about Y

Negative-positive statements are usually in a simple format like:
"I was happy about X, but unhappy about Y for the transaction."
"I was unhappy about X, but was pleased with Y for the transaction."

Examples (Positive then negative):
Good product, but slow shipment.
Great quality, but poor packaging.

Examples (Negative then positive):
Not exactly what I expected, but well packaged.
Slow delivery, but great quality.

Alternative complex NEGATIVE-POSITIVE formats use conjunctions [and, but], prepositions [with], multiple sentences or in combination.

Examples (complex formats):
Good product and slow shipment.
Not exactly what I expected and well packaged.
Good product with slow shipment.
Poor service and good quality.
Not exactly what I expected. Well packaged.

Examples (Complex formats with multiple negative and/or positives):
Good product. Well packaged. Slow shipment.
Good product. Well packaged, but slow shipment.
Good product. Well packaged with slow shipment.
Good product. Poorly packaged. Fast shipment.
Took 7 days to reply to my email. Slow shipment, but well packaged.

Examples that are NOT in negative-positive format:
**** (non-informational)
The weather today was beautiful. (not relevant)
Great seller! (one positive)
Shipped the wrong color! (one negative)
Good product and good shipment. (two positives)
Good product. Good packaging. (two positives)
Took 7 days to reply to my email and poorly packaged. (two negatives)

Example Question #1
Is the following statement in negative-positive format?
(123456) Good packing, but slow delivery.
O NO
O YES
Answer: YES - the statement is in negative-positive format.
Meets the MINIMUM of one positive declarative AND one negative declarative.

Example Question #2
Is the following statement in negative-positive format?
(123457) Good product. Fast shipping. Securely packaged.
O NO
O YES
Answer: NO - the statement is NOT in negative-positive format.
Has 3 positive declarative AND 0 (zero) negative declarative.
Does NOT meet the MINIMUM of one positive declarative AND one negative declarative.

Additional Notes

- The provided statements were made by a BUYER in response to a purchase from a SELLER.
- The provided statements have NOT been edited.
- Natural language communications are variable in form, subject to contextual use, can be incomplete, missing punctuation, can have errors in spelling, and/or can have errors in grammar.
- Your task is first to interpret the provided statement as best as possible.
- Next you are to render your best judgment on whether or not the provided statement is in negative-positive format.
- At the bottom of each HIT is a text field where you have the OPTION to leave a comment or feedback.

CLICK HERE TO DISPLAY/HIDE INSTRUCTIONS

Question:

Is the below statement in negative-positive format?

(123456) Good packing, but slow delivery.

- **No**
- **Yes**

Do you have any feedback or comment? *(optional)*

Submit

Appendix L

Research Production Feedback HIT

Template: Research - coder - q10 - v1

CLICK HERE TO DISPLAY/HIDE INSTRUCTIONS

Question # 1:

Is the below statement in negative-positive format?

(123457) Good product. Fast shipping. Securely packaged.

- **No**
- **Yes**

Question # 2:

Is the below statement in negative-positive format?

(439128) Great seller!

- **No**
- **Yes**

Question # 3:

Is the below statement in negative-positive format?

(019813) Shipped the wrong color!

- **No**
- **Yes**

Question # 4:

Is the below statement in negative-positive format?

(398761) Good product and good shipment.

- **No**
- **Yes**

Question # 5:

Is the below statement in negative-positive format?

(104521) Good product. Good packaging.

- **No**
- **Yes**

Question # 6:

Is the below statement in negative-positive format?

(239501) Took 7 days to reply to my email and poorly packaged

- **No**
- **Yes**

Question # 7:

Is the below statement in negative-positive format?

(449381) Not exactly what I expected, but well packaged.

- **No**
- **Yes**

Question # 8:

Is the below statement in negative-positive format?

(123456) Good packing, but slow delivery.

- **No**
- **Yes**

Question # 9:

Is the below statement in negative-positive format?

(324414) It is raining outside.

- **No**
- **Yes**

Question # 10:

Is the below statement in negative-positive format?

(522414) Slow delivery, but great quality.

- **No**
- **Yes**

Do you have any feedback or comment on the above questions? *(optional)*

Submit

Reference List

2009 Internet crime report. (2010). Washington, DC: Department of Justice.

About feedback. (2010). Retrieved 10/25/10, 2010, from http://pages.ebay.com/help/policies/feedback-ov.html

Advanced research API. (2010). Retrieved 10/25/10, 2010, from http://developer.researchadvanced.com/signup.php

Akerlof, G. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488-500.

