Online Trust: An International Study
Of Subjects’ Willingness to Shop at Online Merchants, Including
The Effects of Promises and of Third Party Guarantees

Eric K. Clemons
University of Pennsylvania
clemsons@wharton.upenn.edu

Josh Wilson
University of Pennsylvania
wilsonjs@wharton.upenn.edu

Christian Matt
Ludwig Maximilians University
matt@bwl.lmu.de

Thomas Hess
Ludwig Maximilians University
thess@bwl.lmu.de

REN Fei
Peking University GSM
fren@gsm.pku.edu.cn

Fujie JIN
University of Pennsylvania
jinfujie@wharton.upenn.edu

Abstract
Significant differences exist among consumers’ online shopping behavior in different international markets. This paper compares consumer behavior in four different markets: (1) The US, the largest and most mature market, (2) Germany, a similarly advanced western market (3) China, the fastest growing online market, but one where consumers are plagued by incidents of counterfeits, forgeries, and spoiled or defective items, and (4) Singapore, an advanced market, culturally similar in some ways to China, but with a strong legal system.

We performed laboratory experiments simultaneously in all four countries. We used three experimental treatments: (1) No assurances of product quality or authenticity, (2) promises of quality and authenticity, and (3) promises backed up by third party assurances. We examined subjects’ responses for all three treatments, and for vendors’ with different degrees of riskiness. We confirmed that significant differences exist in consumer behavior, but these differences were not always what we expected. Chinese consumers do appear to have trust in their best online vendors. US consumers appear to treat online shopping very similarly to the way they treat shopping in physical venues.

1. Introduction

We have conducted a study of consumers’ attitudes towards the risks of online shopping. Understanding trust remains essential to understanding the growth of and limits to online shopping, and understanding national differences is critical as companies attempt to globalize their more localized online retailing strategies. The study was conducted simultaneously, in four nations that we thought were representative of different shopping traditions; we believe this to be the first paper to explore systematic differences in consumers’ trust of online shopping across four countries at the same time. The United States has the most mature and most advanced e-commerce marketplace. Germany is representative of a large portion of e-commerce in the EU and in much of the developed world. We believed that China might be an outlier for two reasons. It is both the largest and the fastest growing online marketplace in the world, which might be indicative of teething problems of various sorts. And it has the highest incidence of product tampering, product fraud, and product counterfeiting of any major economy, which might have led to consumer concerns about the online marketplace. We studied Singapore. Since the vast majority of Singapore’s population is ethnically Chinese, we used Singapore in prior studies to see if it exhibits online behavior more similar to Chinese or more similar to Western behavior [20]. However, since consumer protection laws and rule of law in general are much more advanced in Singapore than in China, we felt that Singapore might exhibit behavior that was significantly different from China, and that it might do so without actually reproducing the behavior of consumers elsewhere. All of these assumptions and intuitions can be tested, and our experiments were designed to test them.

Our study involved three independent variables. One independent variable was the reputation of the online seller. Reputation varied from the buyer’s personal favorite shopping site to unknown and unrated. A second variable was the assurances that the online sellers offered the buyers. Assurances ranged from no assurances, through simple promises, to promises backed up by third party guarantees. The third is the product type. We explored a range of product types, including food and beverage items, clothing, sporting goods, and electronics. The dependent variable was the customer’s WTP for the specific product, from the specified merchants, with the specified assurances.

Our study involved exposing all subjects to two of three treatments. One group was exposed to no assurances and to promises only (the MIN-MID Group). The second group was exposed to no assurances and to promises backed by third party guarantees (the MIN-MAX Group). The final group was exposed to promises alone and to promises backed up by third party guarantees (the MID-MAX Group). Each subject made purchase decisions for the full range of products.

Exposing subjects to two treatments introduces complexity into the data analysis. But it also allows us to control for group heterogeneity in an important way.
This is described more fully in section 6.

Our expectation was that promises alone would have little impact on customers’ WTP. Our thinking was that if customers trusted a website, then promises would not be necessary, and that if customers did not trust a website they would not trust its promises either; either way promises would have no effect. This turned out to be incorrect. Our expectation was that third party guarantees would be most effective. This turned out to be only partly supported. Our expectation was that China would be a significant outlier in one direction, and Singapore would be a significant outlier in the opposite direction. Again, this turned out not to be entirely true. All of this is described more fully in our hypotheses (section 4) and our analyses (sections 7 through 11).

The US online market is the most mature online market in the world, and the behavior of online shoppers in the US is remarkably similar to the behavior of shoppers in traditional physical venues. Shoppers at a range of physical stores, from Wal-Mart and Rite-Aid to Tiffany’s and Brooks Brothers, make their purchases without requesting promises that the goods they buy are authentic, and without requiring third party guarantees. The online behavior of US shoppers seems similar to their offline behavior. In contrast, promises and third party guarantees are significant in other markets, with vendor type and nature of assurances interacting in different ways in different countries. Promises were not always without value, and third party guarantees did not always increase the value of promises.

The structure of the paper is as follows. Section 2 provides a literature review. Section 3 presents the motivation for our hypotheses and section 4 presents the hypotheses. Section 5 describes the experiment. Section 7 and Section 8 present our findings for each of the 6 hypotheses. Section 9 provides further discussion and outlines directions for future research.

2. Literature Review

2.1. Information Asymmetry

The impact of information asymmetry on the behavior of buyers and sellers and on the performance of markets has received considerable attention in the information economics and information-based strategy communities since Akerlof, Spence, and Stiglitz shared the Nobel Prize in Economics. We assume that our readers are familiar with Signaling [29] and its use in marketing [19] to signal high quality.

2.2. Information Asymmetry in E-commerce

Addressing information asymmetry may be even more critical when markets and commerce move online, given the anonymity of parties, the ease of masquerading, and the difficulty of actually inspecting products before purchase. Since the identity of both customers and the store itself may not be known, receiving payment, receiving quality merchandise, and then being able to obtain a satisfactory outcome if problems occur, are all uncertain. Difficulty in identifying and authenticating vendors on the Internet makes it easy for a vendor to profit by cheating, then to exit, and then to reenter under a different name, thus continuing fraudulent activities [10, 15, 16]. Information asymmetry can reduce customers’ willingness to purchase online or reduce the price paid for merchandise sold online [2]. This can impede the growth of online markets or in the limit lead to their total collapse [1].

Different factors attribute to trust in e-commerce during different stages of customers’ online shopping behavior [3]. The prior existence of a trusted physical store serves as a strong signal. A trusted third party’s evaluation of vendors can serve as a signal and facilitate transactions [4]; vendor guarantees can provide a signal and boost the trust of e-commerce consumers, even in the absence of known reputation of the sellers and their products [27]. Third party guarantees from companies like buySAFE [4] can strengthen vendors’ assurances, but are not yet widely available in China.

2.3. Lab Experiments

The experiment is firmly grounded both in theory and in practice. It draws on the literature on information asymmetry, and on signaling mechanisms and reputational capital. It draws also on experimental studies of indicators that consumers use to assess the relative risk of different online sellers, including (1) photographs of products [26]; (2) reputation of an associated physical seller [24, 30]; (3) personal experience with the seller [13]; (4) recommendations from friends [23]; (5) online ratings and online testimonials [6, 17]; (6) direct promises [18]; and (7) attractiveness of the website [14]. We control for these factors by providing the same photographs to all subjects and by providing subjects with the same descriptions for conditions. We do not show actual URLs or websites, so the final factor (attractiveness of website) should not affect subjects’ perceptions or decisions.

3. Motivation for Our Hypotheses

3.1. Differences between China And Other E-commerce Markets

Several papers have studied e-commerce in China and have identified the factors that have increased trust; most find the same factors that increase trust in the United States [12, 22]. Some studies discuss the severity of trust issues in China, and ways in which factors differ in importance in creating trust in e-commerce in China compared to other markets [25, 33]. The entrepreneurs with whom we spoke believed that Chinese consumers had stronger concerns about quality, but they didn’t have concrete data to support this; we used their beliefs as a starting point for this research. We know no other
studies that quantifies differences between trust among Chinese consumers and trust among consumers elsewhere. We chose U.S., Germany, Singapore, and China, as the four markets for our study because of the varying degrees of online maturity and legal protections in these markets.

### 3.2. Case Methodology in the Development Of our Hypotheses

This research started with a small number of semantically rich case studies to explore a theory of trust development in e-commerce in China. The general validity and challenges of this method are explained by Eisenhardt and Dyer [7, 9]. The application of case study methodology in Management Information System is explored by Galliards and Land, and Lee [11, 21].

We conducted several rounds of interviews with three large online shopping sites in China, 360buy, Taobao, and YiHaoDian (The Store), in order to assess how three very different businesses dealt with the problems of information asymmetry and lack of trust among Chinese online shoppers. These discussions helped us understand how each company addressed the problems of building, preserving, and profiting from consumer confidence. These studies suggested that the reputation of an online vendor was vitally important to the vendor's success. We operationalized this by suggesting that reputation would be directly reflected in consumers’ WTP for products. Our cases, and their contribution to the formulation of our hypotheses, is described in detail elsewhere.

### 4. Hypotheses

Our discussions with executives at 360buy, Taobao, and Yihaodian suggested the following: (1) vendors’ reputational capital may be the greatest determinant of consumers’ WTP; (2) vendors’ promises of product quality are of little value from low quality vendors because buyers may not view these promises as binding in the absence of reputational capital; (3) third-party guarantees would help create greater trust and faster growth in online markets; (4) Chinese consumers may be less trusting and Chinese vendors may encounter greater trust penalties; and (5) different risk mitigation mechanisms may have different effectiveness in China, due to different consumer experience with online commerce generally, with third party warranties and performance bonds, and with the legal system when trying to follow up on complaints. These led directly to our six hypotheses.

**Hypothesis 1:** In every country, with purchases aggregated across all product types, as vendor quality decreases, consumer’s WTP will decrease.

The idea that trust is directly observable through subjects’ expressed WTP is central to our experiments. This is critical to establish, because if we cannot observe this in the laboratory then our experiments cannot test any of our more important hypotheses. Hypothesis 1 attempts to determine the extent to which reputation matters to subjects’ WTP for items from different categories of seller.

**Hypothesis 2:** In every country, with purchases aggregated across all product types, promises alone will not increase consumers’ WTP compared to these consumers’ WTP for the same products purchased from the vendors but without promises. (Promises alone refers to promises offered without third party guarantees.)

Hypothesis 2 suggests that promises from low quality vendors will not be credible. The value of promises alone is directly observable through differences in subjects’ WTP for products online, from a range of merchant types, with and without promises. The extent to which differences in WTP vary by merchant type and merchant reputation will indicate the role of reputational capital in determining the value of promises. A promise from Carrefour or Wal-Mart is credible because they have reputations that they do not want to damage; they are likely to back up their promises. Alternatively, if the merchant is unreliable then the promise will probably be seen as unreliable as well.

**Hypothesis 3:** In every country, and across all product types, consumers’ WTP for products from vendors who offer promises backed by a trusted third party will be greater than consumers’ WTP for the same products from the same quality vendors who offer promises not backed up by a trusted third party.

Hypothesis 3 suggests that promises will be effective only if they are backed up by third party guarantees. This hypothesis interacts in complex ways with hypothesis 2. In particular, this suggests that promises from low quality vendors will be credible if they are backed up by third party guarantees, even if these promises alone are not credible (consistent with hypothesis 2). Additionally, hypothesis 3 suggests that promises from high quality vendors are also more effective if backed up by third party guarantees, even if these promises alone are not credible (which is not consistent with hypothesis 2). Combining hypothesis 2 with hypothesis 3 suggests that vendors without established reputations could benefit early in their online evolution through the use of third party guarantees. The value of third party guarantees is directly observable through differences in subjects’ WTP for products online, from a range of merchant types, with and without third party guarantees.

**Hypothesis 4:** For each vendor quality level, the decreases in consumers’ WTP associated with vendors’ quality level will vary across countries.

Hypothesis 4 suggests that the impact of trust and reputation on WTP will vary by country. Based on discussions with online retailers, we expected trust to
be less of a problem among consumers in Singapore, with consumers in Germany and the United States exhibiting intermediate behavior. We expected online trust to be the greatest problem in China, given the incidence of problems with counterfeit and defective merchandise, and with product tampering in China. We test this using subjects’ stated WTP for merchandise from all vendor types under the baseline condition, with no explicit promises or third party guarantees.

Hypothesis 5: Across vendor quality levels, vendor promises without 3rd party backing will increase consumers’ WTP in different degrees across countries.

Hypothesis 6: Across vendor quality levels, vendor promises with 3rd party backing will increase consumers’ WTP in different degrees across countries.

Hypotheses 5 and 6 suggest that mechanisms to increase consumers’ confidence and to increase consumers’ WTP will vary across nations. We expected variation across nations, based on national differences in experience with online shopping, the severity of problems with traditional physical shopping, and the effectiveness of national legal codes in creating a sense of consumer protection.

5. Experimental Design

The experiment is firmly grounded both in theory and in practice. It draws on the literature on information asymmetry, and on signaling mechanisms and reputational capital as mechanisms for dealing with information asymmetry. It draws also on experimental studies of indicators that consumers use to assess the relative risk of different online sellers. Our experimental design has been described in detail in prior publications [5]. We review essential details here.

Subjects were shown only the two treatments relevant to their participation in the experiment. Subjects could choose not to express a WTP for any product they did not have an interest in purchasing. Those subjects who were interested in a product and who expressed a willingness to buy it online from at least some vendors, could choose which vendors they would purchase from and under which treatment. That is, we allowed subjects to select some vendors but not all vendors, based on their perception of risk from individual vendors and their own risk preferences. This is consistent with subsequent work by Gu et al., [33], on elimination by aspects, showing that Chinese consumers would first accept or reject a merchant based on perceived riskiness; they would then express a WTP only for those remaining merchants deemed acceptable to them.

Subjects were shown a photograph of the product and a description of the product as would be found on a trusted vendors’ site. The subjects were given two reference prices, the manufacturer’s suggested retail price and an average online selling price from a highly respected seller. Subjects were then asked to provide what they would be willing to pay for the product, from the set of sellers, under two of three treatments:

1. (MIN) Subjects were asked the amount that they would be willing to pay for the product from each of the described vendors, with no specific assurances from the vendor concerning product quality or authenticity.
2. (MID) Subjects were asked the amount that they would be willing to pay for the product from each of the described sellers, in the presence of explicit promises from the seller that the product is new and genuine, and can be returned for a full refund if the buyer is not fully satisfied.
3. (MAX) Subjects were asked the amount that they would be willing to pay for the product from each of the described sellers, in the presence of explicit promises from the seller that the product is new and genuine, and can be returned for a full refund if the buyer is not satisfied. In addition, sellers were stated to be bonded. Funds for refund are guaranteed to be available, and refunds were administered by a trusted third party with access to an escrow account, instead of by the seller.

Those subjects who were exposed to treatments MIN and MID were considered to have been exposed to the MIN-MID test. Likewise, subjects who were exposed to MIN and MAX were referred to as MIN-MAX and those who were exposed to MID and MAX were referred to as MID-MAX. Through Qualtrics, subjects were randomly assigned to one of the tests. All of the subjects came from behavioral labs at each of the Universities associated with the experiment: the Wharton Behavioral lab, the Singapore Management University Lab, the Ludwig Maximilian University Lab, and Peking University Guanghua School of Management Lab. The subjects recruited for this experiment were students at the participating schools.

In order to determine the different types of vendors that exist in the e-commerce ecosystem, we did extensive online research to determine where certain products could be purchased. Using models from previous academic publications, we developed six vendor descriptions designed so that subjects would perceive them with differing levels of riskiness. The sellers were described as shown below. The phrases in parentheses were used as short identifiers of the vendors throughout the experiment.

1. (Safe Vendor) A well-known online seller in the country in which the experiment is being performed, with a strong reputation, such as Amazon, or a website associated with a well-known physical seller, such as Wal-Mart or Carrefour.
2. (Favorite Vendor) A website in the country in
which the experiment is being performed, that is a
favorite of the subject.
3. **(Friend’s Favorite Vendor)** A website in the
country in which the experiment is being per-
formed, that the subject has not used before, but is
strongly recommended by close personal friends.
4. **(Unknown Rated Vendor)** A website in the coun-
try in which the experiment is being per-
formed, that neither the subject nor friends have used be-
fore, but that has a strong online rating.
5. **(Unknown Unrated Vendor)** A website in the coun-
try in which the experiment is being per-
formed, that neither the subject nor friends have used before, and that does not have an online rat-
ing.

We will post the full vendor descriptions included in
the experiment and a screenshot of an experiment
page online, as they appeared to the subjects.

Product selection involved analysis to ensure that
products across test sites were of similar interest to
subjects in different markets. Thus, in the sporting
goods category, American and German students were
offered an opportunity to buy tennis rackets, and in
China and Singapore they were offered a table tennis
paddle or running shoes. Categories included sporting
equipment, high and low end consumer electronics,
ecosystems, clothing, and food.

We presented the subjects with two prices for each
product, Manufacturers suggested retail price (MSRP)
and Average Online Retail Price. We defined MSRP as
the price that the manufacturers recommended for their
products, and this price could often be found at the
manufacturer’s own website or a trusted online seller’s
website. We defined average online retail price as the
price at which a trusted online seller like amazon.com
or walmart.com lists the product for sale. Participants
were informed that MSRP in the experiment is a proxy
for the price a consumer would get at a physical full
service store, which in some cases was similar to the
price from the most trusted online vendor.

6. **Data Cleaning**

Experiments that involve subjects entering their
WTP require special attention in order to ensure a
clean and comparable dataset for any pursuant anal-
yses. Data cleaning posed a significant challenge for
us, as detailed in a previous conference paper on statis-
tical challenges in e-commerce research [5]. Space
limitations prevent us from duplicating our data clean-
ing procedures here, but they are available online.
Table 1 contains the final number of subjects and ob-
servations that were used for the analyses, after clean-
ing. Of course, the fully intact dataset will be posted
online, with data from subjects who inverted and with
data from both risk-tolerant and risk-intolerant indi-
viduals included, along with the clean dataset that was
used for the analyses.

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>Germany</th>
<th>China</th>
<th>Singapore</th>
</tr>
</thead>
<tbody>
<tr>
<td>MinMid</td>
<td>27</td>
<td>32</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>MinMax</td>
<td>24</td>
<td>41</td>
<td>45</td>
<td>29</td>
</tr>
<tr>
<td>MidMax</td>
<td>21</td>
<td>38</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td># of Subjects</td>
<td>72</td>
<td>111</td>
<td>111</td>
<td>91</td>
</tr>
<tr>
<td># of Obs.</td>
<td>4480</td>
<td>8321</td>
<td>8370</td>
<td>5544</td>
</tr>
</tbody>
</table>

Table 1.— The number of subjects and observations in
the final dataset.

7. **Regression Models**

In order to assess all of the hypotheses, we fit two
different regression models to the data. Since the three
hypotheses concern within country variation, we mod-
el these hypotheses using separate regressions for each
country. And, as the last three hypotheses address
variation across countries, we fit a model that accounts
for the across-country variation and that allows for
direct comparison of each country’s results. Given the
hierarchical structure of the data, with each subject
providing multiple entries for WTP and each country
providing multiple subjects, a hierarchical linear model
is able to account for the multilevel clustering that
results from this data structure [16, 28]. Additionally,
for each model, the WTP data have been normalized
by dividing by the reference price of the product; this
makes data comparable across each of the countries.

The first three hypotheses concern within country
variation due to different vendor quality and variation
due to different types of trust mechanisms (i.e., a
promise, or a promise plus a third party guarantee).
Therefore, we include separate categorical variables
that record each level of vendor quality and each type
of trust assurance. We also want to assess whether
trust mechanisms have a different effect at each level
of vendor quality, and we capture these differences
with interactions. We want to control for any variation
that may come from subject’s risk assessments of the
different products. For example, a fake piece of choc-
olate could be deadly, while a fake Blu-ray may be
both innocuous and common. Since a subject who
decides to purchase a product will enter ten separate
prices, one for each combination of vendor type and
two of the three treatment types, these data points are
clustered by subject. We model this clustering by
allowing the intercepts to vary for each subject. Model
1 is separately fit to each country’s data:

\[
Y_{ij} = \text{Intercept}_j + \text{Vendor} + \text{Promise} + \text{Product} + \text{Promise*Vendor} + \text{error}, \text{for } i = 0, 1, 2, \ldots, n_j
\]

Where each \( i \) is a separate observation for subject \( j \).

\[
\text{Intercept}_j \sim \text{N} \left( \text{Mean ForSubject}, \text{Variance ForSubject} \right) \text{ for } j = 0, 1, 2, \ldots, n
\]

**Promise**, **Product**, and **Vendor** are all vectors.

Where each \( j \) is a separate subject from a country’s
sample.

The last three hypotheses address cross-country differ-
ences in variation that may be due to changes in vendor quality or changes in the trust assurance offered. Model 2 is fit to all of the countries’ aggregated data:

\[
Model 2: \text{WTP} \sim N(\text{Intercept} + \text{Vendor} + \text{Promise} + \text{Promise*Vendor}, \text{WTP Variation})
\]

8. Examination and Discussion
Of Hypotheses 1, 2, and 3

Hypothesis 1 states that WTP decreases as the seller’s reputation decreases. First, we will assess this hypothesis using descriptive statistics and visualizations, and then proceed to the regression analysis. A visualization of the data can be found in figure 1, which contains violin plots for the normalized WTP data by each vendor, for each country.

<<Figure 1 goes here>>

The blue “violin shape” in each of the plots represents the full range of customers’ expressed WTP for all single product types, purchased from a single vendor. Data have been normalized on the reference price for each product so that different products could be combined for each merchant type. The shapes in the figure can be understood as smoothed histograms that have been rotated ninety degrees, and centered on the vertical axis to facilitate the visualization of patterns. Therefore, the maximum width (originally the height of the histogram before rotation) of the shape corresponds to the maximum frequency of customers’ willingness to offer a price within a very small range. For example, the density for the ‘Safe’ vendor in the US plot is thickest where WTP equals the reference price. This suggests that subjects in the US most frequently pay the reference price for a product purchased from a ‘Safe’ vendor. The presence of long tails in both directions suggests that some subjects expect to purchase at a discount to the reference price, while others are willing to offer a premium. In contrast, the shape for the ‘Unknown Unrated’ vendor in the US appears to be bi-modal, with many subjects offering around three-fourths of the reference price, while many others drop out of the market entirely and do not purchase (which is represented as a WTP of zero). White dots represent the mean of customers’ WTP, and the black bars show the interquartile ranges.

Hypothesis 1 suggests that consumers are willing to pay more from vendors with higher reputations. In support of hypothesis 1, the mean and interquartile range can be seen to be monotonically non-decreasing. However, for each country, the difference in WTP between the ‘Safe’ vendor and the ‘Personal Favorite’ vendor is much smaller than the subsequent drops in WTP from ‘Friend’s Favorite’ to ‘Unknown Unrated’.

Table 2 shows each country’s normalized WTP data fit to model 1 discussed in section 7. Hypothesis 1 can be assessed by comparing the additive fixed effect estimations of the non-baseline vendor types to the baseline vendor type ‘Safe’, which has been coded as the intercept. These estimations are labeled in Table 2 with a ‘Vendor’ preceding the specific vendor type. In confirmation of the visual inspection, the US (β = -.02, SE of β = .02) and Germany (β = -.01, SE of β = .02) treat the ‘Personal Favorite’ vendor and the ‘Safe’ vendor as equivalents; that is, subjects do not differentiate between the two vendors and therefore do not apply a discount to the ‘Personal Favorite’ vendor. However, this is not the case in China and Singapore, where even a product purchased from a ‘Personal Favorite’ vendor may be discounted compared to the ‘Safe’ vendor; the stars in the table denote these differences as statistically significant. Comparing the estimate for ‘Friend’s Favorite’ and each subsequent vendor to the baseline ‘Safe’ vendor, we see that the WTP starts to drop precipitously in each country; these differences are also statistically significant. To assess whether other pairwise comparisons of vendors result in large effect sizes and statistical differences, we ran the regression with different vendor types as the baseline. Although not seen in table 2, the pairwise comparisons of the bottom three vendors show that the differences between them are also statistically significant for all four countries. Inspection of the differences in the estimates and the standard errors of the estimates in table 2 provides a crude means of confirming these results; counting the number of standard deviations between estimates for different vendor types suggests that the differences are significant. Table 2 supports hypothesis 1.

<<Table 2 goes here>>

Hypothesis 2 states that WTP is not significantly altered by simple promises from online vendors if those promises are not backed up by third party guarantees. Referring to table 2, hypothesis 2 can be assessed by comparing the baseline ‘No Promise’ to the estimations of fixed additive effects for ‘Promise’, and by inspecting the interactions of ‘Promise’ with the different vendor types. Interactions are at the bottom of the fixed effects portion of the table and are denoted with an ‘x’ between the vendor type and the assurance type. For every country, the estimation of the effect of ‘Promise’ is positive. This means that offering a promise is associated with an increase in WTP. The effect of ‘Promise’ in the US is not statistically different from zero. The effect is statistically different from zero to varying degrees in each of the remaining countries, which is not consistent with hypothesis 2.

The impact of a promise from the lowest quality vendors in China is disproportionately high. This may suggest that the buyers’ expectation is that these vendors are offering low-cost counterfeits, unless vendors explicitly claim otherwise. It is counterintuitive to the authors that these promises should be believed. But this behavior is demonstrably inconsistent with hypothesis 2. Lastly, Singapore data suggest that the
impact of promises is low for both the ‘Unknown Ranked’ and the ‘Unknown Unranked’ vendors than it is for high quality vendors. Results from Singapore are also inconsistent with hypothesis 2.

Thus, we find very little support for hypothesis 2.

Hypothesis 3 states that WTP is increased by promises from online vendors when those promises are backed up by third party guarantees. The regression results in table 3 can be used to assess this hypothesis in much the same way as they were used to assess hypothesis 2. We find that hypothesis 3 is supported in every country except the US. We find that in some countries a third party guarantee produces more of an increase in WTP for low quality vendors than for high quality vendors, again consistent with hypothesis 3.

The impact of third party guarantees differs across the four countries. This will be explored in the discussion of hypotheses 5 and 6.

9. Examination and Discussion of Hypothesis 4, 5, and 6

Hypothesis 4 suggests that the relationship between vendor reputation and consumers’ WTP varies across countries. Referring back to figure 1, we can see by inspection that consumers’ WTP across vendor types does appear to differ from country to country. Table 3 contains the results from fitting model 2 to the data that has been aggregated across all of the countries; it shows that the variations between vendor reputation and WTP are statistically different from country to country. However, these differences are not what we expected. The US and China have similar WTP for most of the vendors types, except Chinese consumers discount more for the lowest quality vendor. Singapore has high WTP from the three highest quality vendors, but subjects’ WTP declines rapidly for the last two vendor types. In Germany the most dramatic decrease in WTP occurs only for the riskiest vendor. These results are also observable from the violin plots shown in figure 1.

We can directly test hypothesis 4 by creating confidence intervals (95%) from the coefficients and standard deviations in table 3, and compare whether the confidence intervals across countries overlap. For example, the upper bound of a confidence interval for the coefficient of ‘Third Party’ in the US would be 0.03, while the lower bound of a confidence interval for this coefficient in Germany would be 0.06. Since these intervals do not overlap, the difference between the US and Germany for the ‘Third Party’ is therefore significant.

Hypothesis 5 suggests that the impact of promises alone will vary from country to country. Hypothesis 6 suggests that the impact of the addition of third party guarantees likewise will vary from country to country. We analyze these hypotheses in two different ways, using graphical tables and statistical regressions.

Figure 2 presents the impact of treatments on consumers’ WTP in all four countries, with a panel for each country and a violin plot for each assurance type within country. An inspection of figure 2 shows that consumer behavior in all four countries is different. In the US, the interquartile ranges do not shrink significantly as the type of assurance changes, while the corresponding ranges shrink much more in China and Germany. The interquartile range in Singapore decreases between the ‘No Promise’ and the ‘Promise’ treatments, but seems to remain unchanged between the ‘Promise’ and ‘Third Party’ treatments.

10. Discussion, Limitations, and Directions for Future Research

We found support for the importance of vendor reputation in all four countries, supporting hypothesis 1. We found that assurances did not matter in the way we expected, providing only limited support for hypotheses 2 and 3. We did find differences across countries in the impact of treatments on consumers’ WTP, supporting hypothesis 4. The impact of both simple promises and promises backed up by third party guarantees did differ by country, providing support for hypotheses 5 and 6, but these differences were not what we had expected, given our discussions with Chinese executives before we conducted our experiments. We do not yet know why results were not what executives expected, but what we do know suggests four possibilities for future research.

1. The data may not reflect as yet unknown systematic differences in shopping experience that we did not capture in our surveys. We will do follow-up research to see if prior experience with e-commerce fraud is in any way correlated with subjects’ responses. We also want to see if we can operationalize ways in which differences across countries, like legal systems or experience with online shopping, may be reflected in our data.

2. The survey data may accurately reflect differences in shopping behavior that are related to the maturity of the online market. The US online market appears similar to the offline market; consumers do not expect promises or third party guarantees when shopping in physical stores, and they do not appear to value them while shopping online. Our results in China and Germany suggest that assurances do matter in some markets with less experience. Further work could assess whether behaviors in China and Germany begin to resemble that in the US more closely. Perhaps other, less mature online markets, like Egypt or Thailand, would be still more different from the US.

3. Singapore may actually be unique, and Singapore shoppers might have very different attitudes to-
wards risk. Singapore has a well-developed legal system. Shoppers might perceive that there is little opportunity for a seller to escape the legal penalties associated with making false claims about product quality and product authenticity; in that case, shoppers would receive little additional value from third party guarantees. We can explore this with attitudinal surveys.

4. China may actually be unique, and China shoppers might have very different attitudes towards risk. The operators of the top Chinese shopping websites, like Yihaodian and 360buy, assured us that they enjoy much stronger reputations than traditional physical stores. In that case, we might expect the behavior of Chinese online shoppers at their most trusted websites to resemble that of German online shoppers, even if the behavior of Chinese shoppers more generally is not similar. We can explore this with attitudinal surveys conducted.

We find that reputation matters. The effects of promises and third party assurance are more complex than we initially thought, because they interact in complex ways with vendor type and with country; differences in country probably reflect differences in experience and legal system in ways that we did not attempt to capture in this study. Differences among countries are indeed significant.

11. References


Figures 1 and 2 and Tables 2 and 3 Follow

![Violin plots of WTP, across merchant type, by country. White dots represent the mean and the black bars represent the interquartile range.](image1)

![WTP, by treatment type, for each country.](image2)
### Regression Results (Model 1)

**Dependent variable:**

<table>
<thead>
<tr>
<th></th>
<th>WTP / Reference Price</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(US)</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept(Safe, No Promise, Cosmetic)</td>
<td>0.97***</td>
</tr>
<tr>
<td>Vendor Personal Favorite</td>
<td>-0.04</td>
</tr>
<tr>
<td>Vendor Friend’s Favorite</td>
<td>-0.14**</td>
</tr>
<tr>
<td>Vendor Unknown Ranked</td>
<td>-0.34***</td>
</tr>
<tr>
<td>Vendor Unknown Unranked</td>
<td>-0.53***</td>
</tr>
<tr>
<td>Promise</td>
<td>0.02</td>
</tr>
<tr>
<td>Third Party</td>
<td>0.03**</td>
</tr>
<tr>
<td>Personal Favorite x Promise</td>
<td>0.004</td>
</tr>
<tr>
<td>Friend’s Favorite x Promise</td>
<td>-0.01</td>
</tr>
<tr>
<td>Unknown Ranked x Promise</td>
<td>-0.04</td>
</tr>
<tr>
<td>Unknown Unranked x Promise</td>
<td>-0.02</td>
</tr>
<tr>
<td>Personal Favorite x Third Party</td>
<td>-0.01</td>
</tr>
<tr>
<td>Friend’s Favorite x Third Party</td>
<td>-0.02</td>
</tr>
<tr>
<td>Unknown Ranked x Third Party</td>
<td>0.01</td>
</tr>
<tr>
<td>Unknown Unranked x Third Party</td>
<td>0.07**</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
</tr>
<tr>
<td>ID (Intercept)</td>
<td>0.2169</td>
</tr>
<tr>
<td>Residual</td>
<td>0.2559</td>
</tr>
<tr>
<td>Observations</td>
<td>3,992</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>867.94</td>
</tr>
<tr>
<td>Akaike Inf. Crit.</td>
<td>-1,683.88</td>
</tr>
<tr>
<td>Bayesian Inf. Crit.</td>
<td>-1,520.29</td>
</tr>
</tbody>
</table>

*Note:* *p < 0.05 **p < 0.01 ***p < 0.001

Table 2.—Estimates and standard errors from Model 1 fit to the data of each country, separately.

### Regression Results (Model 2)

**Dependent variable:**

<table>
<thead>
<tr>
<th></th>
<th>WTP / Reference Price</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(US)</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept(Safe, No Promise)</td>
<td>0.94</td>
</tr>
<tr>
<td>Vendor Personal Favorite</td>
<td>-0.04</td>
</tr>
<tr>
<td>Vendor Friend’s Favorite</td>
<td>-0.13</td>
</tr>
<tr>
<td>Vendor Unknown Ranked</td>
<td>-0.30</td>
</tr>
<tr>
<td>Vendor Unknown Unranked</td>
<td>-0.49</td>
</tr>
<tr>
<td>Promise</td>
<td>0.01</td>
</tr>
<tr>
<td>Third Party</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 3. The country level estimates for the random effect coefficients with standard deviations.