The Role of Trust in Successful Ecommerce Websites in China: Field Observations and Experimental Studies

Eric K. Clemons
University of Pennsylvania

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

Josh Wilson
University of Pennsylvania
joshsw299@yahoo.com

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

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

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

Noisian KOH
Singapore Management University
noisian.koh.2006@phdis.smu.edu.sg

ABSTRACT

The ecommerce market in China is both the largest online market in the world and the market with the greatest number of low quality or counterfeit product offerings. We examine three very successful online companies in China and their very different paths to success and then use our experience to develop testable hypotheses, as proposed by Eisenhardt and Graebner. We present our hypotheses and then discuss the experiments that were conducted in China, the United States, Germany, and Singapore in order to test these hypotheses. We expected to see that the role of reputation is critical in China, and that it is more critical in China than in ecommerce markets elsewhere. Consistent with our hypotheses, we find first that vendor reputation is indeed the most important factor influencing consumers’ willingness to shop at and willingness to pay for goods from a specific vendor in China and elsewhere. We find only limited support for our hypotheses concerning the role of risk mitigation mechanisms in general, but we do find support for the hypothesis that the Chinese online markets differ from online markets elsewhere.

1. Introduction

China has the world’s largest online market, and the online market that is growing most rapidly. Additionally, Chinese consumers face the highest degree of defective products, counterfeit products, and products that suffer from deliberate and potentially harmful or even lethal tampering. Because it is particularly difficult to assess the quality of a product purchased online, and because it is also particularly difficult to assess the quality of a new online vendor, both we and the Chinese entrepreneurs we interviewed believed that there would be unique problems associated with consumer trust in China that might interfere with the development and expansion of online commerce in China. While Chinese consumers may look at the same factors when assessing risk as online shoppers in other markets, it seemed possible that they might weight them differently, or may experience lower willingness to pay than online shoppers in other markets in the presence of the same risks. Moreover, because of the current size and high rate of growth associated with the online Chinese market, both we and the entrepreneurs we interviewed believed that understanding any unique problems associated with consumer trust in China would have significant economic value. Our concerns and the specific hypotheses we developed to test them received only limited experimental support.

We know from prior research that there are numerous factors that affect consumers’ trust in a website and willingness to purchase from a website [25]. We know also that consumers exhibit a risk-adjusted willingness to pay for goods, which varies with their perceptions of risk and with their subjective assessment of the expected quality of the merchandise they will receive [2], [45], [32], [29]. We call the reduction in willingness to pay the trust penalty. Finally, we know that the factors that consumers use when assessing riskiness, like attractiveness of a website, reputation of a seller, or strength of the seller’s offline, physical operations, are similar from market to market [23, 25, 28, 41, 50].

Theoretical results do not always translate directly from the offline to the online world. Indicators like an expensive building, a clean, well-run shop, and well-stocked shelves may be strong signals in a physical world, but their equivalents may be weak signals online, since it is far too inexpensive to create an attractive website in place of a store and far too easy to substitute a photograph of higher quality products in place of a picture of the inferior vendor’s inferior merchandise. Promises may be made binding in a physical world since the buyer can relocate and identify the seller; with known sellers, with known physical assets, the courts may be able to force sellers to honor their promises. In contrast, online promises may be meaningless. Fake products and defective products sold online can be backed up by fake promises and defective promises;
since the buyer may not know who the seller really is, or where the seller really is, the sellers’ promises may be unenforceable and therefore meaningless. Promises may not be sufficient to inform the buyer of quality even when dealing with high quality sellers. Likewise, since online reputations can be easily faked and are less than fully reliable [40] current online reputational ratings may not be sufficient to reduce or eliminate information asymmetry [7, 11, 24, 27].

We believe that for the seller’s promises to be meaningful online, something must make them binding on the seller. The entrepreneurs with whom we spoke when starting this research suggested two mechanisms that were essential to their credibility with their customers, prior reputation and the possibility of a third party backing up their promises [8]. We therefore explore the importance of perceived reputation and of assured-third party enforcement as binding mechanisms, and we assess differences across markets.

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.

Signaling can mitigate information asymmetry, when the high-quality party with private information reveals his or her type by providing signals that would be costly for low quality parties to provide [48]. Kirmani and Rao [30] provide a comprehensive taxonomy of signals of product quality to communicate product quality to buyers. As is well known, signals can be misused, especially when sending a false signal has little cost to the sender, and both information asymmetry and false signals may be greater concerns online.

2.2. Information Asymmetry in Ecommerce

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 [42] 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, 18, 26, 27, 46]. Information asymmetry can reduce customers’ willingness to purchase online or reduce the price paid for merchandise sold online [4]. This can impede the growth of online markets or in the limit lead to their total collapse [1].

Different factors attribute to trust in ecommerce during different stages of customers’ online shopping behavior [6]. Lohse and Spiller [36] argue that 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 [7]; vendor guarantees can provide a signal and boost the trust of ecommerce consumers, even in the absence of known reputation of the sellers and their products [44]. Third party assurances from online companies like buySAFE [7] can strengthen vendors’ assurances, but are not yet available in China.

2.3. Reputational Capital

Reputation is “the extent to which buyers believe a selling organization is honest and concerned about its customers” [13]. In commerce, reputational capital can potentially reduce opportunistic behavior, for the gains of doing so can’t justify the damage to their reputation [5], leading to increased consumer confidence [19]. A good reputation has annuity value [47]. Since short term misbehavior may not produce as much value for the seller as the long term value of its reputational annuity, firms with strong reputations tend to act honorably, and consumers therefore tend to trust the future behavior of firms that enjoy strong reputations.

2.4. Differences between China And other Ecommerce Markets

Several papers have studied ecommerce in China and have identified the factors that have increased trust; most find the same factors that increase trust in the United States [22, 33, 34]. 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 [16, 23, 25, 38, 43, 47]. It is known that social trust and social capital are necessary precondition for the success of companies operating in a global economy [19]; they are also necessary conditions for operating in local markets. 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.
2.5. Case Methodology

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 [14, 17]. The application of case study methodology in Management Information System is explored by Galliards and Land, Lee, and Yin [21, 31, 54]. In the spirit of Van de Ven, we respect the practical knowledge of the senior executives with whom we spoke, developing testable hypotheses from the interviews [53].

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.

- 360buy started as a chain of physical stores selling high quality home electronics, and enjoyed a strong reputation for product quality, price, and service. Because of the SARS epidemic in 2004 they shifted to online only operations, which immediately cost them 75% of their annual sales. They argue that all of the business they retained, as limited as it was, was a result of their prior trusted reputation.
- Taobao is an online marketplace, much like eBay in the United States. They had even greater reputational problems than eBay with counterfeit merchandise and low quality sellers exiting and reentering. They now force vendors to post bond; if a customer is dissatisfied, Taobao repays the customer out of the bonding funds and disciplines the seller.
- Yihaodian is an online grocery store, the third largest website in China and the fastest growing website in China, with monthly growth of 28%. Yihaodian promised customers that any defective or counterfeit product would be replaced at their own expense. While the company Yihaodian did not initially possess a reputation that made these promises credible, the founders did. The founders returned to China after successful careers in the United States, and risked their own considerable fortunes when founding Yihaodian. The investment needed to create Yihaodian made sense only if the founders intended to remain long enough to profit from their investment, which they could do only if they protected their reputation. The founders used advertising and press coverage to ensure that consumers knew this.

The case studies focus on credibility, which is incapable of being measured directly, so in order to operationalize it, we did a baseline, which has to do with a fact that reputation has to do with WTP. We developed additional hypotheses on what other than reputation could make a promise binding. We had the amorphous “reputation matters”, and we had to come up with things that came up with reputation, and things that allowed you to monetize reputation. What we got from the execs was the importance of reputation, promise, and possibly warranty, and then we tried to develop a scale of testable differences based on theory. Our cases, and their contribution to formulation of our hypotheses, is described in detail elsewhere [8].

2.6. 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, 37,]; (2) reputation of an associated physical seller [49,51]; (3) personal experience with the seller [3, 23, 28]. (4) recommendations from friends [15, 28, 40,55]; (5) online ratings and online testimonials [11, 12, 27, 34]; (6) direct promises, and (7) attractiveness of the website [25, 40]. We control for these factors by providing the same photographs to all subjects and by providing subjects with the same descriptions for conditions (2) – (6). We do not show actual URLs or websites, so the final factor (attractiveness of website) should not affect subjects’ perceptions or decisions.

The experiment extends work that has been done previously. One body of experiments asks subjects about factors that contribute to trust when shopping online, without trying to quantify the factors or their interaction. This experiment does indeed demonstrate greater concerns with sellers in China than in the US [43, 55]. Another body of experimental work tries to assess the factors that make promises credible, by stepping through claims (e.g., “We promise that the camera is new, genuine, and has a US warranty.”), argumentation (e.g., “Our promises are credible because we are bonded.”), and support (e.g., “Our bond was issued by the Prudential Insurance Company; the existence of our bond can be verified online at BuySafe.”) [29]. We are not aware of any prior work that fully duplicates our experiments by examining the interaction of seller reputation, strength of promises and assurances, and differences the interactions between these two across different national markets.
3. Hypotheses

Our discussions with executives at 360buy, Taobao, and Yihaodian suggested the following: (1) vendors’ reputational capital may be the greatest determinant of consumers’ willingness to pay; (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 assurances 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 five hypotheses, which follow.

**Hypothesis 1: Reputation matters.**

The idea that trust is directly observable through subjects’ expressed willingness to pay is central to all of our experiments. This is important 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’ willingness to pay for items from different categories of seller.

**Hypothesis 2: Sellers’ promises alone are of little incremental value without prior reputation to create trust, and promises alone are unlikely to reduce consumers’ perceptions of risk.**

The value of promises alone is directly observable through differences in subjects’ willingness to pay for products online, from a range of merchant types, with and without promises. The extent to which differences in willingness to pay 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 may be unreliable as well.

**Hypothesis 3: Sellers’ promises are more convincing if they are backed up by a third party with the power and authority to force sellers to honor their promises.**

The value of third party assurances is directly observable through differences in subjects’ willingness to pay for products online, from a range of merchant types, with and without third party assurances. Promises that are backed up by a trusted third party should be credible and should have a greater impact on customers’ willingness to pay from low quality sellers than promises alone.

**Hypothesis 4: Perceptions of riskiness will vary from country to country. Perceptions of riskiness will be higher among subjects in China, and lower among subjects in the United States.**

We expect trust to be less of a problem among consumers in Singapore, with consumers in Germany and the United States exhibiting intermediate behavior. We test this using subjects stated willingness to pay for merchandise from all vendor types under the baseline condition, with no explicit promises or third party assurances.

**Hypothesis 5: The effectiveness of individual risk reduction mechanisms will vary across countries.**

Hypothesis 5 argues that the interaction of risk reduction mechanisms and their effectiveness will differ across countries. In China, where perception of risk is higher, promises should have a more limited effect on the trust penalty of riskier vendor types than it does in other markets, while escrow and third party assurances should be more useful in reducing the trust penalty in China than in other markets.

4. Description of the Experiment

4.1. Experimental Treatments:

This experiment was conducted in university behavioral labs in four countries, and the subject pool mainly consisted of undergraduate and graduate students. The number of subjects recruited for each country is respectively: 154 (US), 226 (China), 168 (Germany), 178 (Singapore). Subjects were randomly assigned to one of the three experimental groups. The three treatments are designed as:

1. “Min”: Vendors offer no specific assurances concerning product quality or authenticity.
2. “Mid”: Vendors give explicit promises that the product is new and genuine, and can be returned for a full refund if the buyer is not fully satisfied.
3. “Max”: Vendors provide explicit promises that the product is new and genuine, and can be returned for a full refund if the buyer is not satisfied. In addition, such promises are guaranteed by a trusted third party: vendors hold escrow account with the trusted third party, and deposit a certain amount, to be used as refund for unsatisfactory products.

Each subject is exposed to 2 of the above 3 treatments, and we denote the experimental groups who views MIN and MID treatments as “Min-Mid”, “Min-Max” and “Mid-Max” treatments are similarly defined.

Each experiment page shows a photograph and a description of the product, as could be found on a trusted vendors’ site. Two reference prices were
displayed: the manufacturer’s suggested retail price (MSRP) and an average online selling price (AOP) from a highly respected seller. MSRP is found at the manufacturer’s own website; AOP is the price at which a trusted online seller like amazon.com or walmart.com lists the product.

Subjects are asked to assess their willingness to pay (WTP), for 10 different products from a same set of vendors, under two out of three treatments. Subjects could indicate that they won’t purchase a particular product and waive entering their WTP, or indicate unwillingness to purchase a product from a certain vendor, by entering 0 as WTP for this vendor.

We are aware that an experiment exposing subjects to two treatments has both advantages and disadvantages. On the one hand, it allowed us to measure directly the effect we wanted to observe, the difference in a subject’s willingness to pay under two different treatments. We were able to compute average differences across national subject pools, and directly compare differences between or among average differences. Additionally, this experimental design allowed us to minimize the difference in risk aversion in the subjects actually included in our averages. The most risk-averse subjects expressed a non-zero willingness to pay for only the least risky vendor types under weaker treatments, while the more risk-neutral subjects expressed a non-zero willingness to pay for all vendors and under all treatments. This paired treatment design allowed us to ensure that we were comparing the average willingness to pay under all treatments for the same population. In our design we calculated the average willingness to pay only from subjects who reported a non-zero willingness to pay for a product and a vendor under both treatments. A design that exposed subjects to only one treatment would have calculated the average willingness to pay under strong treatments using a larger subset, and the average willingness to pay under weak treatments using a smaller subset. The single treatment experiment would have had significant sample bias inherent in the design. On the other hand we are aware that exposing subjects to two treatments to reduce risk inherently alters the subjects’ perception of risk, and may bias their responses accordingly. We used statistical techniques in our data analysis, derived from Cole (1966) [9, 39, 51] for the repeated measure analysis of variance, to minimize the effects introduced by this mixed within group and between group design.

4.2. Vendor Types, Product Selection
And Design of Reward Function:

We created five hypothetical online vendor types with corresponding descriptions of riskiness as the basis of our experiments, based on prior research [3, 15, 23, 28, 40, 50, 52, 54]. Descriptions of these vendors included their locations and their reputations, i.e., the degree to which subjects should assume that they had experience with or knowledge of the vendors. This is described below:

1) Safe Vendor (Safe): A well-known online seller in the country where the experiment is performed, with a strong reputation, such as Amazon or a website associated with a well-known retailer, such as Walmart.

2) Favorite Vendor (Fav): A website in the country where the experiment is being performed that is assumed to be a favorite website of the subject.

3) Friend’s Favorite Vendor (Rec): A domestic website that the subject has not used before, but that is strongly recommended by close personal friends.

4) Unknown Rated Vendor (UR): A domestic website that neither the subject nor his friends have used before, but that enjoys strong online ratings.

5) Unknown Unrated Vendor (UU): A domestic website that neither the subject nor his friends have used before, and that does not have an online rating.

Products used in our experiments include the following categories: sporting equipment, high and low end consumer electronics, cosmetics, clothing, event ticket and food. While the categories for all four countries remained constant, specific products are adjusted for each country, taking into consideration their comparability across countries in price, appeal to consumers, and risk of being counterfeit. For example, American and German students were offered an opportunity to buy local brands of coffee, while in China and Singapore they were offered tea. We conducted several rounds of pretests until we had tailored the product mix appropriately, to appeal to the subject population in each of our four markets.

A reward function was included in the experiment to encourage realistic behavior from subjects. The subjects were informed that they could receive a bonus cash payment that would be calculated based upon their performance. The subjects were not given the details of how the reward function worked before the experiment; we did not want to provide enough detail for the subjects to attempt to game the reward function. We told them that they would be rewarded by how satisfied they would be had they actually purchased what they said they wanted at the prices that they had been willing to pay. Satisfaction was determined by estimating the quality of the product that we expected they would have received from the vendors they bought from and the prices that they had been willing to pay. We used data from the pretests to calibrate the reward function for each market so that the maximum bonus paid was similar across all test sites. Our German and Chinese authors believed that participants needed extra
motivation to participate in an experiment of this length. In hindsight, it might not have been necessary, and may have distorted the behavior of German and Chinese subjects, the very subjects from whom it was supposed to ensure rational behavior.

All of the test sites screened subjects on the basis of online shopping experience, and the resulting sample of subjects included only those subjects with this experience.

5. Data Cleaning

Our experiment was designed so that we were able to observe differences in paired willingness to pay assessments so that we could estimate the effect of differences in treatments in subjects’ differences in willingness to pay. Importantly, this allowed us to control for selection bias; we dropped all data from subjects who were not willing to purchase from more risky vendors under more risky conditions, ensuring that all data came from comparably risk averse subjects. If we had not paired our treatments, but had offered subjects the opportunity to purchase from vendors under only one treatment, we would have obtained data from more subjects under the stronger treatment than the weaker treatment. Our paired treatment design largely eliminated the effects of selection bias on our data set. We found some anomalies, caused by inversions, when subjects expressed greater willingness to pay under weaker treatments than their willingness to pay for the same product from the same merchant under stronger conditions. Most of the inversions came from a small set of subjects who consistently inverted for every product they decided to purchase. We removed all of the data related to this small set of inverters. For the remaining inversions, we deleted only the data entered by a subject for the product where the inversion occurred. This does not mean that an experimenter interested in behavioral economics should ignore inversions; it might indeed be interesting to understand the behavior of subjects who do not appear to behave rationally or who place negative value on assurances from sellers. But this was a small data set, and it is of less interest when designing websites to appeal to the largest number of online shoppers.

6. Visual Inspection and Preliminary Assessment of our Hypotheses

Below we offer a tentative assessment of the five hypotheses through visual inspection of the data. In the exploration of the hypotheses we also offer behavioral justification, that is, beliefs and behaviors among subjects might have produced the effects observed in the experiments. We are of course aware that correlations are not the same as causality, and that hypothetical explanations of a purported causal mechanism are more suspect still. These explanations are provided because we will attempt to explore them in follow-up work, with longer attitudinal surveys and longer exit interviews than were included in the current experimental design.

Hypothesis 1 (Reputation matters) appears to be supported. Reputation matters and affects willingness to pay. Figure 1 shows willingness to pay, ranged from safest vendor to riskiest vendor, with safest vendor normalized to one in all cases. Willingness to pay is monotonically decreasing in vendor reputation in all four subject populations; visual inspection is confirmed by an F-Test to determine if vendor type affects willingness to pay. We used repeated measures analysis of variance, and ran the following test: WTP = α + β * vendor + γ * condition + δ * product + ε. (i = China, US, Germany, Singapore); H0: β = 0; H1: β ≠ 0. Here “condition” refers to the three assurance designs (MIN, MID, or MAX). Vendor type’s influence on WTP (the vector was significant at a p-value less than 0.0001 in all four countries. See Table 1 in the appendix.

Support for Hypothesis 2 (Promises alone are of little incremental value) appears to be absent when examining the green curves in figures 2 a-d; visual inspection suggests that the value of promises is positive. We calculated difference in WTP in two treatments for each individual, and conducted the following test, using General Linear Model Method:

\[ \text{DiffWTP} = \alpha + \beta \times \text{vendor} + \gamma_1 \times \text{minmax} + \gamma_2 \times \text{midmax} + \delta \times \text{product} + \varepsilon; \]

H0: \( \alpha = 0 \); H1: \( \alpha \neq 0 \) (The base group is min-mid treatment). Analysis shows that all four curves are different from zero and that differences are statistically significant. However, we also used two additional tests to determine if the power of promises, as shown in the green curve, might be an experimental artifact, caused by presenting subjects with two treatments at a time, highlighting the potential power of promises. First, we examined the difference between the blue and the red curves. Algebraically, the
difference between the blue and red curves should be the same as the green curve, except that if all three curves suffer from bias introduced by the experimental design, then subtracting the two should allow us to assess the role of promises without distortion introduced by bias. DiffWTP = \alpha + \beta \cdot \text{vendor} + y_1 \cdot \text{minmax} + y_2 \cdot \text{minmid} + \delta \cdot \text{product} + \epsilon; H0: y_1 - y_2 = 0 H1: y_1 - y_2 \neq 0. Viewed this way, promises alone appear significant in China and Singapore, not significant in the US, and only weakly significant in Germany. A third analysis compared willingness to pay without promises in the MIN to MAX treatment with willingness to pay with promises in the MID to MAX treatment; the idea was to compare willingness to pay of no promises to willingness to pay with promises, under conditions that anchored both as the weaker treatment. WTP = \alpha + \beta \cdot \text{vendor} + \gamma \cdot \text{assurance} + \delta \cdot \text{product} + \epsilon; H0: \gamma = 0 H1: \gamma \neq 0. Under these analyses we see that assurances offer significant value in Singapore, but not in the other three markets. These analyses together suggest that promises alone are significant in Singapore, but not in the US or Germany. Unfortunately, the Chinese data are not internally consistent. See Table 2a-c.

Support for Hypothesis 3 (Sellers promises are more convincing if backed up by a third party with power and authority) would appear to be mixed when examining only the blue curves in figures 2 a-d; this is supported by statistics: DiffWTP = \alpha + \beta \cdot \text{vendor} + y_1 \cdot \text{minmax} + y_2 \cdot \text{minmid} + \delta \cdot \text{product} + \epsilon; H0: \alpha = 0 H1: \alpha \neq 0 (The base group is mid-max treatment). Results are shown in Table 3.

Third party assurances do not appear necessary in Singapore and they add little to the value of a promise. Third party assurances do add value in the US, Germany, and China, again shown by examining the blue line. The blue line in Singapore is, however, statistically significantly different from zero, suggesting the possibility of bias caused by offering subjects pairwise treatments. Once again we used two additional tests to determine if the power of third party assurances, as shown in the blue curve, might be an experimental artifact, caused by presenting subjects with two treatments at a time, highlighting the potential power of third party assurances.

First, we examined the difference between the blue and the red curves. DiffWTP = \alpha + \beta \cdot \text{vendor} + y_1 \cdot \text{minmax} + y_2 \cdot \text{minmid} + \delta \cdot \text{product} + \epsilon. H0: y_1 - y_2 = 0 H1: y_1 - y_2 \neq 0. Algebraically, the difference between the red and green curves should be the same as the blue curve, except that if all three curves suffer from bias introduced by the experimental design, then subtracting the two should allow us to assess the role of assurances without distortion introduced by bias. Viewed this way, assurances appear significant in US and Germany, less significant in Singapore, and not significant in China. A third analysis compared willingness to pay without assurances in the MIN to MID treatment with willingness to pay with assurances in the MIN to MAX treatment; the idea in this case was to compare willingness to pay with promises and without promises, using treatments in which we were comparing average willingness to pay data under treatments where both data sets were reported as the stronger of two trust conditions so that the effect of bias would be reduced. WTP = \alpha + \beta \cdot \text{vendor} + \gamma \cdot \text{condition} + \delta \cdot \text{product} + \epsilon; H0: \gamma = 0 H1: \gamma \neq 0. Under these analyses we see that assurances offer significant value in China, have limited impact in the US, and offer little value in Germany or Singapore. Support from Germany is thus limited. Once again, the data from China are inconsistent. Support for hypothesis 3 is limited.

![Graphs showing the difference in WTP for US, Germany, and China](image)
om A-B testing conducted similarly, we know from our markets: Price premiums F., Wang, Y., and Wilson, Objectively, testing to which argues that different experimental design, was corrected for biases, but not in Singapore. Support for hypothesis 2 was present in the US and Germany when analyses were supported.  Hypothesis 5, which had a value of 48.87 is significant, with p < .0001. In treatment is significant, with 6 degrees of freedom, the see if the cross-term dummy variable for country and explain, but is demonstrably behavior of Chinese subjects once again is difficult to enforce promises. (2) The US and Germany value promises differently from Singapore, treating them as implicit from high quality vendors and strongly value adding from weaker vendors, once again consistent with trust in the legal system. (3) Singapore places little value on third party assurances, as if they are unnecessary in the presence of promises and legal enforcement of them. (4) The US and Germany value third party assurances more than Singapore. (5) The behavior of Chinese subjects once again is difficult to explain, but is demonstrably different from the behavior of consumers in other markets. Objectively, testing to see if the cross-term dummy variable for country and treatment is significant, with 6 degrees of freedom, the F value of 48.87 is significant, with p < .0001. In summary, hypothesis 1, which follows almost tautologically from the experimental design, was supported. Hypothesis 5, which argues that different treatments will have different effects in different markets, appears supported. Support for hypothesis 2 was present in the US and Germany when analyses corrected for biases, but not in Singapore. Support for hypotheses 3 and 4 was mixed.

7. Next Steps

Our next steps will be designed to address the limitations listed above:

- Lack of support for hypothesis 3 is particularly confusing. We know from A-B testing conducted by buySAFE that third party assurances do have significant value in the U.S., as measured by actual purchase price [7]. Similarly, we know from our interviews with Taobao that credible third party assurances have significant value in China, again as measured by actual purchase price [8]. We do not yet know why we are not seeing this in the experiments.
- We are designing live experiments to assess revealed preferences rather than expressed preferences and have begun discussions with important Chinese websites.
- We will repeat the experiments and conduct more detailed exit interviews and surveys in an attempt to understand subjects understanding of the experiments and subjects motivations for their behavior. This would help us assess whether Chinese behavior is best explained by anchoring, by limited understanding of the experiment, by limited experience with third party assurances in ecommerce, or by some other effect. It would help us understand if our behavioral assumptions about the behavior of Singapore subjects are correct or not.

8. References

(1996), 511-535.
[54] Yin, R.K. Research design issues in using the case study method to study management information systems. The information systems research challenge: qualitative research methods, 1 (1989), 1-6.

APPENDIX

TABLE 1. Repeated Measures Analysis of Variance on Vendor Type’s Effect on WTP

<table>
<thead>
<tr>
<th>Factor</th>
<th>US</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF F Value</td>
<td>F Value</td>
<td>F Value</td>
</tr>
<tr>
<td>assurance</td>
<td>2</td>
<td>1.46</td>
</tr>
<tr>
<td>product</td>
<td>9</td>
<td>6.17***</td>
</tr>
<tr>
<td>vendor assurance</td>
<td>10</td>
<td>1.35</td>
</tr>
<tr>
<td>vendor product</td>
<td>45</td>
<td>3.20***</td>
</tr>
<tr>
<td>assurance product</td>
<td>18</td>
<td>0.22</td>
</tr>
<tr>
<td>vendor assurance* product (Continued)</td>
<td>90</td>
<td>0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor</th>
<th>DF F Value</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>0.0506</td>
<td>5.58</td>
</tr>
<tr>
<td>US</td>
<td>0.1018</td>
<td>6.46</td>
</tr>
<tr>
<td>Germany</td>
<td>0.0916</td>
<td>8.67</td>
</tr>
<tr>
<td>Singapore</td>
<td>0.0551</td>
<td>2.76</td>
</tr>
</tbody>
</table>

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

TABLE 2. Testing Power of Promises

Table 2a. Test Results for Whether Difference Between Min and Mid Treatments is Zero

<table>
<thead>
<tr>
<th>Min-Mid</th>
<th>Coefficient</th>
<th>T</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>0.00957</td>
<td>4.11</td>
<td>0.018</td>
</tr>
<tr>
<td>US</td>
<td>0.00147</td>
<td>0.27</td>
<td>0.7719</td>
</tr>
<tr>
<td>Germany</td>
<td>0.00753</td>
<td>2.41</td>
<td>0.0237</td>
</tr>
<tr>
<td>Singapore</td>
<td>0.08855</td>
<td>13.03</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 2b. Test Result for Whether the Differences in WTP for the Min-Max and Mid-Max Treatments Are Different from Each Other

<table>
<thead>
<tr>
<th>(Min-Max)-(Mid-Max)</th>
<th>Difference</th>
<th>T-VALUE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>0.00520</td>
<td>2.13***</td>
<td>0.0528</td>
</tr>
<tr>
<td>US</td>
<td>0.01646</td>
<td>3.31***</td>
<td>0.0056</td>
</tr>
<tr>
<td>Germany</td>
<td>0.01785</td>
<td>5.80***</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Singapore</td>
<td>0.02000</td>
<td>2.67**</td>
<td>0.01926</td>
</tr>
</tbody>
</table>

TABLE 3. Testing Power of 3rd Party Assurances

Table 3a. Test Results for Whether Difference Between Mid and Max Treatment is Zero

<table>
<thead>
<tr>
<th>(Min-Max)-(Mid-Mid)</th>
<th>Coefficient</th>
<th>T</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>0.23</td>
<td>0.6300</td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>0.03</td>
<td>0.8689</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>0.49</td>
<td>0.4834</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>4.43</td>
<td>0.0125</td>
<td></td>
</tr>
</tbody>
</table>

Table 3b. Test Results for Whether the Difference in WTP for the Min-Mid and Max-Mid Treatments Are Different from Each Other

<table>
<thead>
<tr>
<th>(Min-Max)-(Min-Mid)</th>
<th>Difference</th>
<th>T-VALUE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>0.00964</td>
<td>-0.136</td>
<td>0.1748</td>
</tr>
<tr>
<td>GER-CHN</td>
<td>0.02659</td>
<td>4.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SGP-CHN</td>
<td>0.02232</td>
<td>2.50</td>
<td>0.0126</td>
</tr>
<tr>
<td>GER-US</td>
<td>0.03623</td>
<td>5.08</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>GER-SGP</td>
<td>0.00427</td>
<td>0.48</td>
<td>0.6338</td>
</tr>
<tr>
<td>US-SGP</td>
<td>-0.03195</td>
<td>-3.29</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

TABLE 4. Testing Whether WTP Values under No Promise Condition Differ Across Countries

| Parameter | Estimate | t Value | Pr > | |z| |
|-----------|----------|---------|------|---------|
| US-CHN    | -0.00964 | -1.36   | 0.1748 |
| GER-CHN   | 0.02659  | 4.42    | <0.0001 |
| SGP-CHN   | 0.02232  | 2.50    | 0.0126  |
| GER-US    | 0.03623  | 5.08    | <0.0001 |
| GER-SGP   | 0.00427  | 0.48    | 0.6338  |
| US-SGP    | -0.03195 | -3.29   | 0.0010  |