Albert, M. R. (2002). E-buyer beware: Why online auction fraud should be regulated. *American Business Law Journal*, *39*(4), 575-643.

Aldrich, J. (1995). Correlations genuine and spurious in pearson and yule. *Statistical Science*(10), 364-376.

Allen, G. N., Burk, D. L., & Davis, G. B. (2006). Academic data collection in electronic environments: Defining acceptable use of Internet resources. *MIS Quarterly*, *30*(3), 599-610.

Almendra, V., & Schwabe, D. (2009). *Analysis of Fraudulent Activity in a Brazilian Auction Site*. Paper presented at the 18th International World Wide Web Conference Committee (IW3C2). Retrieved from http://www2009.eprints.org/203/

Amazon Mechanical Turk. (2010). Retrieved 10/25/10, 2010, from aws.amazon.com/mturk

Amazon mechanical turk requester best practices guide. (2010). Retrieved from http://mturkpublic.s3.amazonaws.com/docs/MTURK_BP.pdf

Amazon Web Services. (2010). Retrieved 10/25/10, 2010, from aws.amazon.com

Ambach, J. (2008). General announcements: A message from Jim Ambach – more details for sellers. Retrieved 10/25/10, 2010, from http://www2.ebay.com/aw/core/200801.shtml#2008-01-29061132

Ba, S., Whinston, A. B., & Zhang, H. (2002). Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems, 907*.

Bajari, P., & Hortacsu, A. (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions. *RAND Journal of Economics, 34*(2).

Bajari, P., & Hortacsu, A. (2004). Economic insight from Internet auctions. *Journal of Economic Literature, 42*(June), 457-486.

Bajari, P., & Hortacsu, A. (2005). Are structural estimates of auction models reasonable? Evidence from experimental data. *Journal of Political Economy, 113*(4), 703-741.

Bakos, Y., & Dellarocas, C. (2003). Cooperation without enforcement? A comparative analysis of litigation and online reputation as quality assurance mechanisms. MIT Sloan.

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

Bapna, R., Goes, P., Gupta, A., & Jin, Y. (2004). User heterogeneity and its impact on electronic auction market design: An empirical exploration. *MIS Quarterly, 28*(1), 21-43.

Barr, J., & Cabrera, L. F. (2006). AI gets a brain. *Queue, 4*(4), 24-29.

Bartlett, J., Kotrlik, J., & Higgins, C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal, 19*(1), 43-50.

Bhargava, B., Zhong, Y., & Lu, Y. (2003). Fraud formalization and detection *Data warehousing and knowledge discovery* (Vol. 2737, pp. 330-339): Springer.

Bolton, G., Greiner, B., & Ockenfels, A. (2009). *Engineering Trust - Reciprocity in the Production of Reputation Information* (No. 2009-02): The University of New South Wales.

Borg, W. L., & Gall, M. D. (2006). *Educational research: An introduction* (8th ed.). White Plains, NY: Longman.

Bowker, N., & Tuffin, K. (2003). Dicing with deception: People with disabilities' strategies for managing safety and identity online. *Journal of Computer-Mediated Communication, 8*(2).

Brinkmann, U., & Seifert, M. (2001). Face to Interface: To the problem of the confidence constitution in the Internet by the example of electronic auctions. *Journal for Sociology, 30*(1), 23-47.

Brown, J., Fortin, D. R., & Rhodes, P. (2009). The effects of contextual cues on online auction outcomes: A quasi-experimental approach *Journal: International Journal of Internet Marketing and Advertising, 5*(4), 287-308.

Bywell, C. E., & Oppenheim, C. (2001). Fraud on Internet auctions. *Aslib, 53*(7), 265-272.

Cabral, L., & Hortacsu, A. (2004). *The dynamics of seller reputation: Theory and evidence from eBay*. London, UK: University of Chicago.

Callison-Burch, C. (2009). *Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk.* Paper presented at the In Proc. EMNLP 2009, ACL and AFNLP, Suntec, Singapore

Callison-Burch, C., & Dredze, M. (2010). *Creating speech and language data with Amazon's Mechanical Turk.* Paper presented at the North American Chapter of the Association for Computational Linguistics (NAACL) 2010 Workshop, Los Angeles, CA.

Chua, C., & Wareham, J. (2004). Fighting Internet auction fraud: An assessment and proposal. *Computer, 37*(10), 31-37.

Chua, C., & Wareham, J. (2008). Parasitism and Internet auction fraud: An exploration. *Information and Organization, 18*(4), 303-333.

Chua, C., Wareham, J., & Robey, D. (2007). The role of online trading communities in managing Internet auction fraud. *MIS Quarterly, 31*(4), 759-781.

Clarke, A. C. (1983). *2010: Odyssey two*: Granada Publishing Ltd.

Clemons, E. K., Hann, I.-H., & Hitt, L. M. (2002). Price dispersion and differentiation in online travel: An empirical investigation. *Management Science, 48*(4), 534-549.

Cofta, P. (2006). Convergence and trust in ecommerce. *BT Technology Journal, 24*(2), 214-218.

Consumer reports survey of eBay users. (2007). *Consumer Reports, 72*, 12-14.

Cox, B. (2003, 2/14/2003). And the online fraud goes on... *internetnews.com* Retrieved 10/25/10, 2010

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London, England: Chapman and Hall.

Creswell, J. W. (2002). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage Publication.

Crowdflower - gold standard. (2010). Retrieved 10/25/10, 2010, from http://crowdflower.com/docs/gold

Dekel, O., & Shamir, O. (2009). *Good learners for evil teachers*. Paper presented at the Proceedings of the 26th Annual International Conference on Machine Learning.

Dellarocas, C. (2003a). The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. *Management Science, 49*(10), 1407-1424.

Dellarocas, C. (2003b). *Efficiency through feedback-contingent fees and rewards in auction marketplaces with adverse selection and moral hazard.* Paper presented at the Proceedings of the 4th ACM Conference on Electronic Commerce, San Diego, CA.

Dellarocas, C. (2006). Reputation mechanisms. In T. Hendershott (Ed.), *Handbook on economics and information systems*. Amsterdam: Elsevier Science.

Dellarocas, C., & Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science, 54*(3), 460-476.

Diamond, D. (1989). Reputation acquisition in debt markets. *Journal of Political Economy, 97*, 828-862.

Dong, F., Shatz, S. M., & Xu, H. (2009). Combating online in-auction fraud: Clues, techniques and challenges. *Computer Science Review, 3*(4), 245-258.

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). *Are your participants gaming the system?: Screening mechanical turk workers*. Paper presented at the Proceedings of the 28th international conference on Human factors in computing systems.

Easley, R. F., & Tenorio, R. (2004). Jump bidding strategies in Internet auctions. *Management Science, 50*(10), 1407-1419.

eBay buyer protection plan. (2010). Retrieved 10/25/10, 2010, from http://pages.ebay.com/help/policies/buyer-protection.html

eBay feedback scores, stars, and your reputation. (2010). Retrieved 10/25/10, 2010, from http://pages.ebay.com/help/feedback/scores-reputation.html

eBay glossary. (2010). Retrieved 10/25/10, 2010, from pages.ebay.com/help/account/glossary.html

eBay, Inc. v. Bidder's Edge, Inc, C-99-21200 RMW C.F.R. (2000).

eBay's two enormous data warehouses. (2010, 4/30/2009). Retrieved 10/25/10, 2010, from http://www.dbms2.com/2009/04/30/ebays-two-enormous-data-warehouses

Feedback abuse. (2010). Retrieved 10/25/10, 2010, from http://pages.ebay.com/help/policies/feedback-abuse-withdrawal.html

Feng, D., Besana, S., & Zajac, R. (2009). *Acquiring high quality non-expert knowledge from on-demand workforce*. Paper presented at the Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Finch, B. J. (2006). Customer expectations in online auction environments: An exploratory study of customer feedback and risk *Journal of Operations Management, 25*(5), 985-997.

Finin, T., Murnane, W., Karandikar, A., Keller, N., & Martineau, J. (2010). *Annotating named entities in twitter data with crowdsourcing.* Paper presented at the North American Chapter of the Association for Computational Linguistics (NAACL) 2010 Workshop, Los Angeles, CA.

Fisher, R. A. (1995). *Statistical tables for biological, agricultural and medical research* (6th ed.). London, England: Longman Group United Kingdom.

Ford, M. (2010). Auction Inquisitor. Retrieved 10/25/10, 2010, from http://www.auctioninquisitor.com/

Form 10-K for eBay for 2009. (2010). Retrieved 10/25/10, 2010, from http://investor.ebay.com/sec.cfm?DocType=Annual&Year=

Gavin, H. (2008). *Understanding research methods and statistics in psychology* (1st ed.). Los Angeles, CA: Sage Publications, Inc.

Gavish, B., & Tucci, C. L. (2008). Reducing Internet auction fraud. *Communications of the ACM, 51*(5), 89-97.

Ghose, A., Ipeirotis, P. G., & Sundararajan, A. (2005). *Reputation premiums in electronic peer-to-peer markets: Analyzing textual feedback and network structure*. Paper presented at the Proceedings of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems.

Goes, P., Tu, Y., & Tung, Y. A. (2009). Technical opinion: Online auctions hidden metrics. *Communications of the ACM, 52*(4), 147-149.

Grazioli, S., & Jarvenpaa, S. L. (2000). Perils of Internet fraud: An empirical investigation of deception and trust with experienced Internet consumers. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30*(4), 395-410.

Grazioli, S., & Jarvenpaa, S. L. (2003). Deceived: Under target online. *Communications of the ACM, 46*(12), 196-205.

Grazioli, S., & Wang, A. (2001). *Looking without seeing: Understanding naïve consumers' success and failure to detect Internet deception.* Paper presented at the Proceedings of the International Conference on Information Systems, New Orleans, LA.

Gregg, D. G., & Scott, J. E. (2006). The Role of Reputation Systems in Reducing On-Line Auction Fraud. *International Journal of Electronic Commerce, 10*(3), 95-120.

Gregg, D. G., & Scott, J. E. (2008). A typology of complaints about eBay sellers. *Communications of the ACM, 51*(4), 69-74.

Gyongyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). *Combating web spam with trustrank.* Paper presented at the Proceedings of the Thirtieth International Conference on Very Large Databases, Toronto, Canada.

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Heilman, M., & Smith, N. A. (2010). *Rating computer-generated questions with Mechanical Turk.* Paper presented at the North American Chapter of the Association for Computational Linguistics (NAACL) 2010 Workshop, Los Angeles, CA.

Hendershott, T. (2006). Economics and information systems. In A. Whinston (Ed.), *Handbooks in Information Systems* (Vol. 1). Amsterdam: Elsevier Science.

Heymann, P., & Garcia-Molina, H. (2008). *Can tagging organize human knowledge?* . Stanford, CA: Stanford University.

Hijikata, Y., Ohno, H., Kusumura, Y., & Nishida, S. (2006). *Social summarization of text feedback for online auctions and interactive presentation of the summary*. Paper presented at the Proceedings of the 11th international conference on Intelligent user interfaces.

Howe, J. (2006). The rise of crowdsourcing. *Wired, 14*.

Hsueh, P.-Y., Melville, P., & Sindhwani, V. (2009). *Data quality from crowdsourcing: A study of annotation selection criteria*. Paper presented at the Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing.

Johnson, P. E., Grazioli, S., & Jamal, K. (1993). Fraud detection: Internationality and deception in congnition. *Accounting, Organizations and Society, 18*(5), 467-488.

Jurca, R., & Faltings, B. (2007, 6/11/07-06/15/07). *Collusion resistant, incentive compatible feedback payments.* Paper presented at the Proceeding of the ACM Conference on E-Commerce (EC'07), San Diego, CA.

Kamins, M. A., Folkes, V. S., & Perner, L. (1977). Consumer responses to rumors: Good news, bad news. *Journal of Consumer Psychology, 6*(2), 165-187.

Kauffman, R., & Lee, D. (2009). Price rigidity in Internet retailing. SSRN.

Kauffman, R., & Wood, C. A. (2000). *Running up the bid: Modeling seller opportunism in Internet auctions.* Paper presented at the The Proceedings of the 2000 Americas Conference on Information Systems, Long Beach, CA.

Kauffman, R., & Wood, C. A. (2005). The effects of shilling on final bid prices in online auctions. *Electronic Commerce Research and Applications, 4*, 18-31.

Kauffman, R., & Wood, C. A. (2006). Doing their bidding: An empirical examination of factors that affect a buyer's utility in Internet auctions. *Information Technology and Management, 7*(3), 171-190.

Kittur, A., Chi, E. H., & Suh, B. (2008). *Crowdsourcing user studies with Mechanical Turk*. Paper presented at the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems.

Klebanov, B. B., & Beigman, E. (2009). From annotator agreement to noise models. *Computational Linguistics, 35*(4), 495-503.

Kolbe, R., & Burnett, M. (1991). Content analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research, 18*(2), 243-250.

Konrad, R. (2005, 6/27/05). eBay losing allure to some entrepreneurs. *USA Today*.

Kosara, R., & Ziemkiewicz, C. (2010). *Do Mechanical Turks dream of square pie charts?* Paper presented at the BELIV'10: Beyond time and errors: Novel evaluation methods for information visualization, Atlanta, GA.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverley Hills, CA: Sage Publications.

Ku, L.-W., Lo, Y.-S., & Chen, H.-H. (2007). *Test collection selection and gold standard generation for a multiply-annotated opinion corpus*. Paper presented at the Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions.

Lansing, P., & Hubbard, J. (2002). Online auctions: The need for alternative dispute resolution. *American Business Review*(January), 108-116.

Lee, D., Jeong, O.-R., & Lee, D. (2008). *Opinion mining of customer feedback data on the web*. Paper presented at the Proceedings of the 2nd international conference on Ubiquitous information management and communication.

Levine, D. K., & Martinelli, C. (1998). Reputation with noisy precommitment. *Journal of Economic Theory, 78*(1), 55-78.

Lucking-Reiley, D., Bryan, D., Prasad, N., & Reeves, D. (2007). Pennies From eBay: The Determinants Of Price In Online Auctions. *The Journal of Industrial Economics, 55*(2), 223-233.

Mason, W., & Watts, D. J. (2009). *Financial incentives and the "performance of crowds"*. Paper presented at the Proceedings of the ACM SIGKDD Workshop on Human Computation.

McDonald, C. G., & Slawson Jr., V. C. (2002). Reputation in an Internet auction market. *Economic Inquiry, 40*(3), 633-650.

Medved, J. (2010). Medved auction count charts. Retrieved 10/25/10, 2010, from http://www.medved.net/cgi-bin/cal.exe?EIND

Melnik, M. I., & Alm, J. (2002). Does a seller's ecommerce reputation matter? Evidence from eBay auctions. *Journal of Industrial Economics, 50*(3), 337-349.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis *The American Statistician, 54*(1), 17-24.

Merriam-Webster's collegiate dictionary. (2005) (11 ed.). Springfield, MA: Merriam-Webster.

Mierzwa, P. (2005). Squeezing new technology into old laws. *CBA Record, 19*.

Miller, N., Resnick, P., & Zeckhauser, R. (2003). *Eliciting honest feedback in electronic markets*. Harvard, MA: Harvard University.

Miller, N., Resnick, P., & Zeckhauser, R. (2005). Eliciting honest feedback: The peer prediction method. *Management Science*(September).

Morgan, G., Leech, N., Gloeckner, G., & Barrett, K. (2007). *SPSS for introductory statistics: Use and interpretation* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691-692.

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

Nikitkov, A. N. (2006). Information assurance seals: How they impact consumer purchasing behavior. *Journal of Information Systems, 20*(1), 1-17.

Nikitkov, A. N., & Stone, D. N. (2006). *Online auction deception: A forensic case study of an opportunistic seller.* St. Catharines, Ontario, Canada Brock University

Nowak, S., & Ruger, S. (2010). *How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multi-label image annotation.* Paper presented at the Proceedings of the international conference on Multimedia information retrieval.

Ofir, C., & Simonson, I. (2001). In search of negative customer feedback: The effect of expecting to evaluate on satisfaction evaluations *Journal of Marketing Research, 38*(2), 170-182.

Online auction fraud complaints still rising, says consumer watchdog. (2004). Retrieved 10/25/10, 2010, from http://www.nclnet.org/news/2004/internet_fraud_stats.htm

Palmer, J. W. (2002). Web site usability, design, and performance metrics. *Infomation Systems Research, 13*(2), 151-167.

Pandit, S., Chau, D. H., Wang, S., & Faloutsos, C. (2007). *Netprobe: A fast and scalable system for fraud detection in online auction networks*. Paper presented at the Proceedings of the 16th International Conference on World Wide Web.

Papaioannou, T. G., & Stamoulis, G. D. (2005). *An incentives' mechanism promoting truthful feedback in peer-to-peer systems.* Paper presented at the Proceeding of IEEE/SCM CCGRID, Cardiff, UK.

Pavlou, P. (2002). Institutional trust in interorganizational exchange relationships: The role of electronic B2B marketplaces. *Journal of Strategic Information Systems, 11*(3/4), 215-243.

Pavlou, P., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research, 17*(December), 392-414.

Pavlou, P., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research, 15*(1), 35-53.

Pennington, R., Wilcox, H. D., & Grover, V. (2003). The role of system trust in business-to-consumer transactions. *Journal of Management Information Systems, 20*(3), 197-226.

Pinchak, C., Rafiei, D., & Lin, D. (2009). *Answer typing for information retrieval*. Paper presented at the Proceeding of the 18th ACM conference on Information and knowledge management.

Poortinga, W., & Pidgeon, N. F. (2004). Trust, the asymmetry principle, and the role of prior beliefs. *Risk Analysis, 24*(6), 1475-1486.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems: Facilitating trust in Internet interactions. *Communications of the ACM, 43*(12), 45-48.

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye (Ed.), *The economics of the Internet and e-commerce* (Vol. 11). Amsterdam: Elsevier Science.

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics, 9*(2), 79-101.

Shapiro, C. (1983). Premium for high quality products as returns to reputations. *Quarterly Journal of Economics, 98*(4).

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York, NY: Penguin.

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). *Cheap and fast---but is it good?: Evaluating non-expert annotations for natural language tasks*. Paper presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing.

Sorokin, A., & Forsyth, D. (2008, 7/15/08). *Utility data annotation with Amazon Mechanical Turk.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08) Anchorage, AK

Su, Q., Pavlov, D., Chow, J.-H., & Baker, W. C. (2007). *Internet-scale collection of human-reviewed data*. Paper presented at the Proceedings of the 16th international conference on World Wide Web.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*: Anchor.

Sutton, W., & Linn, E. (1976). *Where the money was: The memoirs of a bank robber* New York: Viking Press.

Tang, J., & Sanderson, M. (2010). *Evaluation and user preference study on spatial diversity*. Paper presented at the ECIR 2010 - 32nd European Conference on Information Retrieval.

von Ahn, L., & Dabbish, L. (2004). *Labeling images with a computer game*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Watch out for cyber scrooge this holiday season. (2006). Retrieved 10/25/10, 2010, from http://www.nclnet.org/news/2006/cyber_monday112-2006.htm

Weber, R. P. (1990). *Basic content analysis*. Beverly Hills, CA: Sage.

Weinberg, B. D., & Davis, L. (2005). Exploring the WOW in online auction feedback. *Journal of Business Research, 58*(11), 1609-1621.

Wenzel, J. (2008). *Feasibility study for the crowdsourcing of geometric reasoning in mechanical CAD*. Edinburgh, UK: University of Edinburgh.

What is feedback and how does it affect my reputation? (2010). Retrieved 10/25/10, 2010, from http://pages.ebay.com/help/feedback/questions/feedback.html

Winn, J. (2005). Contracting spyware by contract. *Berkeley Technology Law Journal, 20*(3), 1345-1362

Wood, C. A. (2004). Current and future insights from online auctions: A research framework of selected articles in online auctions. In M. Shaw, R. Blanning, T. Strader & A. Whinston (Eds.), *Handbook on electronic commerce*: Springer-Verlag.

Wood, C. A., Fan, M., & Tan, Y. (2002, 12/14/2002-12/15/2002). *An examination of the reputation systems for online auctions.* Paper presented at the In Proceedings of the Workshop for Information Systems and Economics (WISE 2002), Barcelona, Spain

Yamane, T. (1967). *Statistics, an introductory analysis* (2nd ed.). New York, NY: Harper and Row.

Yan, T., Kumar, V., & Ganesan, D. (2010). *CrowdSearch: Exploiting crowds for accurate real-time image search on mobile phone*. Paper presented at the MobiSys 2010 - 8th Annual International Conference on Mobile Systems, Applications and Services Retrieved from http://www.sigmobile.org/mobisys/2010/keynotes.html

Your user agreement. (2010). Retrieved 10/25/10, 2010, from http://pages.ebay.com/help/policies/user-agreement.html

Zacharia, G., Moukas, A., & Maes, P. (1999). *Collaborative reputation mechanisms in electronic marketplaces.* Paper presented at the Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences.

Zhang, J. (2006). The roles of players and reputation: Evidence from eBay online auctions. *Decision Support Systems, 42*(3), 1800-1818.