Doctoral Dissertation

Essays on Technology-Enabled Platforms

submitted to the
David A. Tepper School of Business
in partial fulfillment for the requirements for the degree of

Doctor of Philosophy
in Business Administration (Information Systems)

Tat Koon Koh
Carnegie Mellon University
Tepper School of Business
Pittsburgh, Pennsylvania 15213

Dissertation Committee:

Mark Fichman (Chair)
Robert E. Kraut
Tridas Mukhopadhyay
Steven Dow

November 2012

Table of Contents

CHAPTER 1:
OVERVIEW

This dissertation consists of four studies that examine firms' usage of online platforms. In the first study, we examine trust formation and development in global buyer-supplier relationships. Trust affects all business relationships, especially global business-to-business (B2B) transactions due to the distances between buyers and suppliers. We use information signaling theory to examine how information indices and signals affect buyers' trust in suppliers in global B2B commerce. Specifically, we look at how buyers' trust is affected by (1) their perceptions of the national integrity and legal structure of suppliers' country, and (2) third-party verifications of suppliers on B2B exchanges. Because buyer-supplier relationships usually evolve over time, we study how the effects of indices and signals change as the number of transactions between the partners increases. A survey of global organizational buyers finds that perceptions of national integrity, legal structure, and supplier verifications are all positively related to buyers' trust. However, the number of prior transactions between buyers and suppliers moderates the impact of perceived legal structure on buyers' trust.

In the second study, we examine how selling and buying activity levels on B2B exchanges drive multi-homing buyers' preferences for exchanges. With the proliferation of B2B exchanges, many firms are multi-homing or using various competing platforms concurrently. Using a unique dataset of 118 buyers' participations in two B2B exchanges over seven months, we find that buyers' preferences are positively associated with selling levels on the platforms. However, buyers' preferences are non-monotonically related to buying levels on the platform. At low levels, an increase in buying level has a positive effect on buyers' preferences. This effect may derive from the principle of social proof, where individual buyers observe and imitate other similar buyers' behaviors. As buying level increases, there is greater competition among buyers on the platform, causing buyers to participate more on the other exchange. We also find that the impacts of buying levels on buyers' preferences attenuate over time. Our results highlight the need to correctly model buyers' homing behavior, and show how market factors and social information conveyed by users on the platforms affects individual buyers' participation and the competition between B2B exchanges.

In the last two studies, we look at how firms can acquire impactful ad designs through crowd-based design contest platforms. Design contests allow firms to acquire a large number of designs that they can consider for use in advertising campaigns. However, the large number of entries brings along a challenge in measuring design distinctiveness: The number of pairwise comparisons that is needed to determine distinctiveness increases at a quadratic rate with the number of designs and may be non-trivial. To tackle this problem, we develop a novel model-based approach to efficiently measure design distinctive in design contests. We also find that ads with more distinctive design achieve more click-through than those with less distinctive designs.

Given that design distinctiveness matters in online advertising campaign, we investigate how firms can influence designers to produce more distinctive work in design contests. Firms often provide examples of ad designs that they like in design projects. Using a randomized design contest experiment that involved experienced graphic designers, we look at how examples provided by the firms influence creative processes and design outcomes in design contests. Specifically, we examine how the number, quality, and design variability of these examples affect designers' exploration for design concepts and their design submissions in the contests. We also look at how the characteristics of designers' exploration and work relate to design distinctiveness.

Chapter 2: Buyer-Supplier Trust in Global Business-to-Business E-Commerce

2.1 Introduction

This paper examines trust formation and development in e-commerce transactions between buyers and suppliers from different countries. In existing research on trust in transactional exchanges, trading partners are generally based in the same country. Yet globalization is changing who and where the trading partners are. The world is "flattening" (Friedman, 2005) as technology drives and facilitates the globalizing of culture and markets. Globalization involves more than reducing technical barriers and transaction costs; it also requires human interaction across cultures and national practices. Even with enabling technology and lower transaction costs, global trade still requires two or more people or firms to interact to develop a cooperative venture across borders. We look at a prototypic example of globalization facilitated by information technology – business-to-business (B2B) e-commerce – where buyers and suppliers across the globe exchange goods and services using information systems, such as online exchanges. These exchanges aggregate, make, and facilitate markets (Bakos, 1998; Dai & Kauffman, 2002), and help firms bypass traditional distribution channels and extend their reach globally (Senn, 2000). We study how the perceived country and firm characteristics of trading partners influence trust formation and development on online exchanges.

Buyers' trust in suppliers is critical in all commerce, but particularly e-commerce, due to more pronounced information asymmetry, where buyers have incomplete information about suppliers. Under these conditions, buyers risk selecting incompetent or opportunistic suppliers. This risk impedes transactions between buyers and suppliers. (The risk is not one-sided because suppliers also face problems such as non-payments by buyers.) In such situations, trust is "an important lubricant" for economic exchanges to take place (Arrow, 1974, p. 23). The separation in time and space between buyers and suppliers in cross-border e-commerce raises the risks associated with information asymmetry and, consequently, increases the value of trust.

One way to increase buyers' trust when there is information asymmetry is to close the information gap. In markets suffering from information asymmetries, buyers can gather information about suppliers (Eisenhardt, 1989), and suppliers could reassure potential buyers of their abilities and intentions by providing credible information to reduce the asymmetries (Nayyar, 1990). There are two types of information suppliers can provide: indices and signals (Spence, 1973). Indices are supplier attributes that are inherently fixed or difficult to alter (e.g., the suppliers' country of origin). In contrast, signals are characteristics that suppliers can more easily invest in or acquire (e.g., web-seals that the suppliers can buy). The aphorism "talk is cheap" captures the idea that signals must go beyond the supplier saying "You can trust me" to be credible; effective signals must create a separating equilibrium (Boulding & Kirmani, 1993) where it is costly for untrustworthy suppliers to acquire the signals. Collectively, indices and signals influence buyers' trust in suppliers only if they are costly to change (indices) or acquire (signals).

Using Spence's (1973) distinction, we examine how information indices and signals influence e-commerce buyers' trust in suppliers. First, how do suppliers' country characteristics influence buyers' trust? Individual suppliers cannot easily change their country's reputation, which is based on public opinion and the behavior of many other suppliers. While a supplier could disguise its country of origin or relocate to another country, such moves are costly, challenging, and potentially disruptive to its business operations or existing industry ties, especially for small and medium-sized enterprises. Hence, we consider suppliers' country of origin as an information index. In particular, we focus on the extent to which buyers' trust is affected by their perceptions of the national integrity and legal structure of the suppliers' country.

Second, how do signals acquired by suppliers affect buyers' trust, particularly in B2B exchanges? Many exchanges provide tools such as feedback mechanisms to help trading partners evaluate each other's trustworthiness (Ba & Pavlou, 2002; Bolton, Katok, & Ockenfels, 2004; Pavlou, 2002; Pavlou & Gefen, 2004). Some B2B exchanges also offer verification or web-seal services that suppliers can use to verify information that they provide. These services help to assure buyers that the information about verified suppliers' is authentic. Do such services increase buyers' trust? Previous studies that examine how web-seals affect buyers' trust h inconclusive results. In this study, we propose conditions for web-seals to be effective trust-building mechanisms.

Third, when do country attributes and web seals have more or less influence on buyer trust? Specifically, could the effectiveness of perceived country attributes and supplier verification be moderated by past transactions between buyers and suppliers? Examining this question complements previous research that looks at initial trust formation in e-commerce (e.g., Lim, Sia, Lee, & Benbasat, 2006; McKnight, Choudhury, & Kacmar, 2002b; Stewart, 2003). Surprisingly, we show that buyers' own experiences with suppliers do not necessarily diminish the value of trust-enhancing indices and signals.

This study makes three contributions to trust research. First, by examining how both information indices and signals serve as antecedents of trust, this study provides a more complete and comprehensive understanding of the role that signaling theory plays in trust development than past research does. Spence's (1973) theory is about information indices as much as it is about information signals. However, existing trust research that applyies signaling theory focuses only on information signals, such as sellers' reputation (e.g., Pavlou & Dimoka, 2006) or competitors' prices (e.g., Trifts & Häubl, 2003). In contrast, information indices as antecedents of trust have received little attention.

Second, we investigate antecedents of trust that are salient in cross-border, global B2B e-commerce contexts. This focus contrasts with most existing research that focuses on localized e-commerce where buyers and sellers are in the same country (e.g., Balasubramanian, Konana, & Menon, 2003; Gefen, Karahanna, & Straub, 2003; Lim et al., 2006; McKnight et al., 2002b; Pavlou & Dimoka, 2006; Stewart, 2003; Sun, 2010). Localized and globalized transactions have different implications for the formation and impact of trust. In localized e-commerce, trading partners in the same country share common knowledge about cultural and legal structures, ways to enforce contracts, and access to legal recourse if transactions fail. Moreover, trading partners can more easily gather information about each other's competencies and reputation. Such conditions, which facilitate trust formation, are more rare in globalized e-commerce, making it more difficult to establish trust in this context. Furthermore, buyers' trust in suppliers during cross-border transactions may be influenced by factors that are otherwise not salient in localized e-commerce contexts. Our focus on such transactions fills an important gap in our knowledge of online trust.

Finally, we examine the moderating role of past transactions on buyer-supplier trust. When assessing suppliers' trustworthiness in the early stages, buyers draw inferences from various sources, including the characteristics of the suppliers' country. Repeated transactions help buyers to accumulate knowledge about their suppliers, and may reduce the influence of country characteristics on buyers' trust. We find that past transactions do mitigate some, but not all, impacts of country characteristics on buyers' trust. This suggests that although suppliers can build up trust through repeated transactions, factors that are outside their direct control also affect buyers' trust in them. This finding has important implications for firms and policy makers.

2.2 Trust

We define buyers' trust in a supplier as the buyers' willingness "to accept vulnerability based upon positive expectations of the intention or behavior of" the supplier (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 295). Mayer, Davis, and Schoorman (1995) identify three components of trustworthiness: ability, benevolence, and integrity. Ability is the supplier's skills and competencies in meeting the buyer's needs. Ability is thus context-dependent. In the case of B2B transactions, the buyer would focus on the supplier's ability to satisfy their purchase requirements such as quality, timeliness, and cost. Benevolence is the overall goodwill of the supplier towards the buyer. A benevolent supplier would not behave opportunistically (in Williamson's (1975) sense of opportunism) towards the buyer for their own benefit. Rather, the supplier is concerned for the buyer's well-being. Integrity is the supplier's adherence to principles (e.g., being honest and fair) that are acceptable to the buyer. The supplier's integrity is judged by the consistency in their behaviors, the credibility of their communication, and their commitment to justice and fairness (Mayer et al., 1995).

We focus on interorganizational trust in this study. This trust can emerge from prior history and expectations of continued relations between the buyers and suppliers (Poppo, Zhou, & Ryu, 2008). Past interactions provide opportunities for partners to build knowledge about each other (Koehn, 2003; Ratnasingam, 2005) and affect their satisfaction and trust in each other

(Kwon & Suh, 2004; Selnes, 1998). Expecting continuity in a relationship improves buyer-supplier trust by extending the time horizon for mutual benefits and discouraging opportunistic short-term gains (Aulakh, Kotabe, & Sahay, 1996; Poppo et al., 2008).

Interorganizational trust affects the organization and coordination of economic activities (McEvily, Perrone, & Zaheer, 2003). It affects transaction costs (Chiles & McMackin, 1996), governance choice, exchange performance (Gulati & Nickerson, 2008), information sharing (Dyer & Chu, 2003), and negotiation and conflict (Zaheer, McEvily, & Perrone, 1998) between the buyer and supplier. Interorganizational trust also has positive transactional effects. Buyers' trust in suppliers is positively related to the buyers' anticipated future interaction with the suppliers (Doney & Cannon, 1997; Pavlou, 2002) and increases their commitment to and cooperation with their suppliers (Morgan & Hunt, 1994). Buyers also allocate a higher share of their business to suppliers whom they trust (Doney, Barry, & Abratt, 2007).

2.3 Antecedents of Buyers' Trust

In most interorganizational transactions, buyers are concerned with micro, supplier-level characteristics, such as the level of specific investments made by suppliers (Heide & John, 1990) or the suppliers' customer orientation (Doney et al., 2007). However, in international sourcing, buyers are also concerned with macro, country-level characteristics, such as the political environment, business practices, and regulations in the supplier's country (Birou & Fawcett, 1993; Min, 1994; Min & Galle, 1993). Therefore, it is important to examine both micro (supplier-level) and macro (country-level) factors in global B2B commerce.

Signaling theory (Spence, 1973) provides a framework to holistically examine how both micro and macro factors affect buyers' trust in suppliers. Information signals encompass micro, supplier-level factors (since suppliers can manipulate their individual characteristics), whereas information indices encompass macro, country-level factors (since individual suppliers cannot change these factors at their own discretion).

2.3.1 Information Indices: Perceived National Integrity and Legal Structure

Information indices are observable, fixed, relatively unalterable attributes of an individual, such as race or nationality (Spence, 1973). Earlier research has shown how foreign partners' nationality – and the corresponding value systems, cultural traits, and institutions – affects others' prior expectations about their behaviors (Ariño, de la Torre, & Ring, 2001). Country of origin influences evaluations of products, people, and firms (Bilkey & Nes, 1982; Madon et al., 2001; Zaheer & Zaheer, 2006), and firms from countries that are viewed as untrustworthy may be perceived as untrustworthy (Zaheer & Zaheer, 2006).

Given the forces of globalization and proliferation of online B2B exchanges, buyers are increasingly exposed to prospective suppliers from different countries with different business practices and orientations (e.g., Ariño et al., 2001; Hofstede, 1980; Xiao & Tsui, 2007). In the context of global B2B e-commerce, two indices that are associated with suppliers' country of origin may be especially important: national integrity and legal structure. National integrity is the extent to which typical actors in a particular country are presumed to adhere to some set(s) of moral or ethical principles in their actions (e.g., fairness and honesty towards others). Legal structure broadly refers to the rules and regulations in a country that govern relationships between entities (e.g., individuals, firms, organizations). These indices respectively relate to the social and formal conditions in partners' countries, which are important considerations in cross-border relationships such as international joint ventures (e.g., Holton, 1989; Luo, 2007) and trade (Birou & Fawcett, 1993). Because individual suppliers cannot easily alter societal norms or modify business regulations on their own, national integrity and legal structure are informative and should affect buyers' beliefs about suppliers (Spence, 1973). Specifically, perceptions of social and formal norms in a supplier's country help to reduce the information asymmetry about the supplier's behaviors and shape a buyer's expectations about the supplier's trustworthiness (e.g., Bachmann, 2001; Bradach & Eccles, 1989; Zaheer & Zaheer, 2006)[1].

[1] There are other considerations influencing cross-border relationships, such as exchange rate fluctuations, and logistics support for longer supply lines (Birou & Fawcett, 1993). While these factors affect buyers' choice of suppliers, they are not relevant to buyer-supplier trust. Buyer can hedge exchange rate risk in futures markets, and use insurance markets and global logistic companies to help handle international logistics. Presently, we do not have sufficient theoretical rationale to explore whether and how these factors affect suppliers' competency, benevolence,

Our concepts of national integrity and legal structure have parallels with the notions of situational normality and structural assurance, respectively. These latter notions have often been used to explain online trust (e.g., Chau, Hu, Lee, & Au, 2007; Gefen et al., 2003; McKnight, Choudhury, & Kacmar, 2002a, 2002b, 2004; Ou & Sia, 2010). Situational normality is the belief that the Internet environment is in proper order and success in online transaction is likely because the situation is normal or favorable. For example, buyers have higher trust in a retailing website when the nature of interaction with the website is typical of other similar websites (Gefen et al., 2003; Ou & Sia, 2010). Structural assurance refers to the belief that "structures like guarantees, regulations, promises, legal recourse, or other procedures are in place to promote success" in e-commerce transactions (McKnight et al., 2002a, p. 339). An online store with sufficient encryption and security capabilities, for instance, is perceived to provide a secured transaction environment, which improves consumers' trust and purchase intention (Chau et al., 2007; Ou & Sia, 2010). However, there is a key distinction between situational normality and structural assurance on one hand and national integrity and legal structure on the other. Situational normality and structural assurance pertain more to the channels in which the online transaction occurs, whereas national integrity and legal structure concern the environment the trading partners are in. Consider a buyer who finds suppliers from different countries on an online exchange. The buyer's situational normality and structural assurance beliefs about the exchange do not vary by suppliers – their perception about whether the exchange is a favorable and safe channel is the same for all suppliers. However, the buyer's perceptions of national integrity and legal structure in different countries are likely to vary and affect the buyer's relationships with suppliers at the dyadic level on the exchange. This is the aspect of situational perceptions that we look at in this study. Because the context of most research in online trust is localized e-commerce (where buyers and sellers are from the same country), the relationship between buyers' trust and their perceptions of foreign suppliers' country has received little attention. Yet given how e-commerce facilitates international trade in today's economy, understanding this relationship is important.

Perceived National Integrity. Most studies of integrity concentrate on how individual actors' integrity affects the trust others place in them. Trust in trading partners can also be related

---

and/or integrity, or buyer-supplier trust at dyadic levels. Hence, we focus on buyers' perceptions of national integrity and legal structure in suppliers' country in this study.

to the perceived level of integrity in the society (i.e., national integrity) to which they belong (Fukuyama, 1995; Mackie, 2001). As noted above, perception of national integrity relates to perceived social norms in suppliers' country. When buyers perceive that norms in a supplier's country encourage positive behaviors such as cooperation or honesty, they expect the supplier to adhere to these norms. Furthermore, societal norms can act as a powerful form of social capital that inhibits deviant actions (Coleman, 1988; Doney, Cannon, & Mullen, 1998). The higher the national integrity of a country, the less likely any particular supplier will be to commit deviant actions that would sully its reputation. Conversely, in a country with lower level of national integrity, deviant behaviors may be more accepted or tolerated.

Thus, perceived national integrity in a supplier's country provides information about expected supplier behavior, and shapes a buyer's beliefs about the moral character of typical suppliers in that country. These expectations and beliefs, in turn, affect the buyer's cognition-based trust in a supplier's reliability and dependability (McAllister, 1995). Suppliers in countries with higher perceived national integrity may be seen to be more likely to adhere to moral or ethical norms and show individual integrity. The country's norms also deter deviant supplier behavior. Therefore, buyers are likely to trust suppliers in countries with higher national integrity.

> *H1: The perceived level of national integrity in the supplier's country is positively related to the buyer's trust in the supplier.*

Perceived Legal Structure. Trust in transactional relationships also depends on stable legal, political, and social institutions (Lane & Bachmann, 1996). As an economy moves from local to national markets, transactions span longer social and geographical distances, which requires institutional, formal trust (Zucker, 1986). Extending this line of argument, we expect institution-based trust to play a significant role when transactions take place in international markets.

The legal structure of a supplier's country provides information about the formal norms in the suppliers' country, and shapes buyers' expectations of suppliers' behaviors in two ways. First, institutional rules and regulations in a country affect various facets of business operations and the types of firms that can operate. For example, when a country has formalized licensing policies that govern businesses formation and operation, opportunities for those that do not meet the requirements to operate are reduced. A buyer may thus expect suppliers from countries with effective business laws and regulations to be more competent. This is consistent with Zucker's finding that the "emergence of licensing standards … increased the certainty of performance characteristics" (1986, p. 94). Licensing provides the buyer some assurance of a licensee's ability to fulfill their purchase requirements.

Second, a country's legal structure affects the extent to which contracts are enforceable, which provides effective legal recourse when disputes arise. Contract laws are broad societal guarantees needed by buyers and suppliers. The availability and effectiveness of these formal mechanisms are important to foreign buyers since trade disputes are more likely to occur given the greater separation in time and space of cross-boundary transactions. Furthermore, more market-oriented societies with more non-familial/tribal transactions have developed institutions to punish those who are not fair and trustworthy (Henrich et al., 2010). Suppliers may be deterred from behaving opportunistically or dishonestly when such legal mechanisms are in place and enforced. A supplier that operates in such an environment could be expected to be more benevolent and ethical.

Therefore, perceptions about legal structure affect expectations about the types of market participants that one is likely to encounter. A buyer may expect a supplier from a country with strong legal structure to be more trustworthy. With stronger legal structures, "undesirable entities" (i.e., those with low ability, integrity, and/or benevolence) are also expected to self-select out from participating in the market given their inability to meet legal requirements or concern for legal penalties for misbehavior.

H2: The perceived level of legal structure in the supplier's country is positively related to the buyer's trust in the supplier.

2.3.2 Information Signals: Supplier Verifications and Web Seals

Signals are information that a supplier can send to better communicate their ability, benevolence, or integrity to the buyer. For such signals to be credible, the cost of signaling must be negatively correlated with the capability being signaled (Spence, 1973). Consider, for example, the provision of product warranties by suppliers. For warranties to effectively signal supplier quality, the costs of providing warranties must be high for low-quality suppliers, and low for high-quality suppliers.

Effective signals create what is known as a separating equilibrium, where high-quality and low-quality suppliers have incentives to choose different signals (Boulding & Kirmani, 1993). Buyers can use effective signals to distinguish between high-quality and low-quality suppliers. Ineffective signals create a pooling equilibrium, in which high-quality and low-quality suppliers share incentives to invest in the same signals. Buyers are then unable to differentiate the suppliers using those signals.

In global B2B e-commerce, many exchanges offer services, such as third-party verifications of suppliers and web seals, as trust-building signals. These signals can play important roles in buyer-supplier trust. B2B exchanges typically maintain low entry costs for suppliers to increase their pool of suppliers and raise the liquidity and activity levels among the exchanges' users. The costs for suppliers to join an exchange can range from nothing (free membership) to between US$300 and US$7,500 per annum (paid memberships). Table 2.1 shows the annual paid membership fees (in addition to free membership options) in three B2B exchanges (as of November 2008). These membership fees are relatively low compared to suppliers' annual sales volumes or the values of typical B2B orders. The ease and affordability of exchange memberships make it easier and attractive for a supplier to (1) engage in identity theft/misrepresentation, where it intentionally and wrongfully submits information of a legally existing supplier, or (2) act as a phantom supplier by creating an account for a nonexistent company. Buyers purchasing from such suppliers face the risks of non-performance and usually have limited legal recourse. It is difficult to locate or take legal actionsagainst a nonexistent company in a foreign country. Moreover, due to the distance between trading partners in global B2B exchanges, buyers have difficulties verifying suppliers' identities, which affects their trust in these suppliers.

Table 2.1 Annual Paid Memberships in Three B2B Exchanges

B2B Exchange	Annual Membership Fee (US$)
Alibaba.com (www.alibaba.com)	$600 for TrustPass; $7,300 for Gold Supplier
EC Plaza (www.ecplaza.com)	$420
Gsm Exchange (www.gsmexchange.com)	$380

Therefore, buyers must rely on B2B exchanges to verify the suppliers' identities. For a fee, a supplier can initiate a third-party verification check through a B2B exchange. This verification check is often out-sourced to independent companies, which verify that particular supplier on the exchange is a registered company. In addition, these services also verify information posted by the supplier in the exchange by inspecting the supplier's production capabilities, premises, and factories. Such verifications signal the legality of the supplier and the authenticity of the information about them in the exchange. A supplier who passes the verification check usually receives a web seal on their company's profile page in the B2B exchange, which indicates that the information has been verified. Typically, the web seal is valid for one year, after which the supplier needs to be re-verified.

The mere presence of web seals, though, may not lead to higher trust. Some studies show that Better Business Bureau Online seals reduce the risk perceived by consumers (Grazioli & Javenpaa, 2000), while other studies find that seals of approval, privacy seals, and industry seals do not significantly affect customer trust (Fisher & Chu, 2009; Houston & Taylor, 1999; McKnight et al., 2004; Ou & Sia, 2010; Pennington, Wilcox, & Grover, 2003). The mixed results from these studies raise an important question about when web seals serve as effective trust-enhancing signals. We believe there are two essential conditions for web seals to improve buyer-supplier trust.

Condition 1. Does a particular web seal create the necessary separating equilibrium for it to be a credible signal? When the costs to obtain verification web seals are substantial, they provide credible signals. Even though the fees to initiate verification checks may be relatively low, the costs associated with having the documentation and capabilities to meet the verification requirements are often high. For instance, to pass the verification checks, suppliers must register their business and subject it to regulations, demonstrate that they have the production capacity, and/or show the certifications they claim (e.g., ISO 9001). Those who cannot incur these costs would either fail the verification checks or avoid undertaking them. Moreover, third-party companies that provide verification services have a continuing reputational stake in the verifications being accurate and untainted. This stake in their reputation is of greater value than acting opportunistically to help any particular supplier. As such, independent verifications are conducted with care. Thus, third-party verifications serve as implicit guarantees (Parkhe, 1998) and contribute to the formation of firm-specific trust (Zucker, 1986), just as outside auditors do in the context of managing the principal-agent problem in management settings (Antle, 1982, 1984; DeFond, Raghunandan, & Subramanyam, 2002).

Condition 2. The presence of a separating equilibrium, however, may be a necessary but insufficient condition for a web seal to be an effective signal. For the web seal to engender trust, buyers must care about the characteristics that are being qualified and signaled. McKnight et al. (2004) suggests that a possible reason why TRUSTe, a privacy web seal, did not improve consumer trust in their study was that the respondents did not consider privacy to be an important web problem. As we point out earlier, the authenticity of counterparties' identities and claims are essential in B2B relationships. Heide and John (1990) found that increased verification efforts by OEM buyers increased their joint action with the supplier (e.g., in the areas of component testing, planning, and forecasting). Similarly, Gefen (2004) found that quality certifications increased client trust in ERP software vendors. Because third-party supplier verification creates a separating equilibrium (see Condition 1 above) and is important to potential buyers in B2B commerce, we posit that:

H3: Supplier verification is positively related to the buyer's trust in the supplier.

2.3.3 Effects of Indices and Signals on Trust over Repeated Interactions

Previous research has examined how initial trust formation is affected by institutional mechanisms (e.g., McKnight, Cummings, & Chervany, 1998; Stewart, 2003; Zucker, 1986) or the specific use of web-seals (e.g., McKnight et al., 2004; Pennington et al., 2003). However, do factors that influence trust early in a buyer-supplier relationship have the same effect later in the relationship? Do the effects of trust indices and signals change as buyer-supplier relationships develop? These are important questions because buyer-supplier relationships can and do evolve over time.

When there are no transactions between the buyer and supplier, categorization processes such as stereotyping should affect the levels of trust between them (McKnight et al., 1998). At this initial stage of their relationship, buyers may expect a supplier to perform or behave like typical suppliers in that country. These expectations are shaped by their perceptions of national integrity and legal structure. Thus, when a buyer is unfamiliar with a supplier, information indices influence the buyer's trust in the supplier.

However, while information indices provide some indications of the typical suppliers' quality, the buyer gains knowledge about the specific supplier through first-hand, repeated interactions (Koehn, 2003; Ratnasingam, 2005). Cooperative history between partners in international alliances affects their trust in each other (Parkhe, 1998). With repeated transactions, buyers should rely less on their perceptions of the supplier's country (i.e. information indices) in evaluating the supplier's trustworthiness. Instead, they should base their evaluation on past performance of the supplier (Ariño et al., 2001; Lane, 1998; Zucker, 1986). Thus, we hypothesize that the influence of information indices on buyer's trust decays with more transactions between the buyer and supplier.

H4a: The effects of the perceived level of national integrity in the supplier's country on the buyer's trust should decline as the number of transactions increases.

H4b: The effects of the perceived level of legal structure in the supplier's country on the buyer's trust should decline as the number of transactions increases.

Do information signals in the form of supplier verifications and web-seals also become less influential with more transactions? Given the different levels of firm-specific information that indices and signals provide, we believe that there are structural differences in how these informational sources affect the development of interorganizational trust. Indices are generic and provide little firm-specific information. When only information about a supplier's location is available, a buyer would treat the supplier as typical of firms with similar attributes in that location (Spence, 1973). Direct experiences with the supplier, however, allow the buyer to move from an average impression of the supplier's quality to a more precise assessment. Relative to indices, signals provide more firm-specific information. Furthermore, because signals are in the firm's control, the absence of a signal in itself is also a signal, albeit a potentially counter-productive one. For instance, companies tend to purchase only from suppliers who are verified or qualified in order to maintain corporate governance and manage liability risks. In our case, the legality of the supplier and authenticity of their claims (ascertained through third-party verifications) are important criteria for the buyer regardless of the length and strength of the buyer-supplier relationship. The buyer may interpret that something is amiss when a long-term supplier is no longer verified – for example, they may wonder whether the supplier's license has been revoked or whether the supplier's production capacity has changed. Trading with this supplier would increase the buyer's exposure to risk[2]. This response is similar to market reactions when experienced professionals (e.g., lawyers or doctors) and established institutions (e.g., schools) lose their licenses or accreditations.

Therefore, unlike country-level indices, we do not expect repeat transactions between buyers and suppliers to moderate the influence of supplier-specific signals on buyers' trust. These signals provide critical information about individual suppliers, and should remain relevant and important even when the buyer has first-hand, direct experiences with the suppliers.

2.4 Method

2.4.1 Overview

We conducted an online survey of buyers on a global B2B exchange in September 2008. This exchange, started in the late 1990s, is operated by a publicly listed firm in Asia. By 2010, it had more than three million international users in its member base. The exchange handles products in multiple industries, including agriculture, electronics, and textiles. Buyers and suppliers can search for and post products, request quotes for price and terms, and contact one another through the exchange. The exchange offers various services for suppliers, such as premium membership (US$4,500 per annum) and third-party verification services (US$1,400 per annum)[3]. It also provides services such as banner advertising and reports on individual verified suppliers.

To develop the survey instrument, we first pre-tested using four organizational buyers from three countries. We obtained feedback about the structure, questions, and cognitive load of

---

[2] On February 21, 2011, Alibaba.com announced that about 1% of its verified suppliers engaged in fraud against its buyers. These fraudulent suppliers evaded the third-party verification process with the help of some Alibaba.com employees. Following the announcement, Alibaba.com's market capitalization dropped by almost US$1b, and its CEO and COO were replaced (although Alibaba.com's internal investigation confirmed that these executives were not involved in the incident). This incident shows that the market places a high value on verified suppliers.
[3] These rates are correct as of November 2008.

the survey from these buyers and refined the instrument. Next, we conducted four rounds of pretests that involved 600 randomly selected active buyers in the B2B exchange. Active buyers are those who had posted at least one buying request in the exchange and had logged in at least once within the three months before the survey. Since communication on the exchange (and in international trade) is primarily in English, we did not translate the instrument into various languages.

Two weeks after the final round of pretesting, the B2B exchange randomly selected and invited 5,250 active buyers (excluding those in the pretests) to take part in the actual survey. We gave respondents two weeks to complete the survey. The B2B exchange sent a first reminder email one week after the initial invitation, and a second reminder email two days before the survey ended. To assure the buyers of their confidentiality and anonymity, we informed them that their responses would be sent directly to the research team, and that the exchange would only receive aggregated results and not individual responses. In addition, we collected no identifying information during the survey. To further encourage participation, respondents who completed the survey received a US$20 credit to purchase reports from the exchange.

Our survey used a within-subject design. We asked the buyers to list company names or initials of two suppliers whom they would consider for an imminent corporate purchase[4]. At least one of the suppliers needed to be a participant in the exchange so that we could examine the influence of third-party verifications. Buyers not making such a purchase could exit the survey and still receive a US$20 credit from the exchange. Buyers whose purchase decisions met these criteria provided information on each supplier's verification status (conditional on the supplier being listed in the exchange)[5], evaluated each supplier's performance in past transactions (if any), and rated their trust in each supplier. Finally, we asked the buyers for their perceptions of the national integrity and legal structure in each supplier's country.

[4] Traditionally, respondents are asked to identify purchase decisions that they have been involved in (e.g., Doney & Canon, 1997). In such cases, it is possible that the measured post-transaction trust could differ from the unobserved pre-transaction trust. For instance, a buyer may have a high level of trust in a particular supplier before a transaction. However, due to a below expectation performance by the supplier, the buyer's trust in this supplier may be lowered after the transaction. To overcome such issues, we asked buyers to consider an imminent purchase that they were making. This approach allows us to better relate (pre-transaction) trust to purchase intention. Although there is a possibility that the buyers' trust in suppliers was biased (where more trustworthy suppliers were being considered for the transaction), we employed a within-subject research design to control for this potential bias (see Appendix G).

[5] As multiple suppliers could use the same company name, it was not feasible for the B2B exchange to provide information of the supplier's verification status. Therefore, we relied on buyers' input for this information.

Two hundred and eighty-seven buyers completed the survey, which provided information about 574 suppliers (two suppliers per respondent). The effective response rate is difficult to determine as not all the 5,250 buyers sampled were making an imminent purchase on the exchange during the survey period. The exchange found 19.95 percent of active buyers sent at least one enquiry to suppliers in a two-week period. Using this type of query as a proxy for whether a buyer was making an imminent purchase, the relevant sample size for this study is 1,048 buyers (5,250 x .1995) and the effective response rate is 27.39 percent.

Appendix 2.A shows the characteristics of our buyers, which include their location and product category of their imminent purchases. The average buyer in our dataset had three to five years of B2B e-commerce experience. On average, the buyer's company had between 10 and 19 employees and sales between US$500,000 and US$999,000 in the previous financial year. The buyer had on average purchased from between one and four other suppliers that are from the referent supplier's country. The median estimated transaction value of the imminent purchase was US$30,000. The B2B exchange reported that these respondents' characteristics and transaction values are representative of those in the exchange. Forty-six percent of buyer-supplier pairs in our sample had prior experiences with each other (see Table 2.2).

Table 2.2 Prior Transactions Between Buyers and Supplies

No. of Past Transactions	Percentage
No prior transaction	54%
Between 1 and 3 transactions	26%
Between 4 and 6 transactions	10%
Between 7 and 9 transactions	4%
Between 10 and 19 transactions	2%
20 transactions or more	4%
Note: Base on 574 pairs of buyer-supplier relationships in our sample.	

Because the B2B exchange did not provide information about non-respondents, we could not compare respondents' attributes with those of non-respondents. Instead, to check for non-response bias, we compared buyers who responded before the final reminder with those who responded after. There were no significant differences between early and late respondents in company's sales, number of employees, purchase value, respondents' education, working experience, and B2B e-commerce experience. This suggests non-response bias is not a problem in our sample (Armstrong & Overton, 1977).

2.4.2 Measures

Appendix 2.B presents the items that this study used. When appropriate, we specified the suppliers' company name or initials in the questions' stem by using information provided by the buyers. This clarified the questions to the buyers, especially since they had to evaluate two suppliers in the survey.

<u>Purchase Intention.</u> Although this study focuses on antecedents of trust, we also measured outcomes of trust so that we could relate our findings to past research and estimate the expected value of information indices and signals. Because buyers evaluated their supplier before actual purchases, a relevant outcome of trust is the likelihood of purchasing from that supplier.

Because the B2B exchange did not track actual transactions, and because some buyers also evaluated suppliers that were not participating on the exchange, we could not use archived purchase data in our analysis. Additionally, because buyers spend different amounts of time making their purchase decisions, it would have been challenging to follow-up with them to get information about their actual purchases. Therefore, we asked buyers to estimate on a 5-point Likert scale the likelihood of making the imminent purchase with the particular supplier. Verbal statements of purchase intentions are excellent predictors of actual purchase behavior (Armitage & Conner, 2001; Sheppard, Hartwick, & Warshaw, 1988; Webb & Sheeran, 2006).

<u>*Buyer's Trust.*</u> We used nine items from Mayer and Davis (1999) to measure the buyer's trust in the supplier. Sample survey items include "Supplier X is well qualified" and "Supplier X would not knowingly do anything to hurt me". We changed one of the items to focus on the supplier's capabilities (instead of skills as per Mayer and Davis' (1999) measures) to make the question contextually relevant. We also asked about the extent to which the supplier can be trusted.

National Integrity. We used two 5-point Likert scale items to measure the buyer's perceptions of the national integrity in a particular supplier's country: the likelihood that suppliers in that country would behave with integrity and do the right things in business deals. These items are similar to those that Morgan and Hunt (1994) use in their study of dyadic retailer-supplier relationships; however, our items focused on the buyer's perception of all suppliers in the country instead on the individual focal suppliers. Appendix 2.C shows the average perceived national integrity ratings for the 50 supplier countries in our sample.

Legal Structure. The measure of the buyer's perceptions of the legal structure in a supplier's country came from two sources. The first comprised three 5-point Likert scale items to measure the buyer's confidence in the legal systems in that country, and the perceived effectiveness of the laws and regulations in that country to govern the suppliers' operations and resolve business disputes, respectively.

The second source was the 2007 corruption perception index (CPI), administered by Transparency International. The CPI is a composite index that provides information about perceptions of corruption within countries. The index score ranges from 0 (high corruption) to 10 (low corruption). The 2007 CPI is based on 14 sources that originate from 12 institutions, such as the Asian Development Bank, the Economist Intelligence Unit, and the World Economic Forum. The average correlations between the sources are .77, which suggests high overall reliability of the CPI (Lambsdorff, 2007). Moreover, Herzfeld and Weiss (2003) found that a positive relationship between countries' CPI scores and the degree to which their citizens are willing to accept the established institutions to make and implement laws and adjudicate disputes. Hence, CPI is a relevant external measure of legal structure perception for this study.

Appendix 2.C shows the average perceived legal structure ratings and CPI scores for the 50 supplier countries in our sample. These two measures correlate at .44 ($p < .01$), which supports the validity of our survey measure of legal structure perceptions.

Supplier Verification. We asked the buyer to indicate the supplier's verification status, provided that the supplier is listed in the exchange. The buyer indicated "not sure" if they could not recall this information about the supplier. The verification status indicator takes the value of 1 if the supplier was verified and 0 otherwise.

Past Transactions. We accounted for the buyer's experience with the supplier using the number of transactions between them over the last 12 months, as reported by the buyer.

Supplier's Performance. To control for supplier's performance, we asked the buyer to compare the referent supplier to other suppliers in terms of three performance criteria: price, product availability, and delivery (Doney & Cannon, 1997). We measured the responses for each on a 5-point Likert scale, ranging from performing much worse than other suppliers to performing much better than other suppliers. The neutral point on the scale was that the supplier's performance was equal to other suppliers'. The buyer indicated "not sure" if they were unable to ascertain the supplier's relative performance.

Supplier Membership. We dummy coded supplier memberships to control for different membership types. We categorized suppliers who were not on the B2B exchange as non-members. Among suppliers who were listed on the exchange, we categorized those with paid memberships as paid members. The buyer indicated "not sure" if they could not recall the supplier's membership type in the B2B exchange. We used suppliers on the exchange with free membership or whose membership types buyers could not recall as the reference group in our analyses.

Same Country. Because cultural or ethnic similarity may influence trust, we controlled for whether the buyer and supplier were from the same country using a dummy variable. Since the buyer indicated their and the supplier's countries during the survey, we matched their responses to code this dummy variable. The variable takes the value of 1 if the buyer and supplier were from the same country and 0 otherwise.

China Supplier. Seventy percent of the suppliers in our sample were based in China, which reflectis the current state of international trade where buyers actively source from China. We added a country dummy that takes the value of 1 if the supplier was from China and 0 otherwise.

2.5 Results and Analyses

Since our respondents were from different countries, we assessed whether we should pool their responses in our analyses. We conducted a Kruskal-Wallis test to assess differences among respondents between countries (Appendix 2.D). The results show that it is reasonable to pool respondents across countries in our analyses. We also assessed the presence of common method variance in two ways (Appendix 2.D). First, we conducted Harman's one-factor test. Second, in a stronger, more refined test that fits our research setting, we compared (1) the covariance of buyer's trust in and likelihood of purchasing from one supplier (i.e., within-supplier covariance), and (2) the covariance of buyer's trust in one supplier and likelihood of purchasing from the other supplier (i.e. between-supplier covariance). The results from both tests indicate that common method variance is not a problem in our data.

2.5.1 Structural Equation Modeling

We analyzed our data using structural equation modeling. Structural equation modeling (SEM) provides the flexibility to properly account for measurement error by having multiple indicators per latent variable. It also allows us to test the overall model and model the error terms. In addition, we can include a consequence of trust (i.e., purchase intention) in the structural model and estimate the expected values of trust indices and signal. We did this using the Mplus (version 5.21) software, which lets us model interaction using the latent moderated structural equations approach (Klein & Moosbrugger, 2000). This approach results in relatively smaller bias of structural parameter estimates and higher power to detect interaction effects than partial least square (Schermelleh-Engel, Werner, Klein, & Moosbrugger, 2010).

Appendix 2.E shows the descriptive statistics and intercorrelations of the items. We mean-centered the items for national integrity, legal structure, and past transactions before creating the interaction terms. Using maximum likelihood, we simultaneously estimated the measurement and structural models. To reduce the number of parameters to estimate, we assigned the ten survey items that measure buyer's trust into three parcels (Trust-A (4 items), Trust-B (3 items), and Trust-C (3 items)), and used them as indicators of the latent variable buyer's trust. Each parcel's score is the average score of the assigned items. The latent variable legal structure comprises four indicators – the three survey items and the supplier's country CPI score. We fixed the error variance of single-indicator variables (i.e., purchase intention and past transactions) with the assumption that the reliability for each of these indicators is .85. Using the Spearman Brown prophecy formula and Cronbach's α for the measure of buyer's anticipated future purchase in Doner and Canon's (1997) study, we estimated a reliability of .90 had we used a single-item scale to measure purchase intention. Our assumed reliability of .85 is therefore conservative.

Also, because each respondent in our dataset provided two supplier-observations, individual respondents' observations may have correlated errors. To obtain robust variance estimate, we clustered the observations by respondent to appropriately adjust the standard errors (Wooldridge, 2002). Lastly, when estimating the structural model, we treated "not sure" responses for supplier's price, product, and delivery performances as missing data. We assumed these responses to be missing at random, which makes the use of maximum likelihood estimation with estimation of missing data values appropriate and strongly preferable to listwise case deletion (Schafer & Graham, 2002).

Measurement Model. We conducted a confirmatory factor analysis and computed the Cronbach's α of the multi-items constructs in our model (Appendix 2.F). The model fit is not significant ($\chi^2 = 55.20$, d.f. = 48, $p > .10$), and the other fit indices also indicate good model fit (CFI = .99; RMSEA = .02; SRMR = .03). All items loaded on their respective constructs. The Cronbach's α estimates suggest the items have good internal consistency. Good convergent validity is shown by higher correlations between items reflecting the same construct than correlations between items reflecting different constructs (see Appendix 2.E). We tested discriminant validity of our constructs using a chi-square difference test (Bagozzi, Yi, & Phillips, 1991). For each pair of constructs, we ran a chi-square difference test that compared an unrestricted model (where correlation of the constructs was freely estimated) and a restricted model (where correlation was fixed to unity). In all pair-wise comparisons, the two models differ significantly on the chi-squared difference test ($p < .001$), with the unrestricted models having better fit, which supports the discriminant validity of the constructs.

Structural Model. We estimated a baseline model with only the main effects. The test of fit for this model is significant ($\chi^2 = 352.04$, d.f. = 113, $p < .01$), but the other fit indices indicate adequate model fit (CFI = .92, RMSEA = .06, SRMR = .05). Based on the modification indices, we added covariance between trust and purchase intention, and between the CPI and whether the supplier is from China. A buyer's trust and purchase intention may share common causes that we did not measure (e.g., buyer's commitment to supplier). Also, given that the proportion of China suppliers in our dataset is high and that the CPI is a country-level score, it is reasonable to allow their error terms to correlate. We re-estimated the model with these modifications (Model 1 in Table 2.3). Although the test of fit is still significant ($\chi^2 = 209.36$, d.f. = 111, $p < .01$), the other goodness of fit indices improved (CFI = .97, RMSEA = .04, SRMR = .04).

We next added two interaction terms, "national integrity x past transactions" and "legal structure x past transactions", to the structural model (Model 2 in Table 3). Including these interaction terms makes the second model a non-nested model relative to the first model, and these two models have overlapping but non-identical variance-covariance matrices (Vandenberg & Grelle, 2009). We compared the models' Akaike information criteria (AIC) and Bayesian information criteria (BIC) to assess whether including the interaction terms is appropriate. These information criteria reward a goodness of model fit and penalize a lack of model parsimony, and the model with the smaller AIC and BIC is the better one (Vandenberg & Grelle, 2009). The AIC declined from 20686.42 in Model 1 to 19482.23 in Model 2, and the BIC (adjusted for sample-size) declined from 20785.38 to 19559.98. The relatively large reductions in information criteria values support the inclusion of these two interaction effects in our structural model.

2.5.2 Main Results

Figure 1 presents the hypotheses testing results using Model 2 in Table 2.3. After controlling for the supplier's performance, the supplier's membership category, whether the buyer and supplier are from the same country, and whether the supplier is China based, we found that the favorable perception of national integrity had a positive effect on buyers' trust in the supplier ($\beta_2 = .18$, $p < .05$). However, this perception did not become less influential with increasing transactions ($\beta_3 = .10$, $p > .10$). Therefore, H1 is supported but H4a is not. Although the perception of national integrity perception appears to have modest statistical impacts, its economic impacts are substantial and meaningful. We discuss the practical impact of our findings in Section 2.6.2.

Table 2.3. Structural Model Results

	Model 1	Model 2	Model
	Coeff.	**Coeff.**	**Coeff.**
β_1: Buyer's trust → likelihood of purchase	0.94** (0.11)	0.93** (0.11)	0.92** (0.11)
β_2: National integrity → buyer's trust	0.17* (0.08)	0.18* (0.08)	0.18* (0.08)
β_3: National integrity x past transaction → buyer's trust	-	0.10 (0.07)	0.10 (0.07)
β_4: Legal structure → buyer's trust	0.06 (0.06)	0.05 (0.06)	0.05 (0.06)
β_5: Legal structure x past transaction → buyer's trust	-	-0.09* (0.05)	-0.11* (0.05)
β_6: Supplier verification → buyer's trust	0.20+ (0.11)	0.24* (0.11)	0.12 (0.14)
β_7: Supplier verification x past transaction → buyer's trust	-	-	0.07 (0.06)
β_8: Past transaction → buyer's trust	0.06* (0.03)	0.06* (0.03)	0.05 (0.03)
Controls			
β_9: Supplier's performance → buyer's trust	0.59** (0.12)	0.58** (0.12)	0.59** (0.11)
β_{10}: Non member → buyer's trust	0.03 (0.07)	0.05 (0.06)	0.06 (0.06)
β_{11}: Paid member → buyer's trust	-0.15+ (0.09)	-0.16+ (0.09)	-0.15+ (0.09)
β_{12}: Same country → buyer's trust	0.06 (0.11)	0.02 (0.13)	0.02 (0.12)
β_{13}: China supplier → buyer's trust	0.04 (0.07)	0.02 (0.07)	0.02 (0.07)
χ^2 (d.f.)	209.36 (111), p < .01	-	-
CFI	0.97	-	-
RMSEA	0.04	-	-
SRMR	0.04	-	-
AIC	20686.42	19482.23	19482.90
BIC	20785.37	19559.98	19561.83

Note: + p < .10 * p < .05 ** p < .01
Note: Standard errors in parentheses. The software package (Mplus) does not provide goodness of fit indices for Models 2 and 3, where we include the interaction terms.

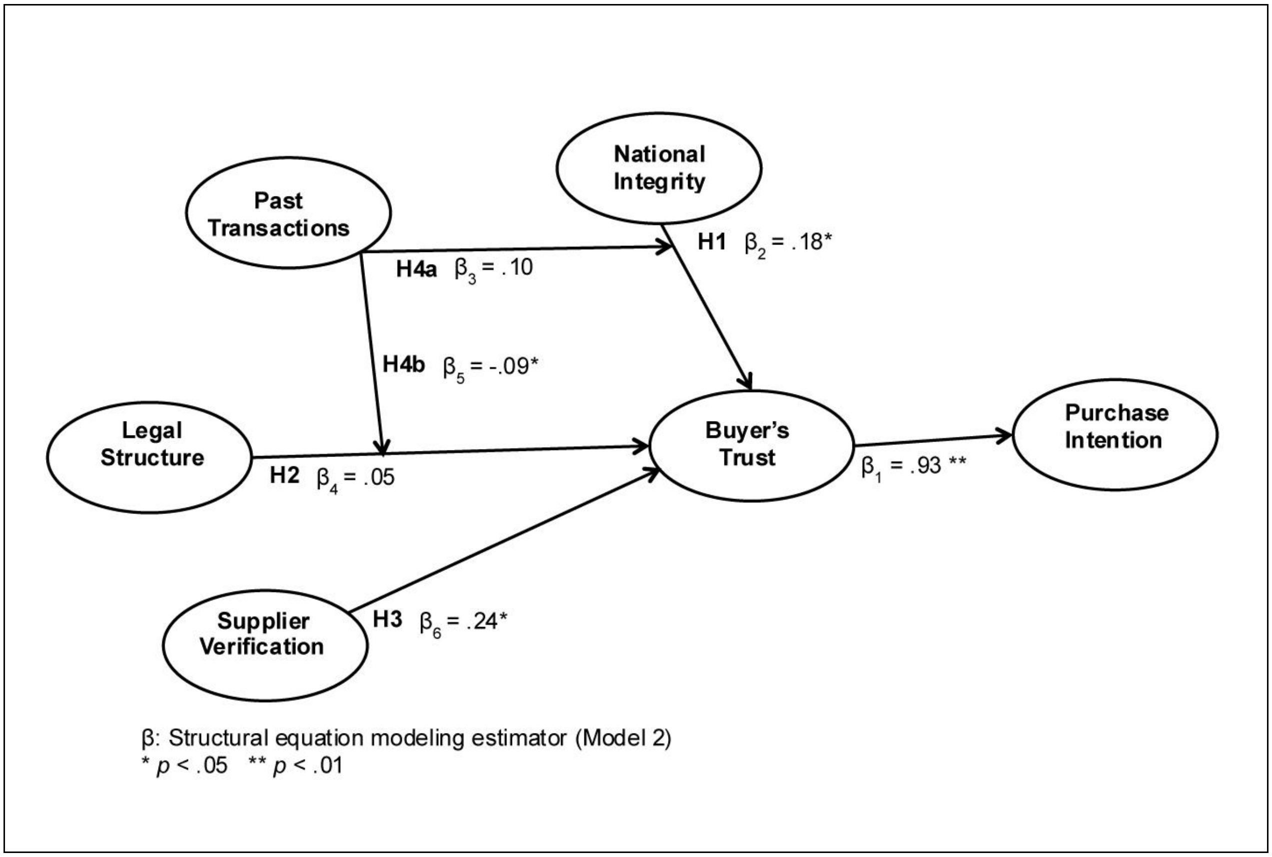


Figure 1. Summary of Results

We also found that (i) the perceived legal structure positively impacted buyers' trust when the buyer had relatively few prior interactions with the supplier, but that (ii) the effects of perceived legal structure on buyers' trust weakened as the number of interactions increased. The number of prior buyer-supplier transactions negatively moderated the relationship between perceived legal structure and buyer's trust (β_5 = -.09, p < .05), although the positive main effect of perceived legal structure on trust was not statistically significant (β_4 = .05, p > .10). These results collectively support H2 and H4b. In addition, buyers' trust was higher in suppliers who had been verified by a third party compared with those who were not verified (β_6 = .24, p < .05), which supports H3. Lastly, a buyer's trust in a supplier was positively related to the likelihood that the buyer would purchase from that supplier (β_1= .93, p < .01). This result is consistent with other studies on outcomes of trust (e.g., Doney & Cannon, 1997; Pavlou, 2002).

We then added the interaction supplier verification x past transactions in Model 3. Since AIC and BIC were higher in Model 3 than in Model 2 (see Table 3), there is no evidence to suggest that Model 3 is a better model (Vandenberg & Grelle, 2009). This interaction term was also not significant ($\beta_7 = .07$, $p > .10$), which implies that the impact of a supplier's verification status on buyer's trust does not diminish with more transactions between them.

2.5.3 Robustness Analyses

We conducted several analyses to check the robustness of our results (see Appendix 2.G for details). First, unobserved effects such as priming, social desirability, and buyer heterogeneity may bias our results. For example, respondents might give positive evaluations of suppliers because they were strongly considering these suppliers in their purchases. The respondents' trust in individual suppliers might also be affected by their disposition to trust (Balasubramanian et al., 2003; McKnight & Chervany, 2001; McKnight et al., 1998) or by their trust in the exchange (Pavlou, 2002; Pavlou & Gefen, 2004). Our within-subject design, where each respondent evaluated two suppliers, allows us to account for these unobserved effects using a random effects model. The results using random effect models show qualitatively similar conclusions as the results using SEM.

Second, we replaced survey data about supplier's membership types with available data from the exchange. We also checked the sensitivity of our SEM results to different model specifications and assumed reliability of single-indicator variables. Finally, we ran a multi-level mixed effects model because the survey responses were nested within buyers. The results in these analyses are consistent with the main results discussed above.

2.6. Discussion

Information systems researchers have examined the roles and impacts of interorganization systems such as electronic data interchange (e.g., Mukhopadhyay, Kekre, & Kalathur, 1995) and electronic infomediaries (e.g., Ghose, Mukhopadhyay, & Rajan, 2007). Online B2B exchanges are also interorganization systems that help buyers and suppliers search for and connect with each other (Pavlou, 2002). Yet firms do not establish relationship with each other simply because the systems to do so are in place. In this study, we examined factors that affect the formation and development of interorganizational relationships on online exchanges, particularly in a global setting.

Using information signaling theory (Spence, 1973), we treated perceived level of national integrity and legal structure in the supplier's country as indices that are difficult for suppliers to alter, and third-party verifications and web seals on B2B exchanges as costly signals that suppliers can manipulate at their discretion. Our results show that supplier indices and signals have positive effects on buyers' trust. Hypotheses 4a and 4b suggested that, with increased experience (more past transactions), the effect of indices such as legal structure and national integrity would decline. This would be evidenced by negative coefficients for interactions of the respective indices and past transactions. We did find such an effect for legal structure by past transactions (β_5= -.09, $p < .05$). However, the estimated coefficient for national integrity by past transactions was not significant (β_3= .10, $p > .10$), which suggests that national integrity is still a consideration even with much past experience. A possible explanation for this non-significant effect is that national integrity and individual suppliers' integrity are more closely associated than we expected – buyers may expect social norms to strongly influence individual suppliers' behaviors, even for those suppliers whom they have transacted with. Since integrity is a key component of trust, and repair of trust due to integrity-related violations (e.g., dishonest behavior) is difficult (Kim, Dirks, & Cooper, 2004; Kim, Ferrin, & Cooper, 2009), buyers' perceptions of national integrity may still matter even when they have completed many transactions with the suppliers in the past. Finally, we also found that buyers' trust positively affects their supplier-selection decisions. Buyers were more likely to purchase from suppliers whom they trust more (β_1 = .93, $p < .01$).

Apart from third-party verifications, another potential signal for suppliers in B2B exchanges is paid membership. Surprisingly, we found a weak negative relationship between buyer's trust and paid membership (β_{11} = -.16, $p < .10$), which indicates that buyers may distrust suppliers that are on paid memberships. We also observed this phenomenon among B2B exchange users. For instance, a participant in an online community shared their experiences with suppliers on paid membership (Gold membership) in a B2B exchange (Alibaba.com):

> *"Bear in mind that I have successfully dealt with an Alibaba Gold member, and still do this day, so I was taken in by the belief that Gold membership meant that the company I was dealing with would be more genuine, than say a free member seller.* ***I now know that this is not the case.****" (Emphasis added) (Robbobb, 2007).*

Our results and the anecdotal evidence suggest that genuine suppliers may not effectively distinguish themselves from non-genuine ones by subscribing to paid memberships – in fact, such services seem to have adverse effects on genuine suppliers. Unlike verification services, paid memberships usually just require suppliers to pay a fee, which may not be effective barriers to untrustworthy suppliers. Membership fees could be too low to separate types of suppliers, which leads to a pooling equilibrium.

Our results add to our understanding of cross-border transactions on online B2B exchanges. Regardless of whether B2B transactions occur within or across borders, or through online exchanges or physical channels, buyers look for certain qualities in suppliers – competency, integrity, and benevolence. However, it is more challenging to identify these qualities in cross-border e-commerce due to information asymmetry in online markets. Moreover, research in localized B2B e-commerce typically focuses on technological structure, particularly the situational normality and structural assurance of the Internet or platforms (e.g., Pavlou, 2002; Pavlou & Ratnasingam, 2003; Ratnasingam, 2005). However, these factors, cannot explain why a buyer's trust in suppliers from different countries may differ on a B2B exchange. A favorable and secured online platform may not be a sufficient condition for global e-commerce transactions to occur because a buyer's trust in foreign suppliers and the buyer's purchase intention also depend on factors that are external to the platform. Our results show that it is necessary to examine social-economic characteristics in partners' countries in globalized B2B e-commerce.

Furthermore, supplier verifications may be more salient when cross-border transactions take place on online exchanges instead of through traditional channels. Earlier studies, particularly those that predate the Internet era, have considered the importance of foreign supplier certifications in international sourcing through physical channels (e.g., Birou & Fawcett, 1993; Scully & Fawcett, 1994). In these cases, an implicit assumption is that the suppliers' certifications are authentic. However, the importance of authenticating trading partners' information, is a relatively new phenomenon with the proliferation of online platforms (e.g., Basu & Muylle, 2003; Lee, 2002). As we note earlier, the relatively low costs of exchange memberships lead to problems such as misrepresentation and phantom suppliers. Thus, in the context of exchange platforms, the authenticity of a firm's information could be as important as the information itself.

2.6.1 Theoretical Implications

With the Internet and e-commerce technology, organizations can now easily look beyond their local markets for new buyers and suppliers. While participating in the global marketplace is attractive, the risks and uncertainties that come with it are qualitatively different from those that arise in domestic exchange. By focusing on globalized B2B e-commerce, we see important factors of trust that are not salient in localized commerce of any kind. Our findings show that perceptions of country and supplier attributes influence buyers' trust. It is important to account for such indices and signals when studying information asymmetry and signaling. Indices are often treated as given. This study suggests that they have informational impact precisely because they are relatively unalterable by suppliers.

We expected that perceptions of supplier-country attributes would have less effect as the buyer gained experience with the supplier. The actual picture is more complicated. Whereas the effects of legal structure on a buyer's trust diminished with repeated transactions, national integrity remained influential. These differences may be due to the different basis for each perception. Perceived national integrity is a cognition-based trust mechanism. It influences judgment of trustworthiness via a categorization process, where an entity in an untrustworthy culture is expected to be untrustworthy (Zaheer & Zaheer, 2006). In cross-border e-commerce, even though a supplier is highly rated in terms of their ability, benevolence, and/or integrity based on their performance in prior transactions, the larger context may make trusting the supplier unwise or indicate the need for lower trust. For example, when opportunism is common or highly tolerated in a particular culture, high integrity and benevolence of an individual may be insufficient assurance that the individual will not be opportunistic (Wicks, Berman, & Jones, 1999). Therefore, the stereotyping of counterparts' national characteristics not only affects initial trust formation (Ariño et al., 2001), but it also influences subsequent trust development as relationships between the partners grow.

Perceived legal structure, in contrast, is an institution-based trust mechanism. Institution-based trust is important when there are limited prior exchanges between buyers and suppliers (Zucker, 1986). However, third-party, institutional mechanisms provide fewer cues about individual suppliers' competency. While the legal structure and licensing requirements in a country provide general indications of the quality of typical suppliers, a buyer learns about the ability of a particular supplier as the exchanges between them increase. Furthermore, although the enforceability of contracts and legal recourses are important in cross-border transactions, these considerations are more important when the partners are unfamiliar with each other. In inter-organizational relationships, buyers and suppliers may avoid invoking legal sanctions when trade disputes occur, as doing so is costly and interferes with their desire to continue doing business with one another (Macaulay, 1963). Instead, they try to resolve disputes through direct negotiations. Thus, when buyers and suppliers can assess each other's trustworthiness through direct means and exchanges, institution-based trust mechanisms become less influential. As a result, buyers' reliance on the legal system in suppliers' countries decreases with repeat transactions. Work in this area should account for the length and strength of buyer-suppliers relationships when examining institution-based trust.

In addition, this study extends research on how online exchanges mechanisms, such as feedback systems, affect a buyer's trust in the community of suppliers on the exchanges (Pavlou, 2002; Pavlou & Gefen, 2004). Just because a buyer trusts the community of suppliers does not mean that they have the same level of trust in every individual supplier in that community. Ultimately, trust at the dyadic buyer-supplier level plays an important role in shaping individual buyers' relationships with their suppliers. We see here that suppliers can differentiate themselves on an exchange by sending credible and important signals of their trustworthiness through third-party verifications.

2.6.2 Managerial Implications

Do the estimated effects in this study have meaningful consequences for buyers and suppliers on B2B exchanges and for policy makers? To address this, we estimated the economic impacts of supplier verification and changes in perceived national integrity using the SEM results in Table 2.3 (Model 2). The expected value of supplier verification on the B2B exchange was 22.32 percent (.93 (β_1) x .24 (β_6)) of the transaction value. Using the median transaction value of US$30,000 and assuming a supplier is considered for 12 purchases per year on the exchange, the expected value for the supplier of being verified was US$80,352 annually[6]. Since the primary context of this study was in cross-border transactions, and perceived country characteristics affect all suppliers within a country, we estimated the impact of perceived national integrity on the external trade of an economy. The expected value of a .01-point improvement in perceived national integrity (measured on a 5-point scale) was .17 percent (.93 (β_1) x .18 (β_2) x .01) of the total export value of a country. Using the 2008 trade statistics for China, for instance, a .01-point increase in its perceived national integrity has an expected value of US$2.43 billion[7]. While this figure is only suggestive, it indicates the potential impact of even small changes in perceived country characteristics.

Therefore, suppliers should be aware of how buyers' perceptions of the legal structure and national integrity in their countries affect buyers' trust. A buyer's trust due to country attributes is essentially beyond an individual supplier's control. Moreover, perceptions of country attributes of a particular country may differ among buyers, which makes it more challenging for individual suppliers to come up with an optimal strategy to engender and sustain trust. As such, what might be needed is a concerted effort to improve the perceived legal structure and national integrity of suppliers at the industry or country level. This can sometimes be done through voluntary business associations developing accreditation standards and self-regulation. Such improvements can benefit both individual suppliers and the respective countries as a whole. For example, Knack and Keefer (1997) found that a 10 percent increase in the level of trust in a society was associated with a .8 percent rise in annual growth in per capita income. The analysis in our study also shows that a small change in perceived national integrity can have a relatively significant impact on a country's external trade.

The negative relationship between buyer's trust and paid membership has important implications for B2B exchanges. Paid memberships are important revenue sources for these exchanges. While there are other reasons why suppliers subscribe to paid membership, such as to

[6] Users' testimonials in various B2B exchanges indicated that suppliers received between 20 and 200 enquiries per month. Our assumption of the supplier being considered for 12 purchases per year is therefore quite conservative.

[7] The value of China's export to the rest of the world was US$1,430.7 billion in 2008. (Source: http://www.uschina.org/statistics/ tradetable.html)

communicate more information or enjoy better customer support on the exchanges, it is nevertheless important for exchanges to explore how they can help those trustworthy suppliers who take up paid membership to increase the buyers' trust and eventually sales through their online marketplaces. One suggestion is for B2B exchanges to create different classes of paid memberships, and entry into certain membership classes requires suppliers to meet additional criteria that credibly signal high trustworthiness, such as suppliers' track record on the exchanges. This could help buyers differentiate among suppliers based on paid memberships.

2.6.3 Limitations and Suggestions for Future Studies

In this study, we examined the relationship between buyers' trust and their perceptions of suppliers' country attributes. Reverse causality is a potential problem: a buyer's trust in a supplier could affect their perceptions of the supplier's country, particularly if that supplier is the only supplier from that country with whom the buyer has interacted. Reverse causality becomes less of a problem as the buyer interacts with more suppliers from that country. In this study, buyers reported that, on average, they purchased from between one and four other suppliers from the referent supplier's country. Therefore, we do not expect that reverse causality was a major problem in this study.

In terms of the theoretical model, a possible extension is to examine the direct and indirect influences of industry-level perceptions on buyers' trust. For example, safety issues in China's toy and dairy industries could have affected the perceived trustworthiness of suppliers in these industries (Fairclough, 2007; Chao, 2008). For example, the Dairy Association of China estimated that it would take about two years to restore consumer confidence following the food safety incident in 2008, where milk and infant formula were adulterated with melamine (Zhou, 2008). There might also be spillover effects to other industries in the country, thus indirectly affecting buyers' perceptions of the country attributes. Including perception of industry-level attributes in the model would strengthen our understanding of how such higher-level perceptions influence trust at dyadic levels.

2.7. Conclusion

Therefore, is the world really flat as Friedman (2005) asserts? Perhaps less so than Friedman thinks. Although physical and geographical boundaries are now less of an obstacle in economic exchanges, they still play important roles in economic agents' attitudes, behaviors, and decisions. Hence there is a need to examine how cross-boundaries exchanges and relationships in all commerce are shaped by country characteristics. Increasingly, transactions are taking place globally, in B2B commerce and elsewhere. Consumers all over the world can now purchase from online retailers based in the US (e.g., Amazon.com) or individuals (e.g., through eBay). Yet most research in e-commerce and sourcing focuses mainly on deals that occur locally (particularly in the US). Adopting a cross-boundary and global perspective in e-commerce studies would enrich research and help to further maximize the benefits that the Internet and e-commerce bring to the global marketplace.

References

Antle, R. (1982). The auditor as an economic agent. *Journal of Accounting Research*, *20*(2), 503-527.

Antle, R. (1984). Auditor independence. *Journal of Accounting Research*, *22*(1), 1-20.

Ariño, A., de la Torre, J., & Ring, P. S. (2001). Relational quality: Managing trust in corporate alliances. *California Management Review*, *44*(1), 109-131.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, *40*, 471-499.

Armstrong, J. S., & Overton, T. S. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research*, *14*(3), 396-402.

Arrow, K. (1974). The limits of organizations. New York: W. W. Norton & Company.

Aulakh, P. S., Kotabe, M., & Sahay, A. (1996). Trust and performance in cross-border marketing partnerships: A behavioral approach. *Journal of International Business Studies*, *27*(5), 1005-1032.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, *26*(3), 243-268.

Bachmann, R. (2001). Trust, Power and control in trans-organizational relations. *Organization Studies*, *22*(2), 337-365.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*(3), 421-458.

Bakos, Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM*, *41*(8), 35-42.

Balasubramanian, S., Konana, P., & Menon, N. M. (2003). Customer satisfaction in virtual environments: A study of online investing. *Management Science*, *49*(7), 871-889.

Basu, A., & Muylle, S. (2003). Authentication in e-commerce. *Communications of the ACM*, *46*(12), 159-166.

Bilkey, W., & Nes, E. (1982). Country-of-origin effects on product evaluations. *Journal of International Business Studies*, *13*(1), 89-100.

Birou, L. M., & Fawcett, S. E. (1993). International purchasing: Benefits, requirements, and challenges. *Journal of Supply Chain Management*, *29*(2), 27-37.

Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, *50*(11), 1587-1602.

Boulding, W., & Kirmani, A. (1993). A consumer-side experimental examination of signaling theory: Do consumers perceive warranties as signals of quality? *Journal of Consumer Research*, *20*(1), 111-123.

Bradach, J. L., & Eccles, R. G. (1989). Price, authority, and trust: From ideal types to plural forms. *Annual Review of Sociology*, *15*, 97-118.

Chao, L. (2008). Tainted baby formula blamed in Chinese kidney cases. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB122115170115523971.html

Chau, P. Y. K., Hu, P. J.-H., Lee, B. L. P., & Au, A. K. K. (2007). Examining customers' trust in online vendors and their dropout decisions: An empirical study. *Electronic Commerce Research and Applications*, *6*(2), 171-182.

Chiles, H., & McMackin, J. (1996). Integrating variable risk preferences, trust, and transaction cost economics. *Academy of Management Review*, *21*(1), 73-99.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95-S120.

Dai, Q., & Kauffman, R. J. (2002). Business models for internet-based B2B electronic markets. *International Journal of Electronic Commerce*, *6*(4), 41-72.

DeFond, M., Raghunandan, K., & Subramanyam, K. (2002). Do non–audit service fees impair auditor independence? Evidence from going concern audit opinions. *Journal of Accounting Research*, *40*(4), 1247-1274.

Doney, P. M., Barry, J., & Abratt, R. (2007). Trust determinants and outcomes in global B2B services. *European Journal of Marketing*, *41*(9/10), 1096-1116.

Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing*, *61*(2), 35-51.

Doney, P. M., Cannon, J. P., & Mullen, M. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review*, *23*(3), 601-620.

Dyer, J. H., & Chu, W. (2003). The role of trustworthiness in reducing transaction costs and increasing information sharing: Empirical evidence from the United States, Japan, and Korea. *Organization Science*, *14*(1), 57-68.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57-74.

Fairclough, G. (2007). New recall of Chinese toys adds to safety concerns. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB118368769117358824.html

Fisher, R., & Chu, S. Z. (2009). Initial online trust formation: The role of company location and web assurance. *Managerial Auditing Journal*, *24*(6), 542-563.

Friedman, T. L. (2005). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: Free Press.

Gefen, D. (2004). What makes an ERP implementation relationship worthwhile: Linking trust mechanisms and ERP usefulness. *Journal of Management Information Systems*, *21*(1), 263–288.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated mode. *MIS Quarterly*, *27*(1), 51-91.

Ghose, A., Mukhopadhyay, T., & Rajan, U. (2007). The impact of Internet referral services on a supply chain. *Information Systems Research*, *18*(3), 300-319.

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.

Gulati, R., & Nickerson, J. A. (2008). Interorganizational trust, governance choice, and exchange performance. *Organizational Science*, *19*(5), 688-708.

Heide, J. B., & John, G. (1990). Alliances in industrial purchasing: The determinants of joint action in buyer-supplier relationships. *Journal of Marketing Research*, *27*(1), 24-36.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker., J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*, 1480-1484.

Herzfeld, T., & Weiss, C. (2003). Corruption and legal (in)effectiveness: an empirical investigation. *European Journal of Political Economy*, *19*(3), 621 - 632.

Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics*, *9*(1), 42-63.

Holton, R. H. (1989). Foreign investments and joint ventures: An American perspective. In R. H. Holton & W. Xi (Eds.), *U.S.-China economic relations: Present and future* (pp. 224-242). Berkeley, CA: Institute of East Asian Studies

Houston, R. W., & Taylor, G. K. (1999). Consumer perceptions of CPA WebTrust[sm] assurances: Evidence of an expectation gap. *International Journal of Auditing*, *3*(2), 89-105.

Kim, P. H., Dirks, K. T., & Cooper, C. D. (2009). The repair of trust: A dynamic bilateral perspective and multilevel conceptualization. *Academy of Management Review*, *34*(3), 401-422.

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The Effects of apology vs. denial for repairing competence- vs. integrity-based trust violations. *Journal of Applied Psychology*, *89*(1), 104-118.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*(4), 457-474.

Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics*, *112*(4), 1251-1288.

Koehn, D. (2003). The nature of and conditions for online trust. *Journal of Business Ethics*, *43*(1/2), 3-19.

Kwon, I.-W. G. & Suh, T. (2004). Factors affecting the level of trust and commitment in supply chain relationships. *Journal of Supply Chain Management*, *40*(2), 4-14.

Lambsdorff, J. G. (2007). *The methodology of the Corruption Perception Index 2007.* Transparency International. Retrieved from Special Investigation Service of the Republic of Lithuania website: http://www.stt.lt/documents/soc_tyrimai/KSI_methodolo gy_2007.pdf

Lane, C. (1998). Introduction: theories and issues in the study of trust. In C. Lane & R. Bachmann (Eds.), *Trust within and between organizations: Conceptual issues and empirical applications* (pp. 1-30). Oxford: Oxford University Press.

Lane, C., & Bachmann, R. (1996). The social constitution of trust: Supplier relations in Britain and Germany. *Organization Studies*, *17*(3), 365-395.

Lee, P.-M. (2002). Behavioral model of online purchasers in e-commerce environment. *Electronic Commerce Research*, *2*(1-2), 75-85.

Lim, K. H., Sia, C. L., Lee, M. K. O., & Benbasat, I. (2006). Do I trust you online, and if so, will I buy? An empirical study of two trust-building strategies. *Journal of Management Information Systems*, *23*(2), 233-266.

Luo, Y. (2007). Are joint venture partners more opportunistic in a more volatile environment? *Strategic Management Journal*, *28*(1), 39-60.

Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American Sociological Review*, *28*(1), 55-67.

Mackie, G. (2001). Patterns of trust in Western Europe and their genesis. In K. S. Cook (Ed.), *Trust in Society* (pp. 245-281). New York: Russell Sage Foundation.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, L. (2001). Ethnic and national stereotypes: The Princeton trilogy revisited and revised. *Personality and Social Psychology Bulletin*, *27*(8), 996-1010.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management. *Journal of applied psychology*, *84*(1), 123-136.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.

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

McEvily, B., Perrone, V. & Zaheer, A. (2003). Trust as an organizing principle. *Organization Science*, *14*(1), 91-103.

McKnight, D. H., & Chervany, N. L. (2001). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, *6*(2), 35-59.

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

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002b). The impact of initial consumer trust on intentions to transact with a web site: A trust building model. *Journal of Strategic Information Systems*, *11*(3-4), 297-323.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review*, *23*(3), 473-490.

Mcknight, D. H., Kacmar, C. J., & Choudhury, V. (2004). Shifting factors and the ineffectiveness of third party assurance seals: A two-stage model of initial trust in a web business. *Electronic Markets*, *14*(3), 252-266.

Min, H. (1994). International supplier selection: A multi-attribute utility approach. *International Journal of Physical Distribution & Logistics Management*, *24*(5), 24-33.

Min, H., & Galle, W. (1993). International negotiation strategies of U.S. purchasing professionals. *Journal of Supply Chain Management*, *29*(3), 40-50.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationships marketing. *Journal of Marketing*, *58*(3), 20-38.

Mukhopadhyay, T., Kekre, S., & Kalathur, S. (1995). Business value of information technology: A study of electronic data interchange. *MIS Quarterly*, *19*(2), 137-156.

Nayyar, P. R. (1990). Information asymmetries: A source of competitive advantage for diversified service firms. *Strategic Management Journal*, *11*(7), 513-519.

Ou, C. X., & Sia, C. L. (2010). Consumer trust and distrust: An issue of website design. *International Journal of Human-Computer Studies*, *68*(12), 913-934.

Parkhe, A. (1998). Understanding trust in international alliances. *Journal of World Business*, *33*(3), 219-240.

Pavlou, P. A. (2002). Institution-based trust in interorganizational exchange relationships: The role of online B2B marketplaces on trust formation. *Journal of Strategic Information Systems*, *11*(3), 215-243.

Pavlou, P. A., & 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*(4), 392-414.

Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, *15*(1), 37-59.

Pavlou, P. A., & Ratnasingam, P. (2003). Technology trust in B2B electronic commerce: conceptual foundations. In K. Kangas (Ed.), *Business strategies for information technology management* (pp. 200-215). Hershey, PA: IGI Publishing.

Pennington, R., Wilcox, D. H., & Grover, V. (2003). The role of system trust in business-to-consumer transactions. *Journal of Management Information Systems, 20*(3), 197-226.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: problems and prospects. *Journal of Management*, *12*(4), 531-544.

Poppo, L., Zhou, K. Z., & Ryu, S. (2008). Alternative origins to interorganizational trust: An interdependence perspective on the shadow of the past and the shadow of the future. *Organization Science, 19*(1), 39-55.

Ratnasingam, P. (2005). Trust in inter-organizational exchanges: A case study in business to business electronic commerce. *Decision Support Systems, 39,* 525-544.

Robbobb. (2007, July 21). Alibaba Gold Member scam [Online forum comment, post 5]. Retrived from http://www.thewholesaleforums.co.uk/forum/bad-supplier-experiences/15465-alibaba-gold-member-scam.html#post110823

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Schafer, J., & Graham, J. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Schaffer, M. E., & Stillman, S. (2010). Xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor. *Ideas.* Retrieved from http://ideas.repec.org/c/boc/bocode/s456779.html.

Schermelleh-Engel, K., Werner, C., Klein, A., & Moosbrugger, H. (2010). Nonlinear structural equation modeling: Is partial least squares an alternative? *AStA Advances in Statistical Analysis, 94*(2), 167-184.

Scully, J. I., & Fawcett, S. E. (1994). International procurement strategies: Challenges and opportunities for the small firm. *Production and Inventory Management Journal*, *35*(2), 39-46.

Selnes, F. (1998). Antecedents and consequences of trust and satisfaction in buyer-seller relationships. *European Journal of Marketing*, *32*(3/4), 350-322.

Senn, J. A. (2000). Business-to-business e-commerce. *Information Systems Management*, *17*(2), 1-10.

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research*, *15*(3), 325-343.

Spence, M. A. (1973). Job market signaling. *Quarterly Journal of Economics*, *87*(3), 355-374.

Stewart, K. J. (2003). Trust transfer on the world wide web. *Organization Science, 14*(1), 5-17.

Sun, H. (2010). Sellers' trust and continued use of online marketplaces. *Journal of the Association for Information Systems*, *11*(4), 182-211.

Trifts, V., & Häubl, G. (2003). Information availability and consumer preference: Can online retailers benefit from providing access to competitor price information? *Journal of Consumer Psychology*, *13*(1/2), 149-159.

Vandenberg, R. J., & Grelle, D. (2009). Alternative model specifications in structural equation modeling: Facts, fictions, and truth. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences* (pp. 167-193). New York, NY: Routledge.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*(2), 249-268.

Wicks, A. C., Berman, S. L., & Jones, T. M. (1999). The structure of optimal trust: Moral and strategic implications. *Academy of Management Review*, *24*(1), 99-116.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Xiao, Z., & Tsui, A. S. (2007). When brokers may not work: The cultural contingency of social capital in chinese high-tech firms. *Administrative Science Quarterly*, *52*(1), 1-31.

Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, *9*(2), 141-159.

Zaheer, S., & Zaheer, A. (2006). Trust across borders. *Journal of International Business Studies*, *37*(1), 21-29.

Zhou, Z. (2008). Be vigilant of questionable dairy product returning to market. *Southern Daily*. Retrieved from http://epaper.nfdaily.cn/html/2008-10/09/content_6692027.htm

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840-1920. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (pp. 53-111). Greenwich, CT: JAI Press.

Appendices

Appendix 2.A

Table 2.A-1 Characteristics of Buyers
(Location, Product Category, Company size, and E-Commerce Experience)

Buyer's continent	*n*	Product category[a]	*n*	No. of employees	*n*
Africa	46	Agriculture & Food	17	1 to 4	56
Americas	21	Apparel & Accessories	11	5 to 9	68
Asia	168	Arts & Crafts	4	10 to 10	61
Europe	24	Auto parts & Accessories	8	20 to 99	65
Oceania	28	Bags, Cases, & Boxes	1	100 to 499	24
		Chemicals	32	500 or more	13
		Computer Products	13		
		Construction & Decoration	17	Sales in previous year	*n*
		Consumer Electronics	17	Less than US$100,000	55
		Electrical & Electronics	28	US$100,000 to US$499,000	71
		Furniture & Furnishing	2	US$500,000 to US$999,000	40
		Health & Medicine	18	US$1 million to US$4.9 million	81
		Lights & Lighting	6	US$5 million to US$9.9 million	15
		Machinery	31	US$10 million to US$49.9 million	12
		Metallurgy, Mineral, & Energy	21	US$50 million to US$99.9 million	5
		Office Supplies	5	US$100 million or more	8
		Security & Protection	1		
		Sporting Goods & Recreation	4	B2B e-commerce experience	*n*
		Textile	13	1 year or less	54
		Tools & Hardware	3	Between 1 and 3 years	85
		Toys	4	Between 3 and 5 years	52
		Transportation	4	More than 5 years	96
		Others	27		

Note: [a] based on the imminent purchases which respondents considered during the survey.

Appendix 2.B

Table 2.B-1 Survey Measures

Construct	Item[a]
Purchase intention	Is it likely that you would buy from *Supplier X* for the purchase that you are thinking about?
Buyer's trust	To what extent do you agree with the following statements: *Supplier X* is very capable of performing its job. I am confident about *Supplier X*'s capabilities. *Supplier X* is well qualified. *Supplier X* would not knowingly do anything to hurt me. *Supplier X* really looks out for what is important to me. *Supplier X* will go out of its way to help me. I never have to wonder whether *Supplier X* will stick to its word. *Supplier X* tries to be fair in dealings with others. Sound principles seem to guide the *Supplier X*'s behavior. *Supplier X* can be trusted.
National integrity	In your opinion, how likely would suppliers from *Supplier X*"s country do the following: Behave with integrity Do the right things in business deals always, even when no one is watching
Legal structure	How confident are you with the legal system in *Supplier X*'s country? In your opinion, how effective are the laws and regulations in *Supplier* X's country concerning the following activities: Governing operations of the suppliers Resolving business disputes
Supplier verification	What is the verification status of *Supplier X* in the exchange?
Past transactions	How many times has your company purchased from *Supplier X* in the last 12 months?
Price performance	How does *Supplier X* compare to other suppliers in terms of price?
Product performance	How does *Supplier X* compare to other suppliers in terms of product availability?
Delivery performance	How does *Supplier X* compare to other suppliers in terms of delivery?
Non-member	Is *Supplier X* listed in the B2B Exchange? b
Paid member	What is the membership type of *Supplier X* in the exchange?
Same country	Are *Supplier X* and buyer from the same country? b
China supplier	Is *Supplier X* based in China? b

[a] "*Supplier X"* in the question stems is the supplier's company name or initial as provided by the buyer.
[b] Not an actual item in the survey – the value is obtained from the buyer's responses to items concerning the buyer's and the suppliers' demographic profiles.

Appendix 2.C

Table 2.C-1 Average National Integrity, Legal Structure, and Corruption Perception Index (CPI) Scores of Supplier's Country

Country	National integrity	Legal structure	CPI	n
Armenia	5.0	4.3	3.0	1
Australia	4.0	3.7	8.6	4
Austria	4.0	4.0	8.1	1
Bahrain	5.0	5.0	5.0	1
Belgium	3.5	4.6	7.1	3
Benin	3.0	3.3	2.7	1
Botswana	2.5	2.3	5.4	1
Brazil	3.2	2.1	3.5	3
Bulgaria	3.0	3.0	4.1	1
Canada	3.5	3.5	8.7	2
China	3.3	2.8	3.5	403
Congo, Dem. Rep.	3.0	2.3	1.9	1
Denmark	5.0	4.3	9.4	1
Egypt	3.0	2.0	2.9	1
Finland	5.0	4.7	9.4	1
France	4.5	4.3	7.3	1
Germany	4.1	4.2	7.8	9
Hong Kong	3.5	3.7	8.3	10
India	3.4	3.3	3.5	34
Indonesia	3.4	2.5	2.3	4
Iran	3.0	2.3	2.5	1
Italy	3.5	3.7	5.2	2
Japan	4.0	3.7	7.5	1
Kuwait	4.0	4.0	4.3	1
Malaysia	2.8	2.8	5.1	2
Mexico	3.5	3.0	3.5	1
New Zealand	5.0	4.3	9.4	1
Nigeria	3.3	3.7	2.2	9
Norway	4.5	4.3	8.7	1
Pakistan	3.5	2.9	2.4	3
Peru	4.0	3.7	3.5	1
Qatar	4.5	4.7	6.0	1
Romania	4.0	4.3	3.7	1
Russia	4.5	3.9	2.3	5
Saudi Arabia	3.0	3.0	3.4	1
Singapore	3.0	3.7	9.3	2
Slovenia	4.0	2.3	6.6	1

South Africa	4.5	4.0	5.1	2
South Korea	3.5	4.0	5.1	3
Spain	5.0	4.3	6.7	1
Switzerland	4.5	3.0	9.0	1
Taiwan	3.9	3.2	5.7	13
Thailand	5.0	4.0	3.3	1
Turkey	4.8	4.5	4.1	4
Ukraine	4.2	4.2	2.7	3
United Arab Emirates	3.5	3.2	5.7	4
United Kingdom	3.0	4.2	8.4	4
United States	4.0	3.7	7.2	19
Vietnam	3.0	2.7	2.6	1
Zambia	3.0	2.7	2.6	1

Note: The items that measure perceived national integrity and legal structure are on a scale of 1 (least favorable) to 5 (most favorable). The CPI index score ranges from 0 (high corruption) to 10 (low corruption).

Appendix 2.D Preliminary Analyses

Our survey respondents were from various countries in different continents. We used a Kruskal-Wallis test to assess whether there are differences among respondents from different countries. We found no significant differences in terms of company's sales ($H = 21.66$, $p > .10$), respondent's working experience ($H = 18.56$, $p > .10$), and B2B e-commerce experience ($H = 13.93$, $p > .10$) among respondents from different countries. However, the number of employees in the respondent's company ($H = 46.33$, $p < .01$) and their education ($H = 52.10$, $p < .01$) are statistically different, which reflects differences in economic and social structures across respondents' countries. Based on these results, we believe it is reasonable to pool our respondents across countries.

Next, we assessed the presence of common method variance in our data. First, we conducted Harman's one-factor test and found that the scale items load onto multiple factors (Podsakoff & Organ, 1986). Second, we compared (1) the covariance of buyer's trust in and likelihood of purchasing from one supplier (within-supplier covariance), and (2) the covariance of buyer's trust in one supplier and likelihood of purchasing from the other supplier (between-supplier covariance). The within-supplier covariance was at least 1.5 times the between-supplier covariance, which indicates that common method variance does not appear to account for the relationship between buyer's trust and purchase intention. The within-supplier covariance of buyer's trust in and likelihood of purchasing from the first supplier was .28; the between-supplier covariance of buyer's trust in the second supplier and the likelihood of purchasing from the first supplier was .18. The within-supplier covariance of buyer's trust in and likelihood of purchasing from the second supplier was .32; the between-supplier covariance of buyer's trust in the first supplier and the likelihood of purchasing from the second supplier was .11.

Appendix 2.E

Table E-1. Item-Level Descriptive Statistics and Correlation Matrix for Structural Equation Modeling

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Purchase intention	1.00																		
2	Trust-A^	0.35	1.00																	
3	Trust-B^	0.37	0.86	1.00																
4	Trust-C^	0.37	0.84	0.85	1.00															
5	National integrity 1*	0.18	0.24	0.28	0.27	1.00														
6	National integrity 2*	0.19	0.22	0.26	0.24	0.53	1.00													
7	Legal S=structure 1*	0.11	0.15	0.24	0.21	0.27	0.36	1.00												
8	Legal structure 2*	0.17	0.24	0.29	0.29	0.35	0.47	0.54	1.00											
9	Legal structure 3*	0.16	0.25	0.30	0.28	0.38	0.48	0.55	0.87	1.00										
10	CPI*	0.03	0.10	0.13	0.13	0.09	0.18	0.22	0.26	0.26	1.00									
11	Supplier verification	0.11	0.08	0.13	0.11	0.07	0.12	0.19	0.13	0.15	-0.05	1.00								
12	Past Transactions*	0.08	0.23	0.19	0.20	0.06	0.09	0.10	0.04	0.05	0.16	-0.02	1.00							
13	Price performance	0.28	0.27	0.32	0.32	0.11	0.13	0.13	0.17	0.17	0.10	0.07	0.13	1.00						
14	Product performance	0.32	0.41	0.46	0.44	0.17	0.17	0.18	0.19	0.19	0.11	0.07	0.14	0.52	1.00					
15	Delivery performance	0.31	0.46	0.49	0.48	0.18	0.17	0.17	0.19	0.22	0.14	0.04	0.14	0.43	0.73	1.00				
16	Non-member	-0.02	0.09	0.08	0.10	0.06	0.09	0.08	0.11	0.11	0.37	-0.20	0.15	-0.02	0.03	0.04	1.00			
17	Paid member	-0.02	-0.10	-0.07	-0.10	-0.04	-0.01	0.01	-0.06	-0.05	-0.11	0.38	-0.08	-0.01	-0.05	-0.03	-0.24	1.00		
18	Same country	0.01	0.06	0.03	0.02	0.02	0.08	0.21	0.10	0.11	-0.06	0.00	0.24	-0.06	-0.06	-0.05	0.19	-0.04	1.00	
19	China supplier	-0.04	-0.12	-0.12	-0.16	-0.10	-0.19	-0.30	-0.27	-0.26	-0.55	0.06	-0.25	-0.04	-0.08	-0.08	-0.51	0.18	-0.41	1.00
	Mean	3.79	3.78	3.88	3.79	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	3.67	3.76	3.69	0.24	0.16	0.09	0.70
	Variance	1.10	0.58	0.66	0.64	1.11	1.44	1.44	1.43	2.18	2.18	0.10	1.65	1.04	0.90	0.95	0.18	0.13	0.08	0.21

Notes: ^: We assigned the ten survey items that measure buyer's trust into three parcels: Trust-A (4 items), Trust-B (3 items), and Trust-C (3 items).
*: Mean-centered.

Appendix 2.F

Table 2.F-1 Measurement Model Results

Construct	Item	Loading	Std. error	Cronbach's α
Buyer's trust	Trust-A	0.92	0.01	.94
	Trust-B	0.93	0.01	
	Trust-C	0.91	0.01	
National integrity	National Integrity 1	0.66	0.07	.69
	National Integrity 2	0.80	0.06	
Legal structure	Legal Structure 1	0.59	0.05	.74
	Legal Structure 2	0.94	0.02	
	Legal Structure 3	0.92	0.02	
	CPI	0.28	0.04	
Supplier's performance	Price Performance	0.57	0.06	.81[a]
	Product Performance	0.87	0.03	
	Delivery Performance	0.83	0.04	

Chi-square = 55.20, d.f. = 48, $p > .10$
CFI = 0.99 RMSEA = 0.02 SRMR = 0.03
All item loadings are significant at $p < .01$
[a] Using casewise deletion for missing data (i.e. "not sure" responses)

Appendix 2.G Robustness Analyses

Random Effects Model. We estimated random effects models with buyer's trust as dependent variable using Stata (version 10.1). We assumed that the unobserved effects affect a buyer's evaluations of both suppliers in the same manner. For instance, the (unobserved) importance of situational normality or structural assurance to a buyer should be the same for every supplier whom they were considering. This assumption justifies the use of random effects model to account for unobserved effects (Wooldridge, 2002) and minimizes the need to include control variables for buyer attributes in our models. We standardized items that measured legal structure and national integrity and then averaged to form the respective variables. We also standardized the item for the number of past transactions. We recoded the responses for each supplier's relative performance criterion into three dummy variables: (1) better than other suppliers, (2) worse than other suppliers, and (3) not sure about the performance, with the base category being the neutral "equal to other suppliers" response. We clustered the observations by respondent to obtain robust variance estimates (Wooldridge, 2002). Tables G-1 and G-2show the descriptive statistics and random effects model results, respectively.

Table G-1. Descriptive Statistics and Correlation Matrix for Random Effects Model

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Buyer’s trust	1.00																	
2	National integrity	0.30	1.00																
3	Legal structure	0.30	0.48	1.00															
4	Supplier verification	0.11	0.11	0.14	1.00														
5	Past transactions	0.22	0.09	0.12	-0.02	1.00													
6	Price performance (better)	0.28	0.10	0.18	0.09	0.19	1.00												
7	Price performance (worse)	-0.17	-0.09	-0.05	-0.02	-0.03	-0.37	1.00											
8	Price performance (not Sure)	-0.17	-0.05	-0.02	0.01	-0.24	-0.44	-0.14	1.00										
9	Product performance (better)	0.40	0.15	0.19	0.09	0.21	0.45	-0.17	-0.30	1.00									
10	Product performance (worse)	-0.32	-0.11	-0.11	0.03	0.10	-0.17	0.39	-0.08	-0.25	1.00								
11	Product performance (not Sure)	-0.24	-0.09	-0.08	-0.02	-0.23	-0.27	-0.10	0.65	-0.45	-0.10	1.00							
12	Delivery performance (better)	0.45	0.17	0.22	0.03	0.31	0.40	-0.11	-0.29	0.65	-0.16	-0.35	1.00						
13	Delivery performance (worse)	-0.28	-0.11	-0.12	0.01	0.04	-0.12	0.26	-0.11	-0.20	0.55	-0.11	-0.25	1.00					
14	Delivery performance (not sure)	-0.21	-0.09	-0.07	0.00	-0.35	-0.24	-0.10	0.59	-0.28	-0.11	0.68	-0.50	-0.15	1.00				
15	Non-member	0.10	0.09	0.22	-0.20	0.15	-0.04	0.06	-0.02	0.04	0.05	-0.06	0.08	0.03	-0.10	1.00			
16	Paid member	-0.10	-0.03	-0.07	0.38	-0.08	-0.02	0.05	0.05	-0.07	0.09	0.04	-0.08	0.00	0.02	-0.24	1.00		
17	Same country	0.04	0.06	0.12	0.00	0.24	-0.01	0.16	-0.08	-0.01	0.16	-0.05	0.04	0.11	-0.10	0.19	-0.04	1.00	
18	China dupplier	-0.04	-0.16	-0.45	0.06	-0.25	-0.05	-0.07	0.03	-0.08	-0.06	0.04	-0.15	-0.05	0.10	-0.51	0.18	-0.41	1.00
	Mean	3.81	0.00	0.00	0.12	0.00	0.53	0.11	0.15	0.53	0.05	0.15	0.45	0.07	0.23	0.24	0.16	0.09	0.70
	Minimum	1.00	-2.17	-1.38	0	-0.68	0	0	0	0	0	0	0	0	0	0	0	0	0
	Maximum	5.00	1.36	2.00	1	3.21	1	1	1	1	1	1	1	1	1	1	1	1	1
	Variance	0.56	0.76	0.59	0.10	1.00	0.25	0.10	0.13	0.25	0.05	0.13	0.25	0.06	0.18	0.18	0.13	0.08	0.21

Table 2.G-2.Random Effects Model Results

Dependent Variable: Buyer's Trust	Model 1	Model 2	Model
	Coeff.	Coeff.	Coeff.
γ_1: Constant	3.69** (0.10)	3.69** (0.10)	3.69** (0.10)
γ_2: National integrity	0.14** (0.04)	0.13** (0.04)	0.13** (0.04)
γ_3: National integrity x past transaction	-	-0.01 (0.05)	-0.01 (0.04)
γ_4: Legal structure	0.06 (0.04)	0.08+ (0.04)	0.08+ (0.04)
γ_5: Legal structure x past transactions	-	-0.06* (0.03)	-0.06* (0.03)
γ_6: Supplier verification	0.23* (0.11)	0.23* (0.11)	0.24* (0.11)
γ_7: Supplier verification x past transactions	-	-	0.08 (0.08)
γ_8: Past transactions	0.08* (0.03)	0.09** (0.03)	0.09** (0.03)
γ_9: Relative price (better)	0.05 (0.07)	0.06 (0.06)	0.06 (0.06)
γ_{10}: Relative price (worse)	-0.04 (0.11)	-0.03 (0.11)	-0.03 (0.11)
γ_{11}: Relative price (not sure)	-0.04 (0.11)	-0.04 (0.11)	-0.04 (0.11)
γ_{12}: Relative product (better)	0.02 (0.07)	0.02 (0.07)	0.03 (0.07)
γ_{13}: Relative product (worse)	-0.69** (0.20)	-0.70** (0.20)	-0.69** (0.20)
γ_{14}: Relative product (not sure)	-0.26* (0.10)	-0.26* (0.10)	-0.25* (0.10)
γ_{15}: Relative delivery (better)	0.35** (0.07)	0.34** (0.07)	0.33** (0.07)
γ_{16}: Relative delivery (worse)	-0.22 (0.16)	-0.22 (0.16)	-0.23 (0.16)
γ_{17}: Relative delivery (not sure)	0.09 (0.09)	0.10 (0.09)	0.09 (0.09)
γ_{18}: Non-exchange member	0.03 (0.07)	0.03 (0.07)	0.04 (0.07)
γ_{19}: Paid membership	-0.11 (0.09)	-0.10 (0.09)	-0.10 (0.09)
γ_{20}: Same country	0.05 (0.12)	0.07 (0.12)	0.06 (0.12)
γ_{21}: China supplier	-0.03 (0.07)	-0.03 (0.07)	-0.02 (0.07)
R-sq (within)	0.27	0.29	0.29
R-sq (between)	0.40	0.39	0.39
R-sq (overall)	0.36	0.36	0.36
Wald χ^2 (d.f.)	233.36 (17)	243.19 (19)	246.99 (20)
Prob > χ^2	0.00	0.00	0.00
Sargan-Hansen statistic	17.61	25.42	25.37
p-value	0.41	0.15	0.19

+ $p < .10$ * $p < .05$ ** $p < .01$
Note: Robustness standard errors in parentheses.

We estimated a baseline model with only the main-effects (Model 1), and a model that included the national integrity and legal structure interaction terms (Model 2). The Sargan-Hansen statistic does not reject the null hypothesis that the orthogonality assumption is valid for both models (test statistic = 17.61, p-value = .41 for Model 1; test statistic = 25.42, p-value = .15 for Model 2). This implies that the random effects estimator is consistent[8].

The results using SEM (Model 2 in Table 3) and random effects models (Model 2 in Table G-2) are consistent with each other. Perception of national integrity was positively related to buyer's trust ($\gamma_2 = .13$, $p < .01$), but this relationship was not moderated by past transactions ($\gamma_3 = -.01$, $p > .10$). These results support H1 but not H4a. We also find that past transactions negatively moderate the relationship between perception of legal structure and buyer's trust ($\gamma_4 = .08$, $p < .10$ and $\gamma_5 = -.06$, $p < .05$). Therefore, H2 and H4b are supported. Similarly, supplier verification had a positive impact on buyer's trust ($\gamma_6 = .23$, $p < .05$), which supports H3.

We also examined whether the relationship between supplier verification and buyer's trust was moderated by past transactions. We added the interaction term "supplier verification x past transactions" in Model 3 (Table G-2), and found that it was not statistically significant ($\gamma_7 = .08$, $p > .10$). Once again, there is no evidence that supplier verification becomes less important to a buyer with increased transactions.

Archival Data. Of the 574 suppliers in our dataset, 439 were listed in the B2B exchange. We could uniquely identify 194 (44.2 percent) of these suppliers in the exchange's online directory, and obtain information about their membership type and third-party verification status. Excluding cases in which buyers were "not sure" about suppliers' membership types or verification status, 61 percent of the membership type indicator and 66 percent of the verification status indicator in our dataset match the information from the exchange[9]. These are conservative estimates of correct matches because membership types and verification status could have changed between the time the buyer saw the signals and the time we obtained the information on

---

[8] If the orthogonality assumption is not satisfied (i.e., unobserved effects do correlate with independent variables), random effects estimators are not consistent but fixed effects estimators are. However, if the orthogonality assumption is satisfied, random effects estimators are consistent and also more efficient than fixed effects estimators. The Sargan-Hansen test statistic is a heteroskedastic- and cluster-robust form of Hausman test that compares random effects and fixed effects estimators (Schaffer & Stillman, 2010). Failure to reject the null hypothesis in the Sargan-Hansen test (as in our case) implies using random effects is appropriate.

[9] Using logistic regression, we found that the odds that a supplier was a paid member according to the exchange's data are three times larger when a buyer indicated that the supplier was a paid member than when he indicated that the supplier was not. The predicted probability that a supplier was a paid member based on the exchange's data was .57 when a buyer indicated that the supplier was a paid member, and .30 when they indicated that the supplier was not. We could not analyze supplier verification data as buyers "predicted" non-verified status perfectly (i.e. when they indicated that a supplier was not verified, this was so according to the exchange's data). Thus, supplier membership and verification status seem to be salient to buyers, and our survey data is reasonably robust.

the exchange. We checked the robustness of our results by using the B2B exchange's data in two cases: (1) whenever exchange and survey data differed, and (2) only when buyers indicated that they were not sure of a supplier's membership type or verification status. In both cases, the results from the recoded dataset are qualitatively similar to those from the original dataset.

Alternative SEM Specifications. We checked the robustness of our structural model to changes in specifications of the latent Buyer's Trust variable. There are two ways to re-specify buyer's trust. First, we could directly set the ten survey items that measure trust as indicators of buyer's trust instead of assigning them to parcels trust-A, trust-B, and trust-C. Second, we could impose a hierarchical structure for the construct buyer's trust. This is achieved by modeling buyer's trust as a second-order factor with presumed direct causal effects on three first-order factors, and then assigning the ten survey items to these first-order factors. We estimated our models using these two re-specifications of buyer's trust and obtained results that are similar to our original findings in both cases. Both the re-specified models have adequate model fit. The fit indices for the model with the first re-specification were CFI = .93, RMSEA = .05, SRMR = .04. The fit indices for the model using the second re-specification were CFI = .94, RMSEA = .04, SRMR = .04.

We also re-specified the measurement model for the latent legal structure variable by using the three survey items and removing CPI from our model. The model fit of the resulting structural model was adequate (CFI = .99, RMSEA = .02, SRMR = .03) and our original results still hold qualitatively. We then dropped legal structure from our model and used CPI as the sole proxy for buyers' perceptions of the legal system in suppliers' country. The estimates for CPI and interaction between CPI and past transactions are not significant statistically. Therefore, it is important to measure perceptions of legal structure at the individual buyers' level and not rely on global indices such as CPI alone.

Next, we checked the sensitivity of the SEM results to lower assumed reliability of indicators when fixing the error variance of single-indicator variables. We reduced the assumed reliability of these variables from .85 to .75 and re-estimated Model 2. The interaction between perceived legal structure and past transactions was then weakly supported at .10 level (β_5 = -.11, p = .054), while the other results continued to hold at same level of statistical significance.

Multilevel Mixed Effects Model. Because the observations of suppliers are nested within buyers, we used multilevel mixed effects model to check the robustness of our results. In the multilevel model, we allowed random coefficients on national integrity, legal structure, supplier verification and past transactions at the buyer level. The multilevel mixed effects results are consistent with the random effects model results.

CHAPTER 3: MULTI-HOMING USERS' PREFERENCE FOR TWO-SIDED EXCHANGE NETWORKS

3.1 Introduction

Two-sided networks are platforms that facilitate interactions between distinct but inter-dependent groups of users, such as buyers and suppliers (Rochet and Tirole, 2003).[10] These platforms are an important trend in today's business environment (Bughin et al., 2010). Various new ventures are based on the two-sided network model (e.g. online exchanges, social lending clubs, "deal-of-the-day" websites). Moreover, some merchants are introducing platform-type services that connect their buyers with other sellers. For example, not only does Amazon sell directly to customers, it now also provides a platform (Amazon Marketplace) for buyers and sellers to transact directly with each other.

With the proliferation of online exchanges, buyers and sellers often participate on several platforms at the same time to fulfill a particular task. For example, a buyer can post purchase requests for a particular product on competing Business-to-Business (B2B) exchanges (see Appendix 3.A). Sellers also use multiple auction platforms to get greater market exposure and reach a larger variety of buyers (Walczak et al., 2006). We say a user is *multi-homing* when she participates on multiple competing platforms concurrently (Armstrong, 2006; Choi, 2010; Rochet and Tirole, 2003). Multi-homing behaviors have important impacts on two-sided platform competition and strategy. Consider an extreme case where all buyers multi-home on two platforms. Suppliers can participate in just one of the platforms to access all the buyers. In

[10] In this paper, we use the terms exchange, network, and platform interchangeably.

this case, multi-homing behavior by buyers intensifies competition among platforms to attract suppliers and affects their pricing strategy (Rochet and Tirole, 2003).

To gain traction and avoid considerable costs of failure, platforms have to achieve high participation levels among their users (Eisenmann et al., 2006; Tucker and Zhang, 2010). In this study, we examine drivers of multi-homing buyers' participation on and preferences for B2B exchanges. We investigate how selling and buying activities on B2B exchanges affect market and social dynamics that multi-homing buyers experience, and impact these users' participation on competing platforms. Our results show that selling activity levels on the platforms have positive effects on multi-homing buyers' preference, whereas buying activity levels can have positive or negative effects. The latter result is a consequence of the joint operation of social proof and between-buyer competition. At very low buying levels, buyers have low willingness to participate because they are uncertain about the quality of the platform. As buying increases, buyers become more inclined to participate as other buyers' participation signals a positive evaluation of the exchange. Beyond a certain buying level, however, further increases in buying on the exchange lead to stiffer competition among buyers, raising prices and discouraging buyers from participating. As such, the relationship between multi-homing buyers' preferences and buying activities on an exchange follows an inverted-U shaped curve. We also find that the impacts of buying activities on buyers' preferences attenuate with increasing platform experience over time. As buyers gain hands-on experience on the platforms, they rely less on other buyers' evaluations. In addition, with time, their network of suppliers on the platforms expands and helps to lessen the negative competition effects from increased buying activities on the exchanges.

We also look at how neglecting users' multi-homing behaviors affects implications for platforms' strategy and competitive actions. Due to constraints in collecting data on competing platforms, researchers and practitioners may misspecify users' behaviors in their analyses when they assume users are single-homing. Our research design overcomes this restriction and allows us to compare results between correctly and incorrectly specified models. We find that incorrect assumptions about users' behaviors bias the picture of competitive dynamics between platforms.

This study contributes to platform research in a few ways. First, relatively few studies explicitly examine the multi-sided nature of these platforms or examine the interactions within and across different sides on the platforms (e.g., Belleflamme and Toulemonde, 2009; Tucker and Zhang, 2010). Successful strategies for traditional markets may not be as effective in two-sided network markets (Eisenmann et al., 2006). Some insights from one-sided networks may not be applicable in two-sided network contexts given the social and market dynamics that take place across and within the distinct sides of the platforms. For instance, the relationship between price and cost on two-sided platforms is complex, such that optimal prices depend on demand elasticity on both sides of the platform and the profit-maximizing price may be below marginal cost (Chandra and Collard-Wexler, 2009; Evans and Schmalensee, 2007).

Second, we focus on multi-homing behaviors, whereas most studies on two-sided platforms look at *single-homing* scenarios where users participate on only one platform. For instance, Tucker and Zhang (2010) examined how the number of buyers and sellers on an online exchange affects sellers' posting of product listings on this exchange. Although multi-homing behaviors have been discussed in two-sided networks research (e.g. Armstrong, 2006; Rochet and Tirole, 2006), little empirical work has looked at such behaviors. One exception is Jin and Rysman's (2010) study of dealers' multi-homing behaviors in sportcards conventions. Jin and Rysman did not have direct data on whether dealers single-home or multi-home; instead, they inferred dealers' homing decisions from (i) convention prices and (ii) the interaction of distances between conventions and dealers costs to participate on multiple conventions. In contrast, we directly observed multi-homing buyers' participation on two exchanges over 7 months, which allows us to make robust inferences of buyers' behaviors on the platforms.

Finally, by investigating how non-price factors affect buyers' platform usage over time, this study complements existing research that examines various pricing structures on platforms and their impacts on users' participation (e.g., Armstrong, 2006; Caillaud and Jullien, 2003; Jin and Rysman, 2010; Rochet and Tirole, 2003, 2006), and those that focus on sellers' behaviors on the platforms (e.g., Jin and Rysman, 2010; Tucker and Zhang, 2010).

3.2 Business-to-Business Exchanges

Firms participate on B2B exchanges such as Alibaba.com, ECEurope.com and ECPlaza.com to connect to potential buyers and suppliers. These platforms serve as market-aggregators, -makers and -facilitators (Bakos, 1998; Dai and Kauffman, 2002), and increase the pool of trading partners by creating centralized marketplaces (Spulber, 1999). B2B exchanges also help firms extend their reach globally (Koh et al., forthcoming), and reduce search costs through information discovery and price matching services (Lucking-Reiley and Spulber, 2001).

To reach trading partners on B2B exchanges, firms broadcast their business needs and offers by posting buying requests and product listings, respectively. Buying requests (or requests for quotations) provide details of buyers' purchase requirements, while product listings describe items that suppliers are selling. These broadcasts are important as they provide information of market opportunities to platform users, and many B2B exchanges highlight the numbers of buying requests and product listings to attract new users (see Tucker and Zhang, 2010).

The main source of revenue for most B2B exchanges is supplier membership fees. For example, almost 90% of Alibaba.com's US$995.7 million in revenue is from membership packages.[11] B2B exchanges usually provide free memberships to buyers, while allowing suppliers to choose between free or paid memberships. Buyers can create a company profile, post buying requests, and view suppliers' product listings on the platforms. Suppliers having a free membership can create company profile, post product listings, and view buyers' information, whereas those who upgrade to paid memberships enjoy additional and/or enhanced services such as priority listings in online directories, and higher limits in the number of product listings that they can post. Most B2B exchanges do not charge transaction-based fees because they cannot reliably observe transactions between firms (Rochet and Tirole, 2006; Roson, 2005). Fixed membership fees (in contrast to per-transaction charges) allow buyers and suppliers to retain the benefits of their transactions on the platforms, and make getting and keeping both sides on board even more critical to the platforms' success (Armstrong, 2006; Evans and Schmalensee, 2007).

As the presence of buyers helps B2B exchanges attract and generate revenues from suppliers, platforms also compete to acquire and retain buyers. In fact, competition for buyers can be intense as many buyers multi-home on various competing platforms. Buyers' participation may vary across platforms where they post more buying requests on certain platforms than on others. Therefore, B2B exchanges not only have to compete to get buyers to join their platform, but they must also get these buyers to prefer their platforms to others. Achieving higher preferences from multi-homing buyers helps exchanges have an edge over their competitors in attracting and retaining suppliers.

[11] Based on Alibaba.com's 2011 Annual Report, available at http://ir.alibaba.com/ir/home/financial_reports.htm (last accessed in April 2012). Although information on revenue sources of other B2B exchanges is not publicly available, an examination of services offered by other exchanges indicates that membership packages are key revenue drivers.

3.3 Drivers of Multi-Homing Users' Preference

Multi-homing users' preference can be inferred from their relative usage of competing platforms. All else equal, users would participate more actively on their preferred platform relative to other platform(s) in which they are participating. We assume that preference between two platforms A and B is shown when one of them is chosen more often than the other. That is, the preference for A over B is indexed by the proportion of times A is chosen over B. This operationalization of user preference is conceptually similar to classical methods for measuring preferences in psychology and marketing (Coombs et al., 1970; Thurstone, 1927). In this section, we look at how multi-homing buyers' preferences for B2B exchanges are affected by actions of other buyers and suppliers through the effects of *network externalities* and *social proof*.

3.3.1 Network Externalities

Network effects occur when the value of a product to a user depends on the number of users of that product (Katz and Shapiro, 1985). Positive network externalities arise when the product becomes more attractive as the number of users increases. Many technology-related products, such as software and communication systems, demonstrate positive network externalities (Brynjolfsson and Kemerer, 1996; Kraut et al., 1998; Tucker, 2008). Network effects can also be negative, where the value of a product decreases as the user base increases. For instance, negative network externalities occur in Peer-to-Peer networks due to consumption of scarce network resources or free riding by users in larger networks (Asvanund et al., 2004).

In two-sided networks, network externalities can be categorized as either inter-network externalities or intra-network externalities.[12] Inter-network externalities refer to how characteristics of one side of the platform affect users on the other side. For instance, the level of selling activities on a B2B exchange is an important inter-network activity for buyers, and it affects the benefits that buyers derive from the platform. In most cases, inter-network externalities on two-sided networks are positive, where the expected gain for users on one side of the platform increases with higher activity levels on the opposite side (Eisenmann et al., 2006).

Positive inter-network externalities lead to a "chicken-and-egg problem" on two-sided platforms: to attract users of one type, a platform needs to have a sufficient number of users of the other type (Caillaud and Jullien, 2003). B2B exchanges thus need to get "both sides on board," attracting buyers and suppliers to adopt their platforms simultaneously. One way to achieve this objective is to use appropriate pricing strategies (Armstrong, 2006; Caillaud and Jullien, 2003; Eisenmann et al., 2006; Rochet and Tirole, 2003, 2006). Many platforms adopt pricing structures that are heavily skewed towards one side of the market; generally, the side that enjoys greater inter-network externalities is more likely to face higher prices than the side that experiences lower inter-network externalities (Eisenmann et al., 2006; Evans, 2003). This pricing strategy is consistent with what we observe in many B2B exchanges, where suppliers face higher membership fees and buyers are heavily subsidized with free memberships.

How does inter-network activity (selling) levels affect multi-homing buyers' preferences? On the one hand, buyers typically join and use platforms for free. Also, as the required

[12] The concepts of inter-network and intra-network externalities are similar to those of indirect network and direct network externalities, respectively. The latter terminology, however, do not depend on the existence of a platform that connects distinct but inter-dependent groups of users. For instance, indirect network effects can refer to how the availability or prices of complementary goods affect the value of a main good. To emphasize the two-sided nature of the platforms that we focus on in this study, we stick to the terminology of inter- and intra-network externalities.

information in buying requests is similar across different B2B exchanges (e.g., buyer's contact information, product category, and request details), buyers frequently copy-and-paste their buying requests across multiple platforms. Hence the cost for buyers to multi-home appears relatively low, and selling activity levels may not affect buyers' preferences *after* they join the various competing exchanges. Since these buyers have already joined multiple exchanges, their strategy could be to post buying requests on as many exchanges as possible to increase their choices of suppliers in fulfilling their purchase requirements.

On the other hand, multi-homing buyers incur time and effort when they participate on various platforms. They need to manage quotations that they receive from suppliers, and update their buying request details on the exchanges. Buyers are also likely to enjoy stronger bargaining power and receive more competitive quotations on platforms with higher selling activities. As these buyers have limited resources (e.g. time, attention) and various platforms to choose from, they should prioritize and actively use beneficial platforms more. Furthermore, as they are using multiple exchanges, they are by definition not locked into a specific platform. As such, multi-homing buyers will reduce the use of "under-performing" exchanges. Hence, we hypothesize a positive relationship between selling activity levels and buyers' preferences for B2B exchanges: buyers prefer and use more of those platforms that have higher levels of selling activity.

H1: A positive relationship exists between buyers' preferences for B2B exchanges and selling activity levels on the exchanges.

The other type of externalities in two-sided network is intra-network externalities, which refer to how characteristics of one side of the platform affect users located on that same side. In our context, intra-network externalities relate to how variations in buying activity levels on the platforms affect individual buyers' benefits. While most studies of two-sided networks focus on inter-network externalities, intra-network externalities are "either abstracted away or not central to the analysis" (Belleflamme and Toulemonde, 2009: 247). We suspect that since users usually derive benefits and value on two-sided networks via interactions that occur across the platforms, there is greater interest and focus on inter-network rather than intra-network dynamics.

Among the few studies that look at intra-network externalities on exchange networks, the general conclusion is that these externalities are negative due to competition effects: as the number of users on one side of a platform increases, these users face greater competition, less attractive prices, and lower benefits on the platform (Anderson et al., 2008; Belleflamme and Toulemonde, 2009; Eisenmann et al., 2006; Tucker and Zhang, 2010). For example, the presence of many sellers has negative effects on whether a potential seller participates on the platform (Tucker and Zhang, 2010). Covisint, a B2B exchange launched by major auto manufacturers in 2000, failed to attract sufficient auto parts suppliers as these suppliers were concerned with rivalry and downward pricing pressure on the exchange (Eisenmann et al., 2006). Negative intra-network externalities suggest that multi-homing buyers' preferences for a particular platform would be inversely related to intra-network activity (buying) level on that platform (Belleflamme and Toulemonde, 2009). When too much buying is occurring on a platform, buyers are likely to seek out other platforms where they might find lower competition and prices.

However, the relationship between buyers' preferences and buying activities might not be monotonic. Competitive interactions among buyers should occur at high levels of buying activity. *Ceteris paribus*, suppliers have greater power to raise prices when there is much buying going on in the platform. At low levels of buying activity, competitive effects are relatively weaker, and a different type of process among buyers may be occurring. Specifically, a certain level of buying must take place to motivate individual buyers to participate on the platform. This phenomenon can be explained by the principle of social proof, which states that observations of other people's behaviors affect one's decisions (Cialdini, 2008), in this case by persuading prospective buyers that other buyers value participation on the platform.

3.3.2 Social Proof

Until now, we have discussed how price factors and competition affect buyers' decisions through inter- and intra-network activity levels. Nevertheless, there is *always* some residual uncertainty in market transactions. To reduce this uncertainty, buyers use various sources of information they can acquire, such as market and seller reputation, contractual safeguards, and verification. Social proof is another informative factor that influences buyers' choice and behavior.[13] Cialdini (2008: 99) points out that individuals in part "determine what is correct by finding out what other people think is correct." Thus, market behavior is a form of social proof, and how other buyers behave on B2B exchanges is informative to individual buyers.

Social proof is particularly salient under two conditions (Cialdini, 2008), both of which often occur in two-sided networks. The first is the similarity of other people to oneself. We are motivated to act like those we observe when we are similar to them. The nature of two-sided networks is such that a clear boundary separates one type of user (e.g., buyers) from another (e.g., suppliers) on the platform. Moreover, B2B exchanges often structure their platforms by

[13] Social proof is not conformity or following fashion. To be part of those who follow fashion or are "in the know" requires others to observe you; otherwise there are no incentives to engage in following fashion or fads. However, while social proof requires us to observe others, it is not necessary for us to be observed by others (Cialdini, 2008; Keizer et al., 2008). Social proof is also different from an information cascade (Bikhchandani et al., 1992) which requires (i) observation of the actions of others, and (ii) the observer's action being conditioned on that observation (making the same choice another makes because you observe their choice) while ignoring private information. While both social proof and information cascades reflect inferences from observing others' behaviors, social proof is only one piece of information in the decisions we discuss in this paper. Observing others' decisions, while informative, will not necessarily lead to a cascade and an observer will still consider private information.

industries or product categories. This structure, together with the clear distinction between user types, facilitates prospective buyers observing other buyers in their category of interest.

The second condition required for social proof is uncertainty, which causes us to draw inferences from more sources of information, such as others' behaviors. With substantial user turnover on B2B exchanges and potential frauds in online transactions, one uncertainty that buyers face concerns the platforms' quality in terms of supplier trustworthiness. For example, Alibaba.com needs to acquire 35,000 suppliers a year to maintain its revenue, as 35% of its suppliers do not renew their annual paid memberships (BusinessWeek, 2010). In 2011, more than 2,300 Alibaba.com suppliers who joined between 2009 and 2010 were found to engaged in fraud (Time, 2011), causing its market capitalization to drop by almost US$1 billion and its key executives to resign (Bloomberg, 2011). This event demonstrated the uncertainty in platform quality that buyers are exposed to, and its impacts on the market's valuation of platforms.

Observing other similar buyers on a B2B exchange provides information to a buyer when there is uncertainty about platform quality. According to the social proof principle, what matters is not just the presence or number of others people that an individual observes; rather it is *the* actions of other people that the individual observes that influence behavior (Cialdini, 2008). In our two-sided platform context, it is important to consider not the number of other users on the same side per se, but what these users do: the intra-network activity level. A buyer who observes that other buyers are actively posting buying requests on a platform is likely to be motivated to do the same (assuming he has purchase requirements to fulfill). Yet this positive social proof effect is likely to have an upper bound and increase at a diminishing rate with respect to the number of observed actions of others. The initial observations of other buyers' behaviors should provide significant informational value about the platform to a particular buyer. Beyond a certain level, additional observations would add little to what the buyer may have already inferred.

Here is how intra-network externalities and social proof work in the B2B exchange setting. A buyer has low inclination to participate on a platform when the product category that he is interested in has few buying activities going on. The lack of other buyers' participation raises his uncertainty about the platform's quality, even if it is beneficial to him when there are few competitors. As buying level in that category increases, the buyer sees evidence of quality in the buying of others and has higher willingness to participate. Although competition is higher with more buying on the platform, the positive signal from other buyers' participation matters more. However, beyond a certain buying level, the buyer will be deterred from participating as the platform turns overly crowded and competition becomes too intense. Thus, we expect a non-monotonic relationship between buyers' preferences for exchanges and buying activity levels.

> *H2: As buying activities increase, buyers' preferences for B2B exchanges (i) increase when activity level is low and (ii) decrease when activity level is beyond a certain point.*

Over time, direct participation and experience on B2B exchanges reduce buyers' uncertainty about the platforms and their reliance on other buyers' behaviors as social proof. Buyers also build up their network of suppliers on the exchanges with time. A larger network of suppliers provides buyers with more supply-side alternatives, which help to lessen the negative competition effects from increased buying activities on the platforms. Furthermore, interactions between trading partners over time promote mutual commitment and cooperation (Doney et al., 2007; Poppo et al., 2008). Suppliers on the platforms may give preferences to buyers with whom they have longer relationships, and preferential treatments from suppliers are valuable to buyers when competition increases. Thus, although competitive buying activities on the platforms should affect multi-homing buyers' preferences for B2B exchanges over time, we hypothesize that the influences of these activities weaken as buyers' experience on the platforms increases.

> *H3: The impacts of changes in buying activity levels on buyers' preferences for B2B exchange attenuate over time as their platform experience increases.*

3.4 Data

We observed buyers' usage on two B2B exchanges between July 2009 and February 2010. These platforms have similar Alexa site popularity rankings.[14] (We could not collect data from the most popular B2B exchange according to Alexa rankings, as that exchange blocked certain information from non-registered users. This constraint would have affected our identification of multi-homing buyers, as we discuss below. However, not using data from the most popular B2B exchange is not a problem since our theory and models explicitly account for site usage, which are indicators of site popularity.) Both platforms cover multiple industries such as chemicals, computers, and electronics. Each industry on the exchanges is further segmented into various product categories. For instance, the chemical industry segments include inorganic and pharmaceutical chemicals. Majority of companies on these exchanges are small and medium enterprises with specific industry specialization. Although these companies may participate in multiple product categories within an industry, they usually focus on a dominant category.

Both B2B exchanges offered free memberships and similar services for buyers (e.g., create profile pages, view suppliers' information). Buyers can also post an unlimited number of buying requests for free on both platforms. In contrast, the platforms provided differentiated services for suppliers. Suppliers can post up to a certain number of product listings depending on

[14] The exchanges ranked second and third in Alexa's site popularity for "Import and Export Portals" in June 2010.

their membership type (free or paid) on the platforms. In addition, the platforms offer multiple categories of paid memberships with different limits on the number of product listings.[15] By focusing on buyers' usage in this study, our results are less likely to be confounded by system constraints due to differentiated supplier services across the two platforms.

We collected data over four time periods. In the first period, we retrieved buying requests posted in all product categories by buyers on both B2B exchanges. We did this on a daily basis for a month. By the end of the month, we identified 690 buyers in Exchange A and 902 buyers in Exchange B who posted at least one buying request. For each of these buyers, we gathered information about the number of buying requests they individually posted and the corresponding platform characteristics at the product category level (the numbers of product listings and buying requests posted by other users, and suppliers on paid membership).

Appendix 3-A shows buying requests posted in the exchanges. To identify multi-homing buyers, we matched information about each buyer (e.g. company name, address, country, and contact information) and their buying requests (e.g. product requested, product category, and product description) across the platforms. We considered buyers to be multi-homing if they posted the same buying request in both platforms in the first period. On average, buyers in our dataset posted the same buying request in both platforms less than 4 days apart, with 50% of buyers posted the same buying request in both platforms on the same day. By the end of first period, we identified 118 multi-homing buyers, who constitute 17.1% and 13.1% of buyers who posted a buying request in the respective platforms in that period. These percentages are

[15] On one platform, suppliers having free membership can post up to 20 product listings, while those having paid memberships can post up to 200 or infinite number of product listings (depending on the membership category). On the other platform, suppliers having free membership can post up to 50 product listings, while those having paid memberships have a limit of 200 or 1,000 product listings (depending on the membership category). However, both platforms offer free membership to buyers, who can post unlimited number of buying requests.

conservative estimates of multi-homing buyers in our sample, since buyers could be using other B2B exchanges beside either of those that we use in this study. The percentages of multi-homing buyers might also be affected by the platforms' popularity (proxied by Alexa site popularity ranking), as being in popular platforms negate some buyers' need to use competing platforms. This is consistent with Walczak et al.'s (2006) observations of online auction sellers in eBay, Amazon, and Yahoo. In that study, 10% of respondents in eBay, which is the most popular among the three auction platforms, indicated that they use multiple auction sites. In contrast, 53% and 74% of respondents in Amazon and Yahoo, respectively, use multiple auction sites.

In the subsequent three time periods, we collected data on buyers' usage and platform characteristics for the 118 buyers at two-month intervals.[16] At the end of the data collection stage, we had 472 observations of 118 buyers in 79 product categories. Our sample includes new and experienced users on the platforms (Table 3.1). Almost 55% of buyers in each platform have more than one year of experience with the platform when we included them in our dataset in the first period. The average membership tenure of buyers in Exchange A and Exchange B was 837 and 827 days, respectively; the median membership tenure of buyers was 498 and 407 days, respectively. Also, the difference in the median buyer's registration dates on the two platforms is approximately 15 days – i.e., the median buyer joined both platforms about two weeks apart. Hence the buyers in our sample have relatively similar experiences with both platforms.

[16] We could observe multi-homing buyers' daily usage on the platforms for one month in the first time period. 86.4% of buyers in Exchange A and 78.8% of buyers in Exchange B posted buying requests on one day during this period. In each exchange, less than 6% of buyers posted requests on three or more days, and none posted requests on more than five days. Given the relatively infrequent postings by buyers, it is reasonable to use a two-month interval during our data collection. In addition, as we could not observe transactions on the platforms, we did not collect data of individual buying requests in subsequent time periods. Nevertheless, by aggregating requests at the buyer level, we still can make appropriate inferences of buyers' behaviors in the setting of our interest.

Table 3.1 Frequency Distribution of Buyers' Membership Tenure

Membership tenure in First Time Period	Exchange A	Exchange B
30 days or less	16 (13.6%)	19 (16.1%)
Between 30 and 89 days	13 (11.0%)	14 (11.9%)
Between 90 and 179 days	13 (11.0%)	13 (11.0%)
Between 180 and 359 days	12 (10.2%)	9 (7.6%)
360 days or more	64 (54.2%)	63 (53.4%)
Average Tenure	837 days	827 days
Median Tenure	498 days	407 days

3.5 Model Specification

To model multi-homing buyers' behavior, we adapt the model specification in Chevalier's and Mayzlin's (2006) study of the effects of consumer reviews on relative sales of books on two online booksellers. We specify the number of buying requests posted in each platform by a multi-homing buyer as a function of (i) the number of product listings posted by suppliers, and (ii) the number of buying requests posted by *other buyers* in the relevant product category. Since buyers are multi-homing, we allow the characteristics of each platform to affect individual buyers' usage in the other platform. Equations 1 and 2 show the function of buying requests posted by a buyer in Exchange A and Exchange B, respectively:

$$\begin{aligned}\log\left(Request_{i,t}^{A}\right) = {} & \alpha_1 \log(Selling_t^A) + \alpha_2 \log\left(Buying_{-i,t}^{A}\right) + \alpha_3 \log\left(Buying_{-i,t}^{A}\right)^2 \\ & + \alpha_4 \log(Selling_t^B) + \alpha_5 \log\left(Buying_{-i,t}^{B}\right) + \alpha_6 \log\left(Buying_{-i,t}^{B}\right)^2 \\ & + \boldsymbol{X_t}\Gamma^A + C_i + \upsilon_i^A + \mu_{i,t}^A \qquad [3.1]\end{aligned}$$

$$\begin{aligned}\log\left(Request_{i,t}^{B}\right) = {} & \beta_1 \log(Selling_t^A) + \beta_2 \log\left(Buying_{-i,t}^{A}\right) + \beta_3 \log\left(Buying_{-i,t}^{A}\right)^2 \\ & + \beta_4 \log(Selling_t^B) + \beta_5 \log\left(Buying_{-i,t}^{B}\right) + \beta_6 \log\left(Buying_{-i,t}^{B}\right)^2 \\ & + \boldsymbol{X_t}\Gamma^B + C_i + \upsilon_i^B + \mu_{i,t}^B \qquad [3.2]\end{aligned}$$

where superscripts *A* and *B* refers to Exchange A and Exchange B, respectively; $Request_{i,t}$ represents the number of buying requests posted by buyer *i* in period *t*; $Selling_t$ is the number of product listings posted in the relevant product category in period *t*; $Buying_{-i,t}$ is the number of buying requests posted by all buyers excluding buyer i in the relevant product category in period t; $\mathrm{X_t}$ is a vector of variables to control for heterogeneity between the platforms in period t; C_i is the unobserved individual effect; and υ_i is the unobserved time-invariant individual-platform effect. We use a log specification for activity levels in our models as the data in their original form have high skewness and kurtosis. For instance, the numbers of buying requests posted in Exchange A and Exchange B by individual buyers have skewness greater than 3.0 and kurtosis greater than 20.0. After log-transformation, the skewness and kurtosis are closer to those in normal distributions: skewness is now between .4 and .9 and kurtosis is between 3.0 and 4.5.

The unobserved individual effect, C_i, in Equations 1 and 2 accounts for factors such as a buyer's computer self-efficacy, which affect his perceptions and use of the platforms (Venkatesh, 2000). To eliminate C_i, we difference Equations 1 and 2. By doing so, we also eliminate industry-level effects that influence platform characteristics and users' preference. For instance, an industry shock may affect the numbers of product listings and buying requests in the platforms, and individual buyers' purchase needs. Equation 3 shows the differenced equation:

$$\begin{aligned} PrefB = {} & \gamma_1 \log(Selling_t^A) + \gamma_2 \log\left(Buying_{-i,t}^A\right) + \gamma_3 \log\,(Buying_{-i,t}^A)^2 \\ & + \gamma_4 \log(Selling_t^B) + \gamma_5 \log\left(Buying_{-i,t}^B\right) + \gamma_6 \log\left(Buying_{-i,t}^B\right)^2 \\ & + \boldsymbol{X_t}\Pi + \left(\upsilon_i^B - \upsilon_i^A\right) + \mu_{i,t} \end{aligned} \qquad [3.3]$$

where $\gamma_j = \beta_j - \alpha_j$, and $PrefB = \log\left(Request_{i,t}^B\right) - \log\left(Request_{i,t}^A\right) = \log\,(\frac{Request_{i,t}^B}{Request_{i,t}^A})$. The term in the parentheses denotes the ratio of buying requests posted in Exchange B to those posted in Exchange A by buyer i. We treat this ratio as buyer i's revealed preference for Exchange B. The log specification in our model ensures the dependent variable is symmetric. Our measure of buyer preference is conceptually similar to the constant-sum scale measure in marketing research (e.g. Amir and Levav, 2008; Carpenter and Nakamoto, 1989; Griffin and Hauser, 1993), where an individual's preference is expressed as ratios of points allocated to different options. However, instead of asking users to allocate points, we observe their relative usages of various options (in this case, the respective platforms). This approach is appropriate as we assume users reveal their preferences pattern by their market behaviors (Samuelson, 1948).

The time-invariant individual-platform effects, υ_i^A and υ_i^B, captures the unobserved heterogeneity in individual buyers' preferences for the platforms.[17] The difference in these effects in Equation 3 will bias the parameter estimates if it is non-zero. To address this, we use a panel structure for our data and estimate our main models using fixed effects (Wooldridge, 2002). As the product category is time-invariant, we cannot introduce a dummy variable for each product category using a fixed effects model. Instead, we cluster our observations by the product category to which buyers belong to appropriately adjust the standard errors (Wooldridge, 2002).

To control for heterogeneity between platforms, we account for the number of suppliers with paid memberships in the respective product categories on each platform ($Paid_{\mathrm{t}}$). B2B exchanges often promote suppliers who have paid memberships; different membership types may affect buyers' relationships with suppliers on the platforms (Koh et al., forthcoming) and their preferences for exchanges. In addition, we add three time period dummies (d_1, d_2, d_3) in Equation 3 to account for systematic differences in buyers' preferences across time.

[17] There were no press releases from the exchanges during our data collection period. This indicates that no major initiatives were launched on the exchanges during this time, and supports our assumption of time-invariant individual-platform effects.

3.6 Results and Analyses

3.6.1 Main Results

Table 3.2 Descriptive Statistics and Correlation Matrix

Variable	1	2	3	4	5	6	7
1 Preference for Exchange B	1.00						
2 Selling activity (Exchange A)	-0.09	1.00					
3 Buying activity (Exchange A)	-0.14	0.60	1.00				
4 Selling activity (Exchange B)	0.10	0.15	0.00	1.00			
5 Buying activity (Exchange B)	-0.04	-0.01	0.24	0.30	1.00		
6 Paid-membership suppliers (Exchange A)	0.01	0.75	0.33	0.25	-0.12	1.00	
7 Paid-membership suppliers (Exchange B)	0.12	0.12	-0.09	0.83	0.24	0.27	1.00
Mean	0.00	2.96	1.75	3.81	2.60	0.55	1.12
Std. Dev.	0.30	0.81	0.78	0.39	0.41	0.64	0.43

N = 472 (4 observations per buyer). All variables are log-transformed.

Table 3.2 shows the descriptive statistics and correlation matrix of the variables across time periods. We added a constant (.5) to all variables before log transformation to avoid logarithm of zero. Selling and buying activities between the exchanges correlate moderately (r = .15 and r = .24, respectively), suggesting that activities in the exchanges do not overlap sufficiently to lead to identification issues. Also, the correlation between buying and selling activities is .60 in Exchange A but .30 in Exchange B. This suggests that there could be structural differences between platforms, and it is necessary to control for unobserved effects as we do in our models.

Table 3.3 Main Results (Fixed Effects Regressions)

	Regression 1	Regression 2
DV: Preference for Exchange B	**Coeff.**	**Coeff.**
γ_0: $Constant$	-1.41 (1.23)	-0.97 (1.54)
γ_1: $\log(Selling_t^A)$	-0.15^{*} (0.06)	-0.13^{*} (0.06)
γ_2: $\log\left(Buying_{-i,t}^A\right)$	-0.04 (0.10)	-0.05 (0.13)
γ_3: $\log\,(Buying_{-i,t}^A)^2$	0.01 (0.10)	-0.02 (0.12)
γ_4: $\log(Selling_t^B)$	0.63^{+} (0.32)	0.80^{*} (0.36)
γ_5: $\log\left(Buying_{-i,t}^B\right)$	1.83^{**} (0.64)	2.19^{*} (0.84)
γ_6: $\log\,(Buying_{-i,t}^B)^2$	-0.73^{**} (0.26)	-1.01^{**} (0.34)
γ_7: $\log(Paid_t^A)$	-0.09^{*} (0.04)	-0.07^{+} (0.04)
γ_8: $\log(Paid_t^B)$	-0.15 (0.11)	-0.14 (0.10)
γ_9: $d1$	0.00 (0.05)	-0.81 (0.83)
γ_{10}: $d2$	-0.01 (0.04)	-1.27^{*} (0.55)
γ_{11}: $d3$	-0.01 (0.02)	-1.95^{*} (0.93)
γ_{12}: $d1$ x $\log\left(Buying_{-i,t}^A\right)$		0.06 (0.08)
γ_{13}: $d1$ x $\log\,(Buying_{-i,t}^A)^2$		-0.03 (0.03)
γ_{14}: $d1$ x $\log\left(Buying_{-i,t}^B\right)$		0.82 (0.70)
γ_{15}: $d1$ x $\log\,(Buying_{-i,t}^B)^2$		-0.20 (0.15)
γ_{16}: $d2$ x $\log\left(Buying_{-i,t}^A\right)$		0.12 (0.09)
γ_{17}: $d2$ x $\log\,(Buying_{-i,t}^A)^2$		-0.05 (0.04)
γ_{18}: $d2$ x $\log\left(Buying_{-i,t}^B\right)$		1.15^{*} (0.44)
γ_{19}: $d2$ x $\log\,(Buying_{-i,t}^B)^2$		-0.26^{**} (0.09)
γ_{20}: $d3$ x $\log\left(Buying_{-i,t}^A\right)$		0.06 (0.07)
γ_{21}: $d3$ x $\log\,(Buying_{-i,t}^A)^2$		-0.03 (0.03)
γ_{22}: $d3$ x $\log\left(Buying_{-i,t}^B\right)$		1.61^{*} (0.73)
γ_{23}: $d3$ x $\log\,(Buying_{-i,t}^B)^2$		-0.33^{*} (0.14)
F-statistics	$p < .001$	$p < .001$
R^2 (within transformation)	0.07	0.13

N = 472. Robust standard errors in parenthesis.
$^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$

We estimated Equation 3.3 with fixed effects transformation at the buyer level. Table 3.3 Regression 1 shows the results from this regression. Using the Sargan-Hansen statistic, we reject the null hypothesis that the orthogonality assumption is valid (chi-sq = 25.60, df = 8, $p < .01$), supporting our use of fixed effects.[18] In addition, to compare fixed effects model with pooled

[18] If the orthogonality assumption is not satisfied (i.e. unobserved effects do correlate with independent variables), random effects estimators are not consistent but fixed effects estimators are. However, if the orthogonality assumption is satisfied, random effects estimators are consistent and also more efficient than fixed effects estimators. The Sargan-Hansen test statistic is a heteroskedastic- and cluster-robust form of a Hausman test that compares random effects and fixed effects estimators (Schaffer and Stillman, 2010). Rejection of the null hypothesis in the Sargan-Hansen test (as in our case) implies using random effects is not appropriate.

OLS model, we estimated the fixed effects using dummy variable regression without clustering the observations by product category. The F-test rejects the null hypothesis that fixed effects are zero (F-statistics = 8.59, $p < .001$), indicating that using fixed effects is appropriate.

As shown in Regression 1, selling activity level in Exchange A has a significant negative effect on buyers' preferences for Exchange B ($\gamma_1 = -.15$, $p < .05$), while selling activity level in Exchange B has a marginally significant positive effect on buyers' preferences ($\gamma_4 = .63$, $p < .10$). These results demonstrate positive inter-network externalities, and support H1. In addition, the level of buying activity in Exchange B has a significant non-monotonic relationship with buyers' preferences for Exchange B ($\gamma_5 = 1.83$, $p < .01$ and $\gamma_6 = -.73$, $p < .01$), supporting H2. Buyers' preferences for Exchange B initially increase with increasing buying levels in the exchange but decrease at higher buying levels, consistent with our hypothesis integrating social proof and competition effects. Although the signs of the estimators for buying activity level in Exchange A are in the expected direction, they are not statistically significant ($\gamma_2 = -.04$, $p > .10$ and $\gamma_3 = .01$, $p > .10$). These estimates could be non-significant because of low variations in buying activities on Exchange A across time periods. The average median change in buying activities on Exchange A across the four time periods is 3.67 buying requests, while the corresponding change on Exchange B is 9 buying requests. The low variations in buying activities on Exchange A would lead to larger standard errors for the estimates, resulting in non-significant results (Wooldridge, 2006).

Next, to examine if the influences of buying activities on buyers' preferences decline over time, we included interactions between time period dummies and buying activity levels in our model (Table 3 Regression 2). The estimates for selling activities in both platforms are now significance at .05 levels ($\gamma_1 = -.13$, $p < .05$ and $\gamma_4 = .80$, $p < .05$). We find interactions that involve buying activities on Exchange A are not significant (p > .10 for $\gamma_{12}, \gamma_{13}, \gamma_{16}, \gamma_{17}, \gamma_{20}$, and γ_{21}). Although the estimates for interactions between buying activities on Exchange B and period 1 are consistent with our expectation, they are also not significant (γ_{14} = .82, p > .10 and γ_{15} = -.20, p > .10). However, we find significant interactions between buying activities on Exchange B and period 2 (γ_{18} = 1.15, p < .05 and γ_{19} = -.26, p < .01) and period 3 (γ_{22} = 1.61, p < .05 and γ_{23} = -.33, p < .05). These results indicate that the impacts of buying activities in Exchange B on buyers' preferences is weaker in period 4 than in periods 2 and 3, supporting H3. The non-significant interactions between period 1 and buying activity levels in Exchange B suggest a possibility of non-linear time trend, which we examine in the robustness analysis section.

What do these results mean practically? Recall that a buyer's preference for Exchange B refers to her use of Exchange B over that of Exchange A. Higher relative usage of Exchange B leads to a higher total dollar value of buying requests and marketplace liquidity in Exchange B than in Exchange A. Suppose each buyer in our sample posts 5 buying requests in Exchange A and the average value of each request is $30,000.[19] A 1% higher relative usage of Exchange B implies that the value of buying requests in Exchange B is greater than that in Exchange A by $177,000 (5 x $30,000 x .01 x 118). Based on the results in Table 3 Regression 2, *ceteris paribus*, a 1% increase in selling activities in Exchange A in period 4 lowers relative usage of Exchange B by the multi-homing buyers by .13% ($23,010 less in buying request value in Exchange B), whereas a 1% increase in selling activities in Exchange B in the same period raises relative usage of Exchange B by these buyers by .80% ($141,600 more in buying request value in Exchange B). Table 3.4 shows changes in relative usage of Exchange B and corresponding

[19] Buyers in the exchanges do not state the purchase values in their buying requests. We based our estimates on Koh et al.'s (forthcoming) survey of buyers in a similar B2B exchange, in which the median purchase value is $30,000.

additional values of buying requests posted by the multi-homing buyers at various buying activity levels in the exchange in period 4.

Table 3.4 Estimated Impacts of Changes to Buying Activity Levels in Exchange B

A 1% increase in buying activity level when number of requests posted by *other buyers* is…	**Change in relative usage of Exchange B**	**Additional value of buying requests in Exchange B**
5	.78%	\$137,720
10	.17%	\$30,090
15	-.19%	-\$32,870
20	-.44%	-\$77,540
25	-.63%	-\$112,189
30	-.79%	-\$140,500

Note: We assume each multi-homing buyer in our sample initially posts 5 buying requests in both exchanges, and the purchase value of each request is \$30,000. A 1% increase in relative usage of (or preference) Exchange B implies that the value of buying requests in Exchange B is greater than that in Exchange A by \$177,000 (5 x \$30,000 x .01 x 118).

3.6.2 Robustness Analyses

A concern in our setting is that buyers' preferences may affect the number of buying requests and product listings posted in the same time period. Using differencing regression, we regressed the change in buyers' preferences between July 2009 and February 2010 on changes in exchange characteristics between July and December 2009. The longer period for our dependent variable (7 months) relative to that for the independent variables (5 months) minimizes the endogeneity concern. This specification is similar to that in Chevalier's and Mayzlin's (2006) study of the effects of books reviews on online sales, where the dependent variable is the change in sales rank of books between May and August 2003, and the independent variables are the changes in book reviews between May and July 2003.

Table 3.5 presents the results from this analysis. The change in selling activities in Exchange B is positively related to the change in buyers' preferences for Exchange B ($\gamma_4 = 1.00$, $p < .05$), and the change in buying activities in Exchange B is non-monotonically related to the change in buyers' preferences ($\gamma_5 = 1.94$, $p < .05$ and $\gamma_6 = -.77$, $p < .05$). These results are qualitatively similar to those in our hypothesized models, supporting H1 and H2. However, the change in buyers' preferences is not significantly related to changes in selling (γ_1 = -.02, p > .10) and buying activities on Exchange A (γ_2 = -.04, p > .10 and γ_3 = -.07, p > .10). One possible explanation for these non-significant estimates is that using differencing regression may reduce variations in the explanatory variables (Wooldridge, 2006). Between July and December 2009, the median changes in selling and buying activities on Exchange A are 120 product listings and 9 buying requests, respectively, whereas the corresponding changes in Exchange B are 929 product listings and 19 buying requests. As in the case of our earlier results, the small variations in activity levels in Exchange A's could result in non-significant estimates. However, the results for Exchange B are still robust and consistent with our theory.

Table 3.5 Robustness Analysis

DV: Change in Preference for Exchange B	**Coeff.**
γ_0: $Constant$	0.00 (0.04)
γ_1: $\mathbf{\Delta}\log(Selling_t^A)$	-0.02 (0.24)
γ_2: $\mathbf{\Delta}\log\left(Buying_{-i,t}^A\right)$	-0.04 (0.11)
γ_3: $\mathbf{\Delta}\log\ (Buying_{-i,t}^A)^2$	-0.07 (0.13)
γ_4: $\mathbf{\Delta}\log(Selling_t^B)$	1.00^* (0.49)
γ_5: $\mathbf{\Delta}\log\left(Buying_{-i,t}^B\right)$	1.94^* (0.79)
γ_6: $\mathbf{\Delta}\log\ (Buying_{-i,t}^B)^2$	-0.77^* (0.31)
γ_7: $\mathbf{\Delta}\log(Paid_t^A)$	-0.32^+ (0.17)
γ_8: $\mathbf{\Delta}\log(Paid_t^B)$	-0.02 (0.16)
F-statistics	p < .01
R^2	0.082

N = 118. Robust standard errors in parenthesis.
$^+$ p < .10 * p < .05

The dependent variable is the change in buyers' preferences for Exchange B between July 2009 and Feb 2010, while the independent variables are changes in exchange characteristics between July and December 2009.

We also checked the robustness of our results by using time trend variables instead of time period dummies in our models (Appendix 3.B). As pointed out earlier, time trend may be non-linear in our model. Therefore, we specified a quadratic function for time trend, and interacted it with buying activity levels on the platforms. We find that the interactions between time trend and buying activities on Exchange A are not significant. However, the relationship between buyers' preferences for exchanges and buying activities on Exchange B is moderated by time. Our results show that the sensitivity of buyers' preferences to changes in buying levels in Exchange B decreases over time, and support H3. These results are similar to those where we used time period dummies as proxies for buyers' platform experience.

3.6.3 Impacts of Model Misspecification

Our models assume that multi-homing behavior influences buyer participation on B2B exchanges. Our results suggest it is a valid assumption. A related question is whether this is a useful assumption for researchers and practitioners. In practice, B2B exchanges may analyze activities on their platforms without considering their users' participation on other exchanges. In doing so, these exchanges essentially assume users are single-homing on their respective platforms. In prior published work, researchers sometimes assume single-homing behaviors even though multi-homing behaviors are possible (e.g. Belleflamme and Toulemonde, 2009; Chandra and Collard-Wexler, 2009). We conducted additional analyses by assuming (contrary to fact) buyers in our sample are single-homing in each platform. Regressions 1a and 2a in Table 3.6 show the estimates of Equation 3.1 (buyers' participation in Exchange A) and Equation 3.2 (buyers' participation in Exchange B), respectively, with a fixed effects model. We also included interactions between time period dummies and buying activities on the platforms in these regressions. Each of these regressions assumes that participation in a focal platform is a function of activity levels of the competing platform but not of its users' specific level of participation there.

Table 3.6 Results of Misspecified Models (Fixed Effects Regressions)

Equation 1			Equation 2		
DV: Buyers' participation in Exchange A	**Regression 1a Coeff.**	**Regression 1b Coeff.**	**DV: Buyers' participation in Exchange B**	**Regression 2a Coeff.**	**Regression 2b Coeff.**
$\alpha_0: Constant$	1.99 (1.96)	0.49^{+} (0.27)	$\beta_0: Constant$	1.02 (1.97)	0.90 (1.83)
α_1:log($Selling_t^A$)	0.26** (0.10)	0.24* (0.10)	β_1:log($Selling_t^A$)	0.13 (0.08)	—
α_2: log($Buying_{-i,t}^A$)	-0.08 (0.23)	-0.07 (0.23)	β_2: log($Buying_{-i,t}^A$)	-0.12 (0.23)	—
α_3: log $(Buying_{-i,t}^A)^2$	-0.14 (0.11)	-0.14 (0.10)	β_3: log $(Buying_{-i,t}^A)^2$	-0.16 (0.12)	—
α_4: log($Selling_t^B$)	-0.47 (0.45)	—	β_4: log($Selling_t^B$)	0.33 (0.53)	0.18 (0.47)
α_5: log($Buying_{-i,t}^B$)	1.09 (0.77)	—	β_5: log($Buying_{-i,t}^B$)	3.28*** (0.89)	3.69** (1.09)
α_6: log $(Buying_{-i,t}^B)^2$	-0.35 (0.22)	—	β_6: log $(Buying_{-i,t}^B)^2$	-1.36*** (0.35)	-1.49** (0.41)
α_7: log($Paid_t^A$)	0.04 (0.05)	0.04 (0.04)	β_7: log($Paid_t^A$)	-0.03 (0.05)	—
α_8: log($Paid_t^B$)	-0.07 (0.12)	—	β_8: log($Paid_t^B$)	-0.21 (0.13)	-0.17 (0.13)
α_9: $d1$	0.30 (1.06)	-0.02 (0.07)	β_9: $d1$	-0.52 (1.01)	-0.32 (1.02)
α_{10}: $d2$	0.49 (1.01)	0.03 (0.06)	β_{10}: $d2$	-0.79 (0.87)	-0.78 (0.86)
α_{11}: $d3$	0.67 (0.96)	-0.01 (0.04)	β_{11}: $d3$	-1.28 (1.26)	-1.22 (1.23)
F-statistics	$p < .001$	$p < .01$	F-statistics	$p < .001$	$p < .001$
R^2 (within transformation)	0.10	0.08	R^2 (within transformation)	0.13	0.10

N = 472. Robust standard errors in parenthesis.

$^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Notes:

Regressions 1a and 2a assume the respective exchanges observe activity levels in competing exchanges, but they do not monitor their users' participation in those exchanges.

Regressions 1b and 2b assume the respective exchanges only analyze internal data that they possess.

The interactions between time period dummies and buying activities on the exchanges are not statistically significant, and are not shown here in the interest of space.

We find a positive relationship between selling activities and buyers' participation in Exchange A (α_1 = .26, $p < .01$ in Regression 1a), and a non-monotonic relationship between buying activities and buyers' participation in Exchange B (β_5 = 3.28, $p < .001$ and β_6 = -1.36, $p < .001$ in Regression 2a). These findings are consistent with what we observe in the main model (Equation 3.3; Table 3.3). However, there are three important differences between the results in the main models and those in the misspecified models. First, while the results in our hypothesized models indicate that activity levels in both exchanges affect the buyers' relative usage of the exchanges, the results in the misspecified models suggest that the activity levels in the competitive exchange do not affect buyers' participation in the focal exchange ($p > .10$ for α_4, α_5, and α_6 in Regression 1a, and for β_1, β_2, and β_3 in Regression 2a). Second, unlike in the main models, the interactions between buying activity levels and all the time period dummies are not statistically significant on both exchanges in the misspecified models.[20] Lastly, the estimated positive effect of selling activities in Exchange B on buyers' participation is significant in our hypothesized model reported (γ_4 = .80, $p < .05$ in Table 3.3 Regression 2), but it is not significant in the misspecified model (β_4 = .33, $p > .10$ in Table 3.6 Regression 2a).

We also estimated Equations 3.1 and 3.2 with the characteristics of the non-focal exchange excluded from the models (Regressions 1b and 2b in Table 3.6). This specification represents the scenario where platforms only analyze the internal data that they individually possess. The results in Regression 1b and 2b are similar to those in Regressions 1a and 2a, respectively. Selling activity is positively associated with buyers' participation in Exchange A (α_1 = .24, $p < .05$ in Regression 1b), and buying activities and buyers' participation in Exchange B are non-monotonically related (β_5 = 3.69, $p < .01$ and β_6 = -1.49, $p < .01$ in Regression 2b).

[20] We do not show the results of these interactions terms in Table 6 in the interest of space.

Comparing the results between the hypothesized and misspecified models, we observe two problems with the latter models. First, the misspecified models paint a biased picture of competition between the platforms. The estimates of the misspecified models suggest that the buyers' usage on one platform is not affected by the activities on the competing platform. Such conclusions may cause a B2B exchange to focus insufficiently on activity levels in its competitors' platforms. Second, the discrepancies in results due to model misspecification differ across platforms. For Exchange A, results in the hypothesized models show that selling activities on the platform is positively related to buyers' preferences, but buying activities do not affect buyers' preferences. These results are qualitatively similar to those in the misspecified models. In contrast, for Exchange B, the estimate for selling activity level is significant in the hypothesized model but not in the misspecified models. The differences in discrepancy of results between the hypothesized and misspecified models across platforms (i) make it difficult for practitioners and researchers to know *a priori* how wrong assumptions of users' homing behaviors bias results, and (ii) could lead to sub-optimal strategies and behavior. For example, based on the misspecified models, Exchange B may underinvest in increasing selling activities on the platform; however, according to the hypothesized model, increasing selling activities has a positive effect on buyers' relatively usage on Exchange B.

3.7 Discussion

In this study, we examine a prevailing trend of users' concurrent participation in multiple competing platforms. We find that multi-homing buyers' preferences for a B2B exchange are positively associated with the inter-network activity (selling) levels. We further find that there are two opposing effects of increasing intra-network activity (buying) levels: (i) a positive social proof effect, and (ii) a negative competition effect. As a result, the relationship between multi-homing buyers' preferences and buying activity levels is non-monotonic. As buying activities on an exchange increase, buyers' preferences for this exchange initially increase and subsequently decrease. This result is surprising as economic theory suggests that higher buying activity levels increase competition among buyers and reduce the benefits to their participation on the platforms (Anderson et al., 2008; Belleflamme and Toulemonde, 2009; Tucker and Zhang, 2010). Lastly, we find suggestive evidence that buyers' preferences become less sensitive to changes in competitive buying levels on the platforms over time. Our explanation is that buyers acquire direct experiences on the platforms and build their network of suppliers with time. Consequently, they rely less on other buyers' behaviors on the platforms for social proof, and leverage on existing relationships with suppliers when competition increases on the platforms.

A key implication in this study is that *we cannot ignore multi-homing behaviors when studying or managing platforms*; such behaviors affect various areas of platforms' growth and success (e.g., user participation, marketplace liquidity, and competitiveness). When thought about more broadly, multi-homing is ubiquitous. Firms and consumers are likely to use multiple platforms with similar functionality in various domains. Some examples outside the B2B domain include online investment communities (e.g., Seeking Alpha and The Motley Fool), recommendation sources (e.g., Facebook, Foursquare, and Yelp), and social group buying websites (e.g., Groupon and LivingSocial). We should therefore account for users' homing behaviors when examining the usage of and competition among respective platforms.

3.7.1 Theoretical and Managerial Implications

Our results show the importance of correctly modeling users' behaviors across online exchanges. The assumption of single-homing behaviors facilitates theoretical and analytical discussions, and makes it easier to conduct empirical studies since researchers do not need to involve competing platforms in the research design. However, as we show in our analyses of misspecified models, abstracting away multi-homing behaviors (when they exist) impacts results and gives an incomplete and potentially biased picture of competitive dynamics. For platform operators, this means that they should recognize the extent to which their users are participating on other platforms. Such behaviors affect the exchanges' positions in users' choice sets and competition. Ignoring multi-homing users' participation on other platforms could lead to sub-optimal strategies. Although monitoring their users' participation on competitors' platforms is an ideal way to get information on these users' behaviors and preferences, it may not be possible for individual online exchanges to do so. A feasible alternative is to survey users to find out their usage on and preferences for competing platforms. Online exchanges can also learn from what other industries do to observe users' multi-homing behaviors. For instance, balance transfer services allow credit card companies to observe cardholders' usage of other credit cards. Similarly, price-matching guarantees help retailers learn about other places where their customers shop. Platforms can adopt similar initiatives to encourage their users to provide information about their homing behaviors.

This study also extends research that examines conditions in which information about platforms' intra- and inter-network sizes affects user participation (Tucker and Zhang, 2010). Under the effect of "imputation under uncertainty", when users only know a platform's intra-network size, they impute a commensurate size for the inter-network and assume the platform is in equilibrium. Consequently, a large intra-network size alone does not deter user participation. However, when users know the sizes of both sides of the platform, the intra-network size signals the extent of competition that they face. In other words, negative intra-network externalities kick in when information about both sides of a platform is shown to users.

While the *conditions* in which intra-network information is shown on the platform are important (i.e., whether it is shown in conjunction with inter-network information), we find that the *levels* of intra-network activity also matter. Intra-network activities have different signaling effects at different activity levels. When the level is low, an increase in intra-network activities sends a positive signal about similar users' evaluations about the platform. However, an increase in intra-network activities when the level is high signals stiff competition among users on the same side on the platform, consistent with the observations in Tucker and Zhang (2010). Thus, online exchanges need to strategically manage users' perceptions of activity levels on both sides of their platforms. This can be challenging as users' preferences are asymmetrically related to inter- and intra-network activity levels. On a platform, high activity levels on side A benefit users on the opposite side B, but not necessarily users on side A. There is also a natural interplay between inter- and intra-network externalities on two-sided exchange networks, and strategies aimed at one side of the platform would affect the opposite side indirectly. Therefore, contrary to what many B2B exchanges are doing, being perceived as the largest exchange may not be the best strategy for influencing users' preference![21] The results in this and related studies (e.g., Tucker and Zhang, 2010) show that it is necessary to consider *what* information to provide (e.g.,

[21] The following are claims made by some B2B exchanges (emphases added):

1. "The world's largest base of suppliers." – www.alibaba.com
2. "ECeurope is the largest source of international trade leads, rfq and tender opportunities from companies and government organizations around the world." – www.eceurope.com
3. "Over 4 million offers are posted in our website, which is the largest scale in the world." – www.ecplaza.com

whether to present information about buying and/or selling activities), and *when* to provide the information (based on the activity levels).

3.7.2 Limitations

There are some limitations in this study. First, since the B2B exchanges from which we gathered data in this study are relatively popular (based on Alexa site popularity ranking), one can argue that buyers on these exchanges are not concerned with what other buyers are doing. Instead, these exchanges' popularity could provide sufficient assurance to buyers to overcome any uncertainty about participating in the exchanges. That is, social proof may not be necessary or salient when users are in popular platforms. However, one must recognize that *just* because a B2B exchange is, by and large, popular does not mean that it is useful for buyers and suppliers in every market segment. While the general popularity of a platform is important to users, what also matters to them is how well suited the platform is for specific product categories. For example, books sellers have a stronger inclination to use Amazon Auction than eBay; even though eBay has better name recognition and higher web traffic among online auction markets, Amazon is perceived to be more efficient in the book segment (Walczak et al., 2006). In a similar manner, buyers on popular platforms still consider what other buyers *in the same product categories* are doing. Thus, by analyzing buyers' behavior at the product category level in this study, we managed to observe phenomenon in specific market segments where other buyers' behaviors matter to the focal buyers.

Second, by using secondary data, we can only infer users' preferences indirectly from the data. Our research design also restricted us from measuring social proof and competition that buyers are exposed to. It would strengthen this study if we could directly establish buyers' preferences, and the social and market dynamics that they experience on the platforms. Finally, we only examine buyers' usage of the B2B exchanges. Although suppliers' incentives to participate on the platforms may be affected by the exchanges' pricing strategy, our hypotheses in general should also hold among multi-homing suppliers. Future research can investigate how pricing and non-price factors affect users' preferences in multi-homing contexts.

3.8 Conclusion

Today, not only can users select from multiple competing platforms, they can also choose to participate on multiple platforms at the same time. In this paper, we tackle a challenging but important research question: what factors affect users' preferences when they are multi-homing on two-sided exchange networks. We explicitly study a situation where individuals use multiple competing platforms concurrently, and address challenges presented by the dynamics of inter- and intra-network externalities on the platforms.

By observing multi-homing users' behaviors in two B2B exchanges over time, we find that users' preferences for exchanges are associated with the levels of activity that take place within and across different sides of the platforms. Our results suggest that there should be a greater emphasis on investigating multi-homing contexts and users' preferences in future platforms research. Doing so helps us to better understand social and market dynamics that drive platform users' behaviors on exchanges, and competition between platforms.

References

Amir, O., and Levav, J. 2008. Choice Construction Versus Preference Construction: The Instability of Preferences Learned in Context. ***Journal of Marketing Research,*** 45(2): 145-158.

Anderson, R. M., Ellison, G., and Fudenberg, D. 2008. ***Location Choice in Two-Sided Markets with Indivisible Agents.*** Working Paper.

Armstrong, M. 2006. Competition in Two-Sided Markets. ***The RAND Journal of Economics,*** 37(3): 668-691.

Asvanund, A., Clay, K., Krishnan, R. and Smith, M. D. 2004. An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks, ***Information Systems Research,*** 15(2): 155-174.

Bakos, Y. 1998. The emerging role of electronic marketplaces on the Internet. ***Communications of the ACM,*** 41(8): 35-42.

Belleflamme, P., and Toulemonde, E. 2009. Negative Intra-Group Externalities in Two-Sided Markets. ***International Economic Review,*** 50(1): 245-272.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. ***Journal of Political Economy,*** 100(5): 992-1026.

Bloomberg. 2011. Alibaba Shares Tumble After Fraud Leads to CEO Departure. February 22.

Brynjolfsson, E. and Kemerer, C. F. 1996. "Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market," ***Management Science,*** 42(12): 1627-1647.

Bughin, J., Chui, M., and Manyika, J. 2010. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. ***McKinsey Quarterly***, 4: 26-43.

BusinessWeek. 2010. Alibaba Fraud Scandal May Help Google, Global Sources. February 28.

Caillaud, B., and Jullien, B. 2003. Chicken and Egg: Competition among Intermediation Service Providers. ***The RAND Journal of Economics,*** 34(2): 309-328.

Carpenter, G. S., and Nakamoto, K. 1989. Consumer Preference Formation and Pioneering Advantage. ***Journal of Marketing Research,*** 26(3): 285-298.

Chandra, A., and Collard-Wexler, A. 2009. Mergers in Two-Sided Markets: An Application to the Canadian Newspaper Industry. ***Journal of Economics and Management Strategy,*** 18(4): 1045-1070.

Chevalier, J. A., and Mayzlin, D. 2006. The Effect of Word of Mouth on Sales: Online Book Reviews. ***Journal of Marketing Research,*** 43(3): 345-354.

Choi, J. P. 2010. Tying in Two-Sided Markets with Multi-Homing. ***The Journal of Industrial Economics,*** 58: 607-626.

Cialdini, R. B. 2008. ***Influence: Science and Practice*.** Pearson, Boston.

Coombs, C. H., Dawes, R. M. and Tversky, A. 1970. ***Mathematical Psychology: An Elementary Introduction*.** Prentice-Hall, Englewood Cliffs, NJ.

Dai, Q., and Kauffman, R. J. 2002. Business models for internet-based B2B electronic markets. ***International Journal of Electronic Commerce,*** 6(4): 41-72.

Doney, P. M., Barry, J. and Abratt, R. 2007. Trust determinants and outcomes in global B2B services. ***European Journal of Marketing***, 9/10: 1096-1116.

Eisenmann, T., Parker, G., and Van Alstyne, M. W. 2006. Strategies for Two-Sided Markets. *Harvard Business Review,* 84: 92-101.

Evans, D. S. 2003. The Antitrust Economics of Multi-Sided Platform Markets. ***Yale Journal on Regulation,*** 20(2): 325-82.

Evans, D. S., and Schmalensee, R. 2007. Industrial Organization of Markets with Two-Sided Platforms. ***Competition Policy International,*** 3(1): 151-179.

Griffin, A., J. and Hauser, R. 1993. The Voice of the Customer. ***Marketing Science,*** 12(1): 1-27.

Jin, G. Z., and Rysman, M. 2010. ***Platform Pricing at Sportscard Conventions.*** Working Paper.

Katz, M. L., and Shapiro, C. 1985. Network Externalities, Competition, and Compatibility. ***American Economic Review,*** 75(3): 424-440.

Keizer, K., Lindenberg, S., and Steg, L. 2008. The Spreading of Disorder. ***Science,*** 322(5908): 1681-1685.

Koh, T. K., Fichman, M., and Kraut, R. Forthcoming. Trust Across Borders: Buyer-Supplier Trust in Global B2B E-Commerce. ***Journal of the Association for Information Systems***.

Kraut, R. E., Rice, R. E., Cool, C. and Fish, R. S. 1998. Varieties of Social Influence: The Role of Utility and Norms in the Success of a New Communication Medium, ***Organization Science,*** 9(4): 437-453.

Lucking-Reiley, D., and Spulber, D. F. 2001. Business to Business Electronic Commerce. ***Journal of Economic Perspectives,*** 15(1): 55-68.

Poppo, L., Zhou, K. Z. and Ryu, S. 2008. Alternative Origins to Interorganizational Trust: An Interdependence Perspective on the Shadow of the Past and the Shadow of the Future. ***Organization Science***, *19*(1): 39-55.

Rochet, J.-C., and Tirole, J. 2003. Platform Competition in Two-Sided Markets. ***Journal of the European Economic Association,*** 1(4): 990-1029.

Rochet, J.-C., and Tirole, J. 2006. Two-Sided Markets: A Progress Report. ***The RAND Journal of Economics,*** 37(3): 645-667.

Roson, R. 2005. Two-Sided Markets: A Tentative Survey. ***Review of Network Economics,*** 4(2): 142-160.

Samuelson, P. A. 1948. Consumption Theory in Terms of Revealed Preference. ***Economica,*** 15(60): 243-253.

Schaffer, M. E., and Stillman, S. 2010. xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor. http://ideas.repec.org/c/boc/bocode/s456779.html. (accessed: August 9, 2011).

Spulber, D. F. 1999. ***Market microstructure: Intermediaries and the theory of the firm.*** Cambridge University Press, Cambridge.

Time. 2011. Fraud Scandal Hits China's Online Giant Alibaba. February 22.

Thurstone, L. 1927. A law of comparative judgment. ***Psychological Review,*** 34: 273-286.

Tucker, C. 2008. Identifying Formal and Informal Influence in Technology Adoption with Network Externalities. *Management Science,* 54(12): 2024-2038.

Tucker, C., and Zhang, J. 2010. Growing Two-Sided Networks by Advertising the User Base: A Field Experiment. ***Marketing Science,*** 29(5): 805-814.

Venkatesh, V. 2000. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model**. *Information Systems Research,*** 11(4): 342-365.

Walczak, S., Gregg, D. G., and Berrenberg, J. L. 2006. Market Decision Making for Online Auction Sellers. ***Journal of Electronic Commerce Research,*** 7(4): 199-220.

Wooldridge, J. M. 2002. ***Econometric Analysis of Cross Section and Panel Data*****.** The MIT Press, Cambridge, MA.

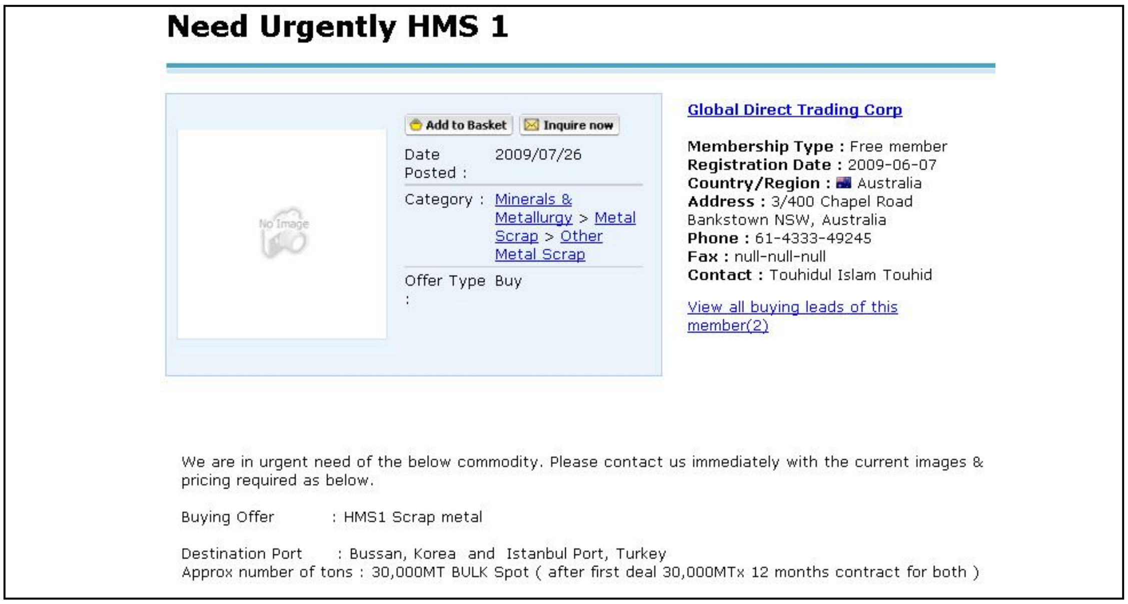
Wooldridge, J. M. 2006. ***Introductory Econometrics: A Modern Approach*** (3rd ed.). Thompson South-Western.

Appendices

Appendix 3.A Sample Buying Requests

Below are sample buying leads posted on the two exchanges. We compared request details and buyer's information listed in the buying requests to identify multi-homing buyers.

Buying Request in Exchange A:



Need Urgently HMS 1

Add to Basket | Inquire now

Date Posted :	2009/07/26
Category :	Minerals & Metallurgy > Metal Scrap > Other Metal Scrap
Offer Type :	Buy

Global Direct Trading Corp

Membership Type : Free member
Registration Date : 2009-06-07
Country/Region : Australia
Address : 3/400 Chapel Road Bankstown NSW, Australia
Phone : 61-4333-49245
Fax : null-null-null
Contact : Touhidul Islam Touhid

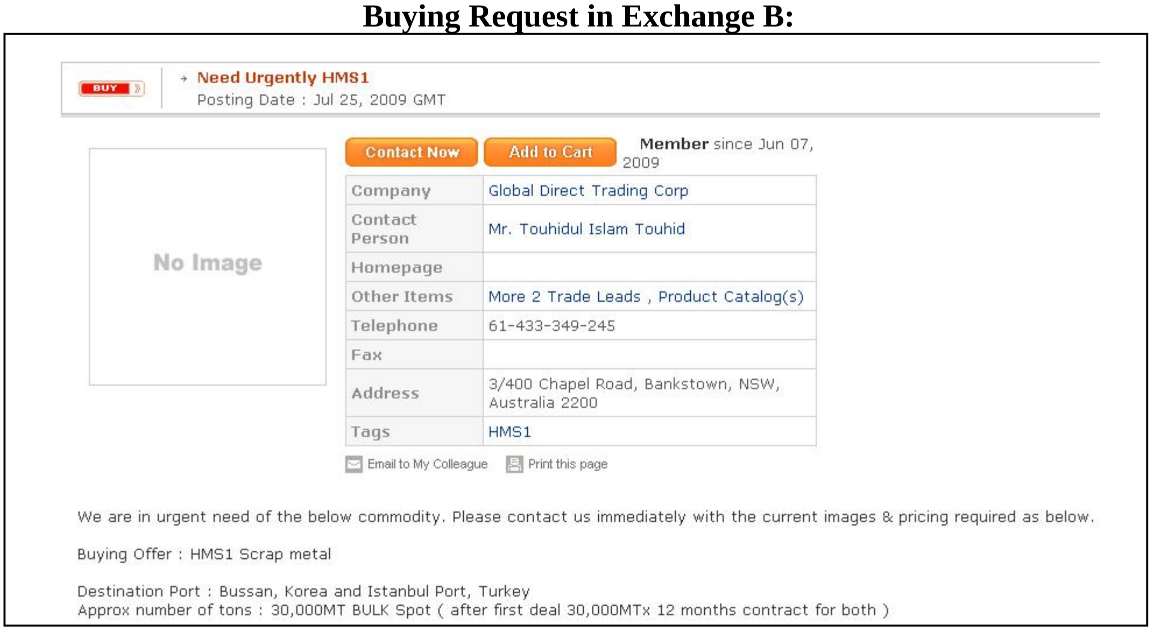
View all buying leads of this member(2)

We are in urgent need of the below commodity. Please contact us immediately with the current images & pricing required as below.

Buying Offer : HMS1 Scrap metal

Destination Port : Bussan, Korea and Istanbul Port, Turkey
Approx number of tons : 30,000MT BULK Spot (after first deal 30,000MTx 12 months contract for both)

Buying Request in Exchange B:



BUY

Need Urgently HMS1
Posting Date : Jul 25, 2009 GMT

No Image

Contact Now | Add to Cart | **Member** since Jun 07, 2009

Company	Global Direct Trading Corp
Contact Person	Mr. Touhidul Islam Touhid
Homepage	
Other Items	More 2 Trade Leads , Product Catalog(s)
Telephone	61-433-349-245
Fax	
Address	3/400 Chapel Road, Bankstown, NSW, Australia 2200
Tags	HMS1

Email to My Colleague | Print this page

We are in urgent need of the below commodity. Please contact us immediately with the current images & pricing required as below.

Buying Offer : HMS1 Scrap metal

Destination Port : Bussan, Korea and Istanbul Port, Turkey
Approx number of tons : 30,000MT BULK Spot (after first deal 30,000MTx 12 months contract for both)

Appendix 3.B Robustness Analysis using Non-Linear Time Trend

We check the sensitivity of our results by using time trend variables instead of time period dummies in our models. As the results in our main analysis indicate the possibility of a non-linear time trend, we specified a quadratic function for time trend (*Time* and *Time*^{2}). The time trend variable takes the value from 1 to 4, corresponding to our data collection periods. Table B.1-1 presents the results of fixed effects regressions using this alternative specification.

Table B.1-1 Robustness Analysis – Including Non-Linear Time Trend (Fixed Effects Regressions)

	Regression 1	**Regression 2**
DV: Preference for Exchange B	**Coeff.**	**Coeff.**
γ_0: $Constant$	-1.38 (1.18)	0.42 (1.75)
γ_1: $\log(Selling_t^A)$	-0.15* (0.06)	-0.13* (0.06)
γ_2: $\log\left(Buying_{-i,t}^A\right)$	-0.04 (0.10)	-0.06 (0.16)
γ_3: $\log\,(Buying_{-i,t}^A)^2$	0.01 (0.10)	-0.01 (0.11)
γ_4: $\log(Selling_t^B)$	0.63^{+} (0.32)	0.78* (0.35)
γ_5: $\log\left(Buying_{-i,t}^B\right)$	1.83** (0.64)	1.32 (1.63)
γ_6: $\log\,(Buying_{-i,t}^B)^2$	-0.73** (0.26)	-0.88^{+} (0.52)
γ_7: $\log(Paid_t^A)$	-0.09^{+} (0.04)	-0.07 (0.04)
γ_8: $\log(Paid_t^B)$	-0.15 (0.11)	-0.12 (0.10)
γ_9: $Time$	-0.03 (0.06)	-2.76^{+} (1.47)
γ_{10}: $Time^2$	0.01 (0.01)	0.59* (0.27)
γ_{11}: $Time$ x $\log\left(Buying_{-i,t}^A\right)$		0.12 (0.14)
γ_{13}: $Time$ x $\log\,(Buying_{-i,t}^A)^2$		-0.05 (0.05)
γ_{14}: $Time^2$ x $\log\left(Buying_{-i,t}^A\right)$		-0.03 (0.03)
γ_{15}: $Time^2$ x $\log\,(Buying_{-i,t}^A)^2$		0.01 (0.01)
γ_{16}: $Time$ x $\log\left(Buying_{-i,t}^B\right)$		2.16^{+} (1.22)
γ_{17}: $Time$ x $\log\,(Buying_{-i,t}^B)^2$		-0.42^{+} (0.25)
γ_{18}: $Time^2$ x $\log\left(Buying_{-i,t}^B\right)$		-0.47* (0.22)
γ_{19}: $Time^2$ x $\log\,(Buying_{-i,t}^B)^2$		0.09* (0.04)
F-statistics	$p < .001$	$p < .001$
R^2 (within transformation)	0.07	0.12

N = 472. Robust standard errors in parenthesis.
$^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$

Regression 1 shows the main effects results, while Regression 2 shows the results with the interaction terms included in the model. The results in Regression 2 support our hypothesis of positive inter-network externalities (H1). Buyers' preference for Exchange B is negatively related to selling activities in Exchange A (γ_1 = -.13, p < .05), but positively related to selling activities in Exchange B (γ_4 = .78, p < .05).

We find that the time trend variables significantly moderate the relationship between buying activities in Exchange B and buyers' preferences (γ_{16} = 2.16, p < .10, γ_{17} = -.42, p < .10, γ_{18} = -.47, p < .05, γ_{19} = .09, p < .5). Together with the estimates of the main effects of buying activities in Exchange B on buyers' preferences (γ_5 = 1.32, p > .10, γ_6 = -.88, p < .10), these results indicate that buyers' preferences for Exchange B initially increase with increasing buying levels in the exchange but decrease at higher buying levels, and support H2.

The coefficient of the quadratic term for buying activities in Exchange B (log $(Buying^{A}_{-i,t})^2$) represents the rate of change in buyers' preferences for Exchange B for a given change in buying level in that exchange. The larger the absolute coefficient of this quadratic term, the more sensitive are buyers' preferences to changes in buying levels in Exchange B. Based on the estimates for γ_{17} and γ_{19}, the absolute value of the quadratic term for buying activities in Exchange B decreases over time. Therefore, there is evidence that the influences of buying activities on buyers' preferences attenuate with time, supporting H3. All these results are consistent with those that we obtained when we used time period dummies instead.

CHAPTER 4:
DESIGN DISTINCTIVENESS IN DESIGN CONTESTS AND ONLINE ADVERTISING CAMPAIGNS

4.1 Design Distinctiveness and Advertising Performances

In this chapter, we examine why and how advertisers should identify distinctive designs during design contests. The proliferation of crowd-based design contest platforms allows advertisers to tap into a large group of designers with diverse experience and skills. As advertisers only pay for designs that they like in the contests, design contests reduce information asymmetry (where it is difficult for advertisers to *a priori* determine designers' quality) and costs. By making it more affordable and easier for advertisers to source for ad designs, such contest platforms are likely to contribute to the growth in online advertising. The number of banner ads served to online users is substantial, with more than 1.1 trillion display ad impressions delivered to U.S. Internet users during the first quarter in 2011.[22] Advertisers are estimated to spend US$22 billion on online display advertisements by 2015, overshadowing the spending on search advertising.[23]

Although advertisers want to achieve high click-through rates (CTR) with their online ads, existing statistics show that they are typically not too successful in doing so. The estimated CTR is approximately .09% to .2%, and an ad campaign with 2% CTR is consider to be successful.[24] Therefore, it is crucial that advertisers identify and use ads with potentially high click-through performance in their advertising campaigns. For example, advertisers often consider using high quality ad designs that are visually appealing and attractive. Poorly designed ads, such as those with low quality graphics or mismatched color scheme, reflect badly on the advertisers. Although one would expect high quality ads to do well in attracting viewers' attention, recent studies did not find significant relationships between design quality (e.g., raters'

[22] http://www.comscore.com/Press_Events/Press_Releases/2011/5/U.S._Online_Display_Advertising_Market_Delivers_1.1_Trillion_Impressions_in_Q1_2011
[23] http://www.emarketer.com/PressRelease.aspx?R=1008432
[24] http://www.mediamind.com/sites/default/files/MediaMind_Global_Benchmark_Q4_2010.pdf

evaluations of visual appeal and tastefulness of the ads) and CTR (Dow et al., 2010; 2011). As the numbers of online display ads and their design quality have increased over the years, high design quality could be an insufficient condition for superior ad performance. While high design quality might have helped online ads to differentiate themselves during the early days of the Internet, this factor alone may not lead to high click-through performance given the clutter of quality ads that consumers are exposed to.

Facing an increasingly crowded advertising space on the Internet, advertisers should also consider the *design distinctiveness* of the ads in their advertising campaigns. Design distinctiveness is a design's contrastive value relative to other designs (Jacoby and Craik, 1979; Rosenkrans 2009). Ads with distinctive designs evoke a sense of surprise and unexpectedness (Jackson and Messick, 1965), which raise viewers' consciousness and overcome their resistance to the ads (Kover, 1995). Distinctive ad designs also draw more attention and promote viewers' exploration of the ads. In an experiment that used eye-tracking technology, viewers paid more attention to ads that are more unique and distinct (Pieters et al., 2002). Online display ads that used distinctive formats such as animations (instead of static ads) generally performed better in terms of click-through, ad recall, and attitude towards the ads and/or brands (Diao and Sundar, 2004; Li and Bukovac, 1999; Rosenkrans, 2009; Sundar and Kim 2005). At times, however, advertisers cannot use certain ad formats because advertising platforms impose restrictions on permissible ads. For example, Facebook does not support animated or flash ads, and Google limits animation length in banners to 30 seconds.[25] As a result, advertisers compete on a relatively level field in terms of ads formats on these platforms, and they have to use distinctive ad designs (e.g., images, colors, and/or content) to stand out from the competition.

Therefore, among a set of ad designs that are being considered for an advertising campaign, we expect designs that are more distinctive in the set to achieve more click-through than designs that are less distinctive. Furthermore, since the design quality of online ads has is improving, ads with high quality design need to be unique and different from competing ads in order to attract viewers' attention and click-through. Thus we hypothesize that high quality

[25] See http://www.facebook.com/help/?page=245316378826196 for Facebook's advertising guidelines, and http://support.google.com/adwordspolicy/bin/static.py?hl=en&topic=1310862&guide=1308145&page=guide.cs&answer=176108&rd=1 for Google's.

designs that are more distinctive would achieve more click-through than high quality designs that are less distinctive.

H1: Ads with higher (lower) design distinctiveness achieve more (fewer) click-through in online ad campaigns.

H2: High design quality ads with higher (lower) design distinctiveness achieve more (fewer) click-through in online ad campaigns.

To measure design distinctiveness in design contests, advertisers need to compare all the designs that they receive. Designs that are more dissimilar from other designs, on average, should be more distinctive. Given a set of n designs, there are $\frac{n(n-1)}{2}$ pairs of designs to compare. That is, the number of pairwise comparisons increases at a quadratic rate with respect to the number of designs. When n is large, which is not uncommon in design contests, the number of pairwise comparison to make is non-trivial (e.g., Kornish and Ulrich, 2011). For example, in a contest with 120 design submissions, the advertiser needs to compare 7,140 pairs of designs in order to measure the distinctiveness of each design.[26] In this study, we developed a novel approach to efficiently measure distinctiveness of ad designs in design contests. We built a model that estimates a design's distinctiveness using the differences in characteristics among designs. Our approach can potentially be implemented by design contest platforms to help advertisers measure design distinctiveness.

4.2 Design Contests

Examples of design contest platforms that facilitate crowdsourcing of creative designs include 99designs (www.99designs.com) and CrowdSpring (www.crowdspring.com). 99designs was started in 2006 and it has hosted more than 170,000 designs contests as of November 2012. On average, each contest received 112 design entries. In November 2012, 99designs paid $1.5 million to designers. Similarly, CrowdSpring has hosted more than 32,000 contests since 2008, where each contest received more than 140 entries on average.

[26] For example, see http://www.crowdspring.com/project/2309576_new-hat-style-promo-on-home-page/ (141 ad designs received) and http://99designs.com/banner-ad-design/contests/banner-ad-creation-contest-171475 (121 ad designs received).

Here is a typical process in a design contest for banner ads. Before launching a contest, the advertiser decides on the duration of the contest and monetary award for the winning designs. Contests have relatively short durations (e.g., between 3 to 14 days), and most design marketplaces provide recommendations for minimum awards. Next, the advertiser provides a creative or project brief that describes what he looks for in the design. He can specify the content and dimensions of the banner, as well as provide information about his business and target audience. The advertiser can also provide some examples of designs that he likes.

Once the contest is launched, designers can submit their design entries based on the information in the project brief. During the contest, the advertiser can give ratings (e.g., between one to five stars) and feedback for some or all submitted designs at his discretion. At the end of the contest, the advertiser chooses the designs that he wishes to acquire, and awards the prizes to the winning designers. [27]

4.3 Design Contest Experiment

We conducted a design contest where participants ("designers") were asked to design banner ads to promote an online wedding photography directory ("aweddinglist.com"). We recruited participants from various online communities for graphic designers. During the recruitment, we stated that the purpose of the study was to understand how designers create impactful ad designs. We invited all individuals, regardless of their graphic design experience, to participate in this study. We informed potential participants that they would also need to complete two online surveys in this study. To make our experiment realistic to design contests, we did not compensate designers for participating in this study. Instead, the designers who submitted the top three designs would receive between US$250 and US$600. These amounts were within the market rates on various design contest platform at the time of the study. Individuals who were interested to participate in this study could register for it by providing their names and email addresses.

---

[27] Advertisers that are looking for one-time and/or stand alone ad designs would find design contests attractive and appropriate for their needs. Other advertisers that require integrated marketing services (e.g., from developments to implementation of banner ads) or have recurring design requirements may still prefer hiring in-house designers or engaging advertising agencies. Also, crowdsourcing their design tasks may require firms to reveal information about their marketing activities and plans; advertisers that wish to maintain secrecy for strategic reasons may prefer to use other channels instead.

We emailed the registrants once we launched the design contest. The registrants first answered a pre-contest survey, where we asked them design- and contest-related questions. After they completed the survey, we emailed them the passwords to login to the design contest platform that we developed. On the platform, they could see the project brief that described the wedding photography directory and its target audience. We asked designers to submit ad designs for the online directory that are attractive and would achieve high ad recognition performance and click-through rate. In the project brief, we also showed participants some examples of online banners that we selected prior to the experiment. (We manipulated the characteristics of the design examples that we showed to the participants. However, these manipulations are outside the scope of this chapter. In the next chapter, we will discuss in detail the design examples and their impacts.)

We provided the participants a logo for the online directory, and ten photos that they could include in their ad designs (Figure 4.1).[28] These photos show different wedding-related images such as the bride and/or groom (in various poses and different settings), wedding bouquet, and wedding gown. Due to legal and copyright concerns, participants must only use the photos that we provided in their designs. Designers could create and use ad copy, such as tagline and phrases, in their submissions. We specified that the ad dimensions must be 300 (width) x 250 (height) pixels, with file size less than 50kb in file size.

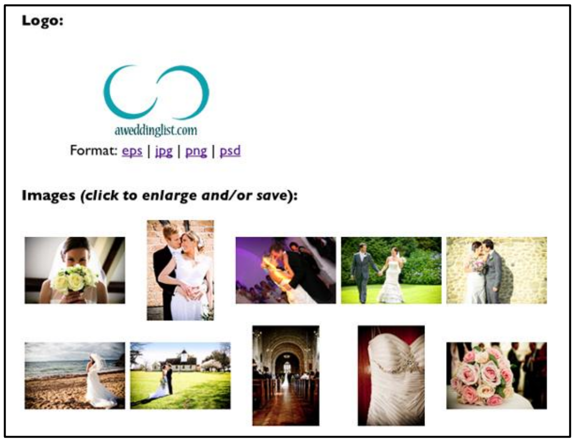


Figure 4.1: Logo and Photos in Project Brief

[28] A wedding photographer, who was blinded to the experiment, granted us the permission to use these photos. This photographer took all these photos during different weddings that she covered.

Designers had ten days to submit their designs through the platform, starting from the day they first logged on to the website. We emailed reminders to designers who had not submitted any designs by the third day. On the eighth day, we notified all designers that there were two more days for them to submit their designs. Designers could access the project brief, logo, and photos any time during the duration of the contest. They could also withdraw designs that they submitted any time before the contest ended. We did not impose any restrictions or quota in terms of the number of ads that designers should submit. We also did not indicate the number of participants who were taking part in this study, nor the number of designs that had been submitted. In addition, designers could not see other designers' submissions during the contest.

252 individuals pre-registered for our study, but only 180 of them completed the pre-contest survey. 105 of these individuals submitted at least one design during our contest. We received 385 ads at the end of the contest. However, 18 ads included photos that we did not provided and/or URL of other websites instead of the online directory that they were supposed to design the ads for. The dimensions of 27 other ads were not within the specified width and/or height. As we could not resize these ads without removing key elements of the ads, we excluded them from our sample. Therefore, we have 340 usable ad designs from 99 participants in our sample (3.43 designs per designer on average).

Among the designers in our sample, 71.7% have or are pursuing graphic design-related certificate or degree programs. 86.9% of these participants were from United States, and the rest were from Canada, India, Indonesia, Jamaica, Malaysia, Nigeria, Pakistan, and Singapore. The average designer had 8.4 years of design experience, took part in 7.3 design contests, and won .8 contests. He also participated in 3.7 wedding-related graphic design projects in the last two years, and had 6.8 design project deadlines over the next 4 weeks during the time of the experiment. Using the non-parametric Wilcoxon-Mann-Whitney test, we compared these statistics with those of the 81 individuals who completed the pre-contest survey but did not submit any designs. The two groups of individuals differ only in terms of design experience: on average, designers who did not submit any designs had fewer years of design experiences (mean = 6.8, std. dev. = 8.5) than those who submitted at least one design (mean = 8.4, std. dev. = 7.5) ($z = -2.309$, $p < .05$).

4.4 Measures

4.4.1 Design Distinctiveness.

A traditional approach of evaluation the distinctiveness of a design is to ask raters to rate the creativity, novelty, and/or originality of a design (e.g., Dow et al., 2010; Heiser et al., 2008). We take a different approach in this study, and developed a three-step procedure to measure design distinctiveness. First, we codified the characteristics of individual designs in our contest. Second, we measured the difference between each pair of designs in our sample by comparing their characteristics. Finally, we measured the distinctiveness score of each design by averaging the design's pair-wise distances with all the other designs in the contest.

Step 1: Codify Design Characteristics. We codified the color scheme, photos, logo, and text in the respective designs.

Color scheme: We extracted the color of each pixel in the designs so that we could quantify the color scheme used in each design. The color of each pixel can be expressed as an RGB triple that represents the amount of red (R), green (G), and blue (B) that the color has. The values for each RGB component range from 0 to 255. For example, black has rgb(0,0,0) whereas white has rgb(255,255,255). Using the RGB decimal values allows us to precisely describe the ad colors. For example, we could distinguish different shades of gray in a design, where some shades are more similar to black (e.g., rgb(24, 24, 24)) and others to white (e.g., rgb(168, 168, 168)). With the RGB decimal values for each pixel in a particular design, we calculated the proportion of each RGB triple in that design.

Photos, logo, and text: During the contest, designers could use different photos that we provided in the project brief for their work. They could also resize the logo, and use different amount of space for the ad copy. Two coders extracted information about these various elements in the designs. On a web-based interface, we displayed the designs individually and a set of "element-boxes" that represent the different elements that appear in the designs (Figure 4.2a). The coders first "activated" the respective element boxes by placing them over the corresponding elements (Figure 4.2b). (We created the element boxes with opacity of 80% in Adobe Photoshop and saved the files in transparent GIF format so the coders could see the actual elements on the designs underneath the boxes.) Next, coders adjusted the size of the element boxes to capture the area of the corresponding elements. Using the height and width of the boxes for logo and text, we obtained the areas occupied by these elements in the designs. We also used the activated element-boxes to infer the *specific photos* and the *number of photos* in each design. The coders coded all the designs individually, and we averaged their scores for each element in the respective designs. The inter-coder agreement is high: The concordance correlation coefficient is 0.99 (Lin 1989, 2000), and the correlation between difference and mean is .04 (where a value near zero implies concordance).

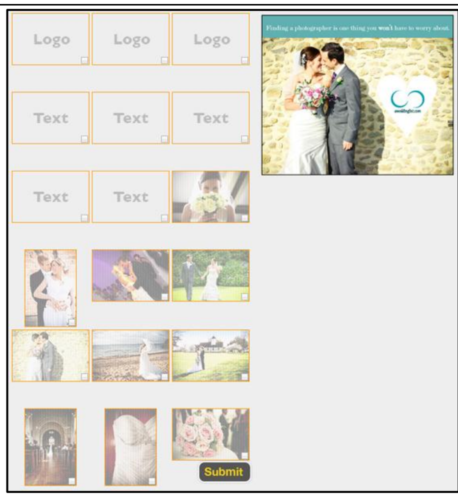


Figure 4.2a

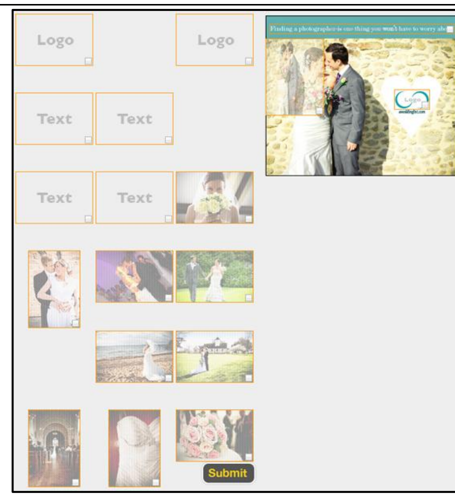


Figure 4.2b

Figure 4.2a The coder sees the ad design (right) and the element boxes (left).

Figure 4.2b The coder drags and places the element onto the design. Next, he resizes the boxes to fit the sizes of the corresponding elements on the design.

<u>*Step 2: Measure Pairwise Differences in Design Characteristics.*</u> Next, we measured the differences in various design aspects between each pair of designs. First, we calculated the difference in color schemes between two designs, *i* and *j*, by:

$$colors_{i,j} = \sum_x \sum_y c_{x,i} c_{y,j} \sqrt{(r_{x,i} - r_{y,j})^2 + (g_{x,i} - g_{y,j})^2 + (b_{x,i} - b_{y,j})^2}$$

where $c_{x,i}$ is the proportion of color x in design i, and $r_{x,i}$, $g_{x,i}$, and $b_{x,i}$ are the respective decimal values for red, green, and blue of color *x* in *i*. (The same description applies to terms with subscripts y and j.) The square-root term is the Euclidean distance between colors x in design *i* and color *y* in design *j*. Colors that are relative similar have a shorter distance between them. We weighted the Euclidean distance by the product of the proportions of the respective colors in each design; we assume that colors that appear frequently in the designs have greater influence on the perceived dissimilarity between the designs. For each pair of designs, we compared every color in one design with all the colors in the other design. We summed all the weighted color comparisons between the two designs to derive the difference in their color schemes. A higher value indicates that the color schemes of the two designs are relatively different.

The pairwise difference in terms of the numbers of photos, size of logo, and size of text-space is expressed by:

$$\ddot{k}_{i,j} = \frac{|k_i - k_j|}{k_i + k_j}$$

where $\ddot{k}_{i,j}$ is the difference between designs *i* and *j* in terms of characteristic *k*. We normalized the absolute difference between designs *i* and *j* in terms of *k* (numerator) by the sum of *k* in the two designs (denominator).[29] Since the designs in our sample have the same dimensions (300 x 250 pixels), we do not need to normalize the differences further by the ad size. $\ddot{k}_{i,j}$ lies between 0 and 1, where a larger value indicates greater difference between the two designs in terms of *k*.

Lastly, we measured the differences in terms of the specific photos in the two designs:

$$\ddot{photos}_{i,j} = \frac{number\ of\ photos\ unique\ to\ i\ only + \ number\ of\ photos\ unique\ to\ j\ only}{number\ of\ distinct\ photos\ in\ i\ and\ j}$$

The value of this measure ranges between 0 (when designs *i* and *j* used the exact same photo images) and 1 (when the two designs have completely different photo images).

Step 3: Measure Design Distinctiveness. The calculation of the distinctiveness of individual designs in the contest involved two stages. First, we compared the designs with one

[29] Suppose there are two pairs of designs, designs A-B and C-D. In the first pair, the sizes of the logo in A and B are 1000 pixel^2 and 950 pixel^2, respectively. In the second pair, the sizes of the logo in C and D are 100 pixel^2 and 50 pixel^2, respectively. Although the difference in the logo sizes in both pairs is 50 pixel^2, this difference is likely to be perceived as larger in the second pair.

another to measure the distances between all pair of designs. Second, we averaged all the pairwise distances for a specific design to obtain its distinctiveness score. Conceptually, the more different a design is from other designs, the higher would be its averaged pairwise distance and the more likely that it would be perceived as distinctive.

Stage I – Estimate pair-wise differences between designs: With 340 designs in our sample, we needed to make 57,630 pairwise comparisons. To do this efficiently, we built a model to estimate the dissimilarity between designs using the differences in design characteristics (Step 2). We randomly chose 74 designs from our sample to use as a learning set for our model. In generating this sub-sample, we selected at most one design from each designer. (We also used stratified sampling such that distribution of designers in the learning set is proportionate to the number of designers in the respective experimental conditions.)

Five raters evaluated the distances (or dissimilarity) among designs in our learning set using the spatial arrangement method (SpAM) (Goldstone, 1994; Hout et al., 2012). Using SpAM, raters arranged multiple stimuli simultaneously such that similar stimuli were placed closer to each other. This approach of collecting similarity/dissimilarity data is relatively fast and efficient. In a lab experiment to scaled 25 to 27 stimuli, participants took about 5 minutes to complete the task using SpAM, and 25 to 30 minutes using traditional pairwise procedure (comparing the similarity of two items at a time using a Likert scale) (Hout et al., 2012). Moreover, Hout et al. (2012) found that the results using SpAM were comparable to those using pairwise procedures. In their study, the correlations between SpAM and pairwise procedure results ranged from .44 to .96, with an average correlation of .60.

We developed a web-based SpAM interface for raters to organize the 74 designs in our learning set (Figure 4.3). Because of the size and numbers of designs, we displayed six randomly selected designs on the webpage at a time. We scaled the designs to 180 x 150 pixels so that raters could work on the task without scrolling the webpage. All raters indicated that they could see all the designs clearly on their screens. We asked raters to arrange the designs on a 750 x 750 pixels white canvas, such that designs that were more similar were to be placed closer together. The raters could do the task at their own pace, and were instructed to keep their arrangements of the designs within the borders of the white canvas. We trained the raters before they started on the actual task to familiarize them with the online interface and procedure.

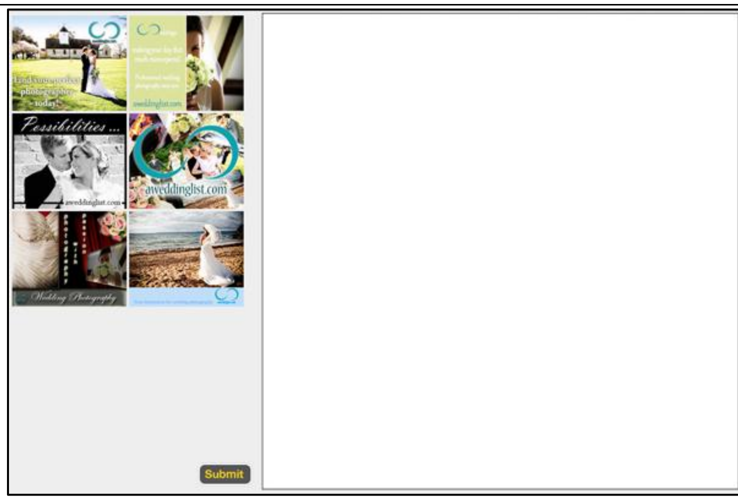


Figure 4.3 Spatial Arrangement Method (SpAM) Interface

Six ads were randomly selected in each round. Raters would drag the ads onto the white canvas, and arrange the ads such that similar ads were placed closer to gather.

On average, each rater organized 245 sets of six designs. Each time a rater completed organizing one set of designs, our system would record the coordinates of the top-left corner of each design on the white canvas. Because the designs have the same dimensions, the distance between the top-left corners of two designs represents the distance between these designs. For each pair of designs, we averaged the distances across all raters to derive its pairwise distance. Table 4.1 shows the descriptive statistics for 2,701 pairs of designs in the sub-sample that we used for the SpAM procedure. The SpAM distance is the pairwise distance that we obtained from the SpAM procedure, and the other variables were based on what we obtained in Step 2, where we measured the differences in characteristics between pairs of designs.

Table 4.1 Descriptive Statistics and Correlation Matrix for Estimating Pairwise Distances

SpAM Distance	1.00					
Diff. in Color Scheme	0.19	1.00				
Diff. in Photos Used	0.51	0.08	1.00			
Diff. in No. of Photos	0.17	-0.10	0.07	1.00		
Diff. in Logo Size	0.12	0.03	0.02	0.04	1.00	
Diff. in Text Area	0.00	-0.09	-0.02	0.07	-0.03	1.00
Mean	389.44	79.13	0.87	0.30	0.53	0.41
Std. Dev.	94.01	36.52	0.25	0.31	0.31	0.27
Min	77.66	21.15	0.00	0.00	0.00	0.00
Max	614.54	280.63	1.00	1.00	1.00	1.00

We regressed the SpAM distances on the differences in design characteristics between pairs of designs using Stata (version 12.1). The differences in all the design characteristics, except that for the area of text in the designs, are positively and significantly related to the pairwise distances (Table 2). Our model explains almost a third of the pairwise distances variability (adjusted R-squared = .31). Based on the beta (standardized) coefficients, we find that the differences in the specific photos used in the designs have the strongest effects on pairwise distances (beta = .48). This is followed by the differences between the designs in terms of color scheme (beta = .17), number of photos (beta = .15), and logo size (beta = .10).

Table 4.2 Results for Estimation of Pairwise Designs

DV: SpAM Distance	Coef.	Beta
Constant	163.86^{***} (7.29)	-
Diff. in Color Scheme	0.43^{***} (0.04)	0.17
Diff. in Photos Used	183.24^{***} (6.13)	0.48
Diff. in No. of Photos	44.54^{***} (4.93)	0.15
Diff. in Logo Size	30.72^{***} (4.83)	0.10
Diff. in Text Area	5.58 (5.70)	0.02
F-statistics	244.14^{***}	
Adjusted R-squared	0.31	
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$		

Using the unstandardized coefficients from the regression, we calculated the 57,630 pairwise distances for our sample of 340 designs. To validate our model, we randomly selected five pair of designs within the 95% confidence interval at the 5th, 25th, 50th, 75th and 95th percentile in terms of the estimated pairwise distance. We also selected five pairs of designs with the highest estimated distance, and five pairs with the lowest estimated distance.[30] We recruited users on Amazon Mechanical Turk (MTurk) to rate the similarity of these 35 pairs of designs on a 7-point scale ("not similar at all… extremely similar"). We reverse-coded the scores to obtain the dissimilarity scores. The correlation between the MTurk scores and our model-estimated pairwise distances is .91 ($p < .001$).

The SpAM procedure and the MTurk validation used subsets of designs from our full sample, which resulted in restrictions of range in these sub-samples (Table 3). After accounting for restricted range (Cohen et al., 2003), the estimated correlation between the model-estimated pairwise distances and SpAM ratings is .57, and that between the model-estimated pairwise distances and MTurk ratings is .87. These correlations are relatively strong, providing support for the validity of our model and estimates.

Table 4.3 Correlations of Distances between Model-Based Estimation and (i) SpAM and (ii) MTurk.

Sample	Model-Estimated Pairwise Distances			Correlation with Model-Estimated Pairwise Distances	
	Min	Max	Std. Dev.	Not Adjusted for restricted range	Adjusted for restricted range^
Full	175.6754	750.9149	54.34979	-	-
SpAM	183.6835	541.679	52.48868	.56	**.57**
MTurk	175.6754	541.679	93.57102	.91	**.87**

^ Range-adjusted correlations are estimated based on Cohen et al. (2003).

Stage II – Calculate distinctiveness score: To measure the distinctiveness of each design, we averaged its estimated pairwise distances with all other designs in our sample. A higher distinctiveness score implies that the design is, on average, relatively dissimilar to other designs.

4.4.2 Advertising Campaign Performance

We launched an advertising campaign on Google Display Network (GDN) to collect data on the actual market performance of the designs. GDN allowed us to include multiple ads in one advertising campaign. However, having more than 80 ads in one campaign might be difficult to

[30] There is a high concentration of certain designs at these extreme ends, particularly among pairs with the largest pairwise distances. For example, two particulars designs appear in the first 46 pairs with the longest pairwise distances. Hence we chose five pairs of designs at the extreme ends such that all chosen designs were selected only once.

manage from an experimental standpoint.[31] We decided to launch a campaign with the 74 designs that we used to estimate the pairwise distances. (We could not select the ads for the GDN campaign using the distinctiveness scores because we launched the advertising campaign while we were building our pairwise distance estimation model.) The mean and standard deviation of the distinctiveness scores of this sub-sample are very similar to those for the full sample (Table 4). Furthermore, the distinctiveness scores range in this sub-sample is wider than the 3th (366.66) to 98th (425.99) percentile in the full sample. Hence the 74 designs is representative of our sample for the advertising campaign.

Table 4.4 Comparison between Samples

Distinctiveness Score	Full sample (340 ads)	Sub-sample/Learning Set (74 ads)
Mean	388.38	389.25
Std. Dev.	16.30	14.85
Min, Max	360.47, 490.91	366.60, 456.55

In GDN campaign settings, we chose the option to rotate all ads in the campaign more evenly; the other options were for Google to optimize the campaign for clicks or conversions. Even though GDN indicated that our ads might not perform well should we rotate the ads evenly in our campaign, this option removed a possible confound that the click-through performance could be driven by GDN's algorithm. We also chose to place our ads on wedding-related websites in GDN. This minimized the likelihood that our ads would appear in irrelevant websites, which could affect ad performance.

We ran the advertising campaign for 11 days in September 2012, and tracked the number of impressions and click-through that each ad received. Each ad received between 634 and 2,090 impressions (mean = 1352.01, std. dev. = 377.80), and between zero and six click-through (mean = 1.57, std. dev. = 1.38). 18 ads in our sample (24.3%) received no click-through. The CTR for individual ads ranged from 0% to .32%. At the campaign level, we received 100,049 impressions and 115 clicks in total, achieving a CTR of 0.11%.

4.4.3 Design Quality

Six advertising industry professionals ("experts") evaluated the 340 designs in terms of the criteria that we stated in the project brief. We randomly assigned between 170 and 228

[31] Per conversation with Steven Dow in May, 2012.

designs to each expert. (Due to work commitment, one of the experts dropped out after evaluating 84 designs.) The experts rated the designs remotely through our experiment website. To make the evaluation task manageable for the experts (given their work commitment), we allowed them to complete the evaluation over a few days.

When the experts logged on to the design evaluation website for the first time, we showed them the project brief and details about the judging criteria. First, the experts were to evaluate a design's *attractiveness* (or how visually appealing the design is) on a scale of 0 (not attractive at all) to 100 (extremely attractive). Second, to rate the *potential ad recognition performance* of the designs, they were to estimate that the likelihood that a potential user would recognize a design one week after he or she see it for the first time. Lastly, the experts were to predict the *potential click through rate (CTR)* of the designs, between 0% to 5%, in an online campaign on GND. (We highlighted to the experts that the 0-5% CTR range was typical in online ad campaigns.)

After they had read the project brief and judging criteria, the experts went through a training phase to evaluate ten designs from our sample. This phase exposed the experts to the variety of designs in our sample, and gave them an opportunity to calibrate their individual scoring. (We discarded these scores at the end of the training phase for each expert.) After the experts completed the training phase, the website directed them to begin the actual evaluations. In both training and actual evaluations, the experts rated on the three criteria simultaneously for one randomly chosen design at a time (Figure 4.4).

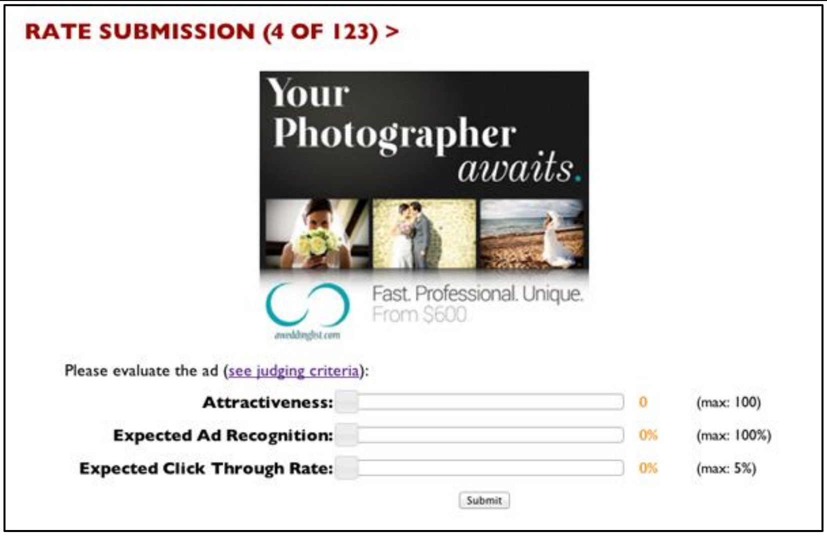


Figure 4.4 Expert Evaluation of Designs

Note: Experts could access the judging criteria any time during the evaluation process.

The experts were reasonably consistent in their evaluations over time despite the number of designs that were assigned to them. Depending on the number of designs that were assigned to them, the experts rated up to ten designs twice during the actual evaluation process. The second evaluation of these designs served as retests and allowed us to check within-rater reliability. We scattered the ten retests throughout the actual evaluation process, approximately one retest point after every 15 to 20 evaluations. When a retest point was reached, a design was randomly chosen from all the designs that the expert had rated up to that point but excluding the last fifteen that she rated. As such, all professional were tested using different designs at the same retest interval. The correlations between the first and second evaluations of the designs in terms attractiveness, predicted ad recognition performance, and predicted CTR are .769, .724, and .791, respectively (n = 54; all significant at $p < .001$).

We averaged the experts' ratings for the respective criteria, and then divided the scores for attractiveness and ad recognition by 20 so that the three criteria have the same range (0 to 5). We then created a design quality index by averaging the scores of the three criteria for each design (Cronbach's alpha = .88). We assumed the three criteria have equal weights in this index. (However, we would examine the individual relationships between the criteria and design distinctiveness in Section 4.5.2.)

4.5 Results and Analyses

4.5.1 Main Results

Table 4.5 shows the descriptive statistics and correlation matrix of the variables. We mean-centered all the independent variables.

Table 4.5 Descriptive Statistics and Correlation Matrix for Estimating No. of Click-Through

No. of Click-Through	1.00						
No. of Impressions ('000)	0.55	1.00					
Design Distinctiveness	0.10	0.10	1.00				
Design Quality	0.07	0.00	-0.04	1.00			
Design Attractiveness^	0.06	-0.06	-0.05	0.95	1.00		
Predicted CTR	0.10	0.04	-0.10	0.84	0.70	1.00	
Predicted Ad Recognition^	0.05	0.04	0.03	0.95	0.86	0.72	1.00
Mean	1.57	0.00	0.00	0.00	0.00	0.00	0.00
Std. Dev.	1.38	0.38	14.85	0.42	10.97	0.30	10.08
Min	0.00	-0.72	-22.65	-0.80	-21.75	-0.54	-17.42
Max	6.00	0.74	67.30	0.99	26.34	0.92	23.06

N = 74
^ The ranges for these variables were scaled from 0-100 to 0-5 when we computed an index for Design Quality. We show these variables at their original scales in this table, and used them as such in our robustness analyses.

Using linear regression, we first estimated a base model where we regressed the number of click-through on the number of impressions (in thousands) and design quality index (Model 1 in Table 4.6). We controlled for the six conditions in our experiment with five dummy variables (*c0*, *c2low*, *c2high*, *c4low*, *c4high*).[32] This model explains 29% of the variance in the number of click-through (adjusted R-squared = .29). On average, the ads get two clicks for every thousand impressions ($\beta = 1.99$, $p < .001$). The relationship between number of click-through and design quality is not statistically significant at .05 levels.

[32] In the experiment, participants were randomly assigned to six conditions that differed in the number and design variability of examples that they saw in the project brief. These conditions are no examples (*c0*), one example, two examples with low design variability (*c2low*), two examples with high design variability (*c2high*), four examples with low design variability (*c4low*), and four examples with high design variability (*c4high*). The base group in our regressions is the one-example condition.

Table 4.6 Results for Estimation of No. of Click-Through

DV: No. of Click-Through	Model 1	Model 2	Model 3
	Coef.	**Coef.**	**Coef.**
Constant	1.13^{**} (0.31)	0.95^{**} (0.3)	1.04^{**} (0.31)
No. of Impressions (‘000)	1.99^{***} (0.37)	2.01^{***} (0.35)	1.95^{***} (0.37)
Design Quality	0.38 (0.35)	0.62^{+} (0.34)	0.66^{+} (0.35)
c0	0.64 (0.43)	0.62 (0.41)	0.65 (0.41)
c2low	0.75^{+} (0.44)	0.83^{*} (0.41)	0.84^{*} (0.41)
c2high	0.02 (0.50)	-0.25 (0.49)	-0.26 (0.50)
c4low	0.74 (0.50)	0.5 (0.48)	0.42 (0.49)
c4high	0.36 (0.48)	0.43 (0.46)	0.36 (0.46)
Design Distinctiveness		-0.02^{*} (0.01)	-0.02^{+} (0.01)
Design Distinctiveness^2		0.001^{**} (0.000)	0.001 (0.001)
Design Distinctiveness x Design Quality			-0.02 (0.03)
Design Distinctiveness^2 x Design Quality			0.000 (0.001)
F-statistics	5.32^{***}	5.87^{***}	4.88^{***}
Adjusted R-squared	0.29	0.38	0.37
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$			

Next, we included design distinctiveness in quadratic form in Model 2. This specification allows for a non-linear relationship between design distinctiveness and click-through performance. Relative to Model 1, Model 2 explains an additional 9% of the variance in the number of click-through (adjusted R-squared = .38). The variance inflation factor ranged from 1.10 to 2.07, which suggests that multicollinearity may not be influencing our regression estimates. After controlling for design distinctiveness, design quality now has a moderately positive impact on number of click-through ($\beta = .62$, $p < .10$). We find a significant non-linear impact of design distinctiveness on number of clicks ($\beta = -.02$, $p < .05$ for *design distinctiveness*, and $\beta = .001$, $p < .01$ for *design distinctiveness^2*). The results indicate a U-shaped relationship between design distinctiveness and click-through. Hence, ads have to be significantly distinctive in order to attract viewers’ attention and clicks. Being marginally or incrementally different from competing ads may not be sufficient for ads to stand out in a cluttered online advertising space.

Lastly, we added interaction terms between design distinctiveness and quality in Model 3. These interactions are not statistically significant at the .10 levels. Hence there is no evidence that design distinctiveness moderates the relationship between design quality and click-through performance.

4.5.2 Robustness Analyses

Earlier field research that examined ad distinctiveness only included ads that were implemented in the marketplace (e.g., Pieters et al., 2002); ad designs that were in the advertisers' consideration sets but not chosen could not be observed. Similar to the experiment design in Dow et al. (2010; 2011), we used all ad designs in our consideration set during the advertising campaign. Hence we overcame potential sample selection bias, where the performances of non-selected designs were unobserved. This research design therefore increases the validity of our findings.

Nevertheless, we checked the sensitivity of our results to different model specifications. In our main analyses, we used an index of design quality that consisted of the three criteria evaluated by the experts (design attractiveness, predicted CTR, and predicted ad recognition performance). When computing the design quality index, we scaled down the measures for design attractiveness and predicted ad recognition performances so that they had the same range as the measure for predicted CTR. We also assigned equal weights to these three criteria. We now examine the individual impact of these criteria on design distinctiveness (Table 4.7). In Model 1, we replaced the design quality index with the three criteria in their original scales. Experts' evaluation of design attractiveness has a weak positive effect on number of click-through ($\beta = .05$, $p < .10$), whereas their predictions of CTR and ad recognition performance are not statistically significant.

We then added the quadratic form for design distinctiveness in Model 2. The results are similar to what we obtained in the main analyses. We find a significant U-shaped relationship between design distinctiveness and the number of click-through ($\beta = -.02$, $p < .10$ for *design distinctiveness*, and $\beta = .001$, $p < .01$ for *design distinctiveness*^{2}). In addition, the estimate for design attractiveness is now significant after we included design distinctiveness in the model ($\beta = .05$, $p < .05$).

We also added the interactions between the individual criteria and distinctiveness (the results are not shown in the interest of space). None of these interactions are statistically significant. Therefore, we did not find any evidence that design distinctiveness moderates the relationship between components of design quality and number of click-through in this study.

Table 4.7 Robustness Analysis Results

DV: No. of Click-Through	Model 1	Model 2
	Coef.	**Coef.**
Constant	1.04^{**} (0.31)	0.87^{**} (0.30)
No. of Impressions (‘000)	2.08^{***} (0.37)	2.09^{***} (0.36)
Design Attractiveness	0.05^{+} (0.03)	0.05^{*} (0.02)
Predicted CTR	0.46 (0.67)	0.23 (0.64)
Predicted Ad Recognition	-0.05 (0.03)	-0.03 (0.03)
c0	0.79^{+} (0.44)	0.75^{+} (0.41)
c2low	0.92^{*} (0.44)	0.98^{*} (0.42)
c2high	0.06 (0.50)	-0.16 (0.49)
c4low	0.82 (0.50)	0.6 (0.48)
c4high	0.37 (0.48)	0.46 (0.46)
Design Distinctiveness		-0.02^{+} (0.01)
Design Distinctiveness^2		0.001^{**} (0.00)
F-statistics	4.62^{***}	5.13^{***}
Adjusted R-squared	0.31	0.38
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$		

4.6 Discussion

Crowd-based design contests allow advertisers to acquire a large number of designs that they can consider for use in advertising campaigns. The likelihood that advertisers would receive some novel and distinctive designs during the contests increases with the number of design submissions. The results in this study show that design distinctiveness matters in online advertising campaigns. We find that ads that were more distinctive performed better than those that were less distinctive in an advertising campaign on Google Display Network. However, the large number of design entries in contests brings along a challenge in measuring distinctiveness: The number of pairwise comparisons that is needed to determine distinctiveness increases at a quadratic rate with the number of designs. Hence as more designs are received in the contest, trying to identify distinctive designs is like searching for a needle in the haystack.

In this chapter, we presented an efficient approach to measure design distinctiveness in design contests. Our procedure required human intervention only for codifying individual designs, while it used an algorithm for the more resource intensive task of estimating pairwise differences between designs. (It is also possible to write a program to automate the design codification process, but a discussion of how this automation could be done is beyond the scope of this study.) Hence this method is scalable and can handle a reasonably large number of design submissions in contests. However, there are certain limitations in our current procedure. For example, if designers were allowed to use images from other sources (e.g., stock images) in their work, it would make it more challenging to measure distinctiveness using our approach. Future studies should look into extending our procedure to contests that impose fewer restrictions in terms of design elements that could be included in the submissions.

Nevertheless, design contests platforms can consider implementing our approach to help advertisers identify distinctive designs among the received entries. In most contests for banner ad designs, the advertisers provide a set of graphic resources (e.g., logos and images) that designers can incorporate into their submissions. Moreover, because of copyright and legal concerns, designers are discouraged from using graphics from other sources. For example, 99designs.com recommends designers not to use stock images or clip art in the designs that they submit, and designers who failed to follow rules around stock images may risk having their account "suspended or closed forever".[33] Such design constraints thus make our procedure appropriate in contests for ad designs.

The performance data in our advertising campaign shows that selecting the right ads during design contests is important. Almost one-quarter of the ads in our advertising campaign received no click-through, and only 11 ads (14.8%) achieved CTR greater than the high-end of the industry average of .20%. Among ads with non-zero clicks, the CTR of the best performing ad (.35%) is 2.5 times higher than the average CTR (.14%), and 7 times higher than the CTR for the worst performing ad (.05%). Therefore, choosing the right (or wrong) ad designs in the contests would impact the advertisers' payoff significantly. Yet, this is not an easy task for advertisers: The rules on design contest platforms require advertisers to decide on the wining designs within a certain number of days, and they cannot test-drive design submissions before making their decisions.

We believe our approach can help advertisers make better decisions in the ad design selection process. We propose that advertisers use our measure of design distinctiveness in

[33] http://99designs.com/help/stock-image-and-clip-art-policy

conjunction with other selection criteria, such as design attractiveness and quality of ad copy. In doing so, it increases the likelihood that they would select designs with potentially higher click-through performance. For instance, an advertiser could shortlist a set of ad designs that are highly distinctive in the contest, and then choose the one that she likes the most based on the other criteria. Alternatively, she can reverse the process: She first picks out the designs that she likes, and then selects the one that is most distinctive in the contest. In either case, it is critical that the advertiser can measure and pick out distinctive designs during the contest.

Lastly, we find that attractiveness has a positive impact on the number of click-through after we controlled for design distinctiveness ($\beta = .05$, $p < .05$; see Model 2 in Table 4.7). This finding suggests that advertisers should also consider design attractiveness when selecting ads in design contests. However, we need to interpret this result cautiously. In this study, design attractiveness was evaluated by a group of advertising professionals. In contrast, the typical advertisers-clients in design contests are likely to be non-professionals (in advertising) and lack the background and experience in evaluating ad designs. Dow et al. (2010) found that clients and advertising professionals evaluated the same set of ad designs differently in their study: Clients gave higher ratings on average than professionals did. Hence, a design that is not so attractive from the professionals' perspective may appear attractive to a client, who would then expect high click-through performance from this particular design. Advertisers who are sourcing for ad designs through contests should therefore be aware that selecting designs based on their judgments of attractiveness might not necessarily lead to the desired outcomes.

4.7 Interlude

Up to this point, we described a novel way to measure design distinctiveness in design contests. We also showed that design distinctiveness is a driver of click-through performance in advertising campaigns. A relevant question then is: how can advertisers get designers to produce distinctive design in design contests? We will touch on this question in the next chapter.

References

Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Routledge Academic, 2003.

Diao, F. and Sundar, S. S. "Orienting Response and Memory for Web Advertisements:Exploring Effects of Pop-Up Windowand Animation," *Communication Research* (31:5), 2004, pp. 537-567.

Dow, S. P., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D. L. and Klemmer, S. R. "Prototyping Dynamics: Sharing Multiple Designs Improves Exploration, Group Rapport, and Results", 2011.

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L. and Klemmer, S. R. "Parallel prototyping leads to better design results, more divergence, and increased self-efficacy," *ACM Trans. Comput.-Hum. Interact.* (17:4), 2010, pp. 18:1-18:24.

Goldstone, R. "An efficient method for obtaining similarity data," *Behavior Research Methods* (26), 1994, pp. 381-386.

Hout, M. C., Goldinger, S. D. and Ferguson, R. W. "The Versatility of SpAM: A Fast, Efficient, Spatial Method of DataCollection for Multidimensional Scaling," *Journal of Experimental Psychology: General*, 2012.

Jackson, P. W. and Messick, S. "The person, the product, and the response: conceptual problems in the assessment of creativity1," *Journal of Personality* (33:3), 1965, pp. 309-329.

Jacoby, L. L. and Craik, F. I. M. "Levels of Processing in Human Memory", *in* Cermak, L. S. and Craik, F. I. M., ed.,, Lawrence Erlbaum Associates, Hillsdale, NJ, 1975, pp. 1-21.

Kornish, L. J. and Ulrich, K. T. "Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas," *Management Science* (57:1), 2011, pp. 107-128.

Kover, A. J. "Copywriters' Implicit Theories of Communication: An Exploration," *Journal of Consumer Research* (21:4), 1995, pp. 596-611.

Li, H. and Bukovac, J. L. "Cognitive Impact of Banner Ad Characteristics: An Experimental Study," *Journalism & Mass Communication Quarterly* (76:2), 1999, pp. 341-353.

Pieters, R., Warlop, L. and Wedel, M. "Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity for Brand Attention and Memory," *Management Science* (48:6), 2002, pp. 765-781.

Rosenkrans, G. "The Creativeness and Effectiveness of Online Interactive Rich Media Advertising," *Journal of Interactive Advertising* (9:2), 2009, pp. 18-31.

Sundar, S. S. and Kim, J. "Interactivity and Persuasion: Influencing Attitudes with Information and Involvement," *Journal of Interactive Advertising* (5:2), 2005, pp. 5-18.

CHAPTER 5: IMPACT OF CLIENT-PROVIDED DESIGN EXAMPLES IN DESIGN CONTESTS

5.1 Introduction

In Chapter 4, we developed an approach to measure design distinctiveness of ads in design contests. We found that ads that were more distinctive obtained more click-through in an online ad campaign than those that were less distinctive. Earlier research also found other benefits of distinctive ads: such ads achieved higher recall of advertised claims, better attitudes toward the ads and brands, and higher purchase intention compared to competing ads (Heiser et al., 2008; Keller, 1991). Therefore, advertisers sought to use distinctive ads to differentiate from competition and obtain better responses from users (Diao and Sundar, 2004; Li and Bukovac, 1999; Sundar and Kim 2005).

Since design distinctiveness matters, how can advertisers influence designers in design contest platforms to produce more distinctive design? In this chapter, we focus on drivers of design distinctiveness in crowd-based design contests. In design projects, advertisers or clients often provide examples of designs that they like. We investigate how such clients-provided examples influence creative processes and design outcomes in design contests. Specifically, we examine how the number, quality, and design variability of these examples affect designers' exploration for design concepts and their design submissions in the contests. We also look at how the characteristics of designers' exploration and work relate to design distinctiveness.

5.2 Design Examples

Design examples provide inspirations for creative works, and help designers to assess the originality of their ideas and identify flaws or limitations to avoid (Herring et al., 2009). Designers can also study how others have approached a design problem by referring to existing designs. One source of design examples is through designers' self-exploration of the design solution space. For example, in a project to design banner ads for a wedding photography-related website, designers could search for relevant banners ads on the Internet. Another source of design examples is from the project clients. In many design projects, clients provide examples of designs that they like (see Figure 5.1), and these examples could serve as a starting point for designers in their search for potential design solutions.

What 3 things would you like to communicate to your audience through your ad banner design?

The simplicity of the use of the product
The functionality size of the product
The ease of signing up and getting it operational.

How is your company different from your competitors?

Relatively unique offering in the marketplace as a SAAS pick pack and dispatch software solution.

What ad banner design styles do you like?

Like these:
http://www.digitalwaybill.com/index.php/technol...

http://www.sonyericsson.com/cws/products/mobile...

http://www.ibm.com/systems/info/aprannounce.htm...

Figure 5.1 A Creative Brief in CrowdSpring[34]
This figure shows a partial screenshot of a project brief provided by a client during a design contest. The client provided information of his design preference, and listed three examples of banner design styles that he liked.

5.3 Deviatory Exploration and Design Deviation

Two important activities in creative tasks are the exploration and exploitation of solution spaces. Audia and Goncalo (2007:3) contrasted exploration and exploitation, and related these activities to creative ideas:

> *"Like divergent creativity, exploration involves the search for knowledge that departs from an established direction, the potential generation of a completely new principle, and breaking with accepted modes of thought. Like incremental creativity, exploitation involves continuity with existing solutions, improvement through modification, and generating ideas within an established framework."*

In creative design task, exploration involves searching for major new design concepts that are applicable for the project objective (e.g., Dorst and Cross, 2001; Dow et al., 2010). By exploring the design solution space, a designer is more likely to identify unique concepts that few other designers might consider in design contests. In contrast, exploitation involves applying

[34] http://www.crowdspring.com/project/2294683_banner-for-pick-pack-and-dispatch-marketing-site/details/

patches to initial or existing design concepts to achieve slightly improved versions (Ball et al., 1994; Ullman et al., 1988). Typically, exploitation is comparatively less costly than exploration in design projects; creating multiple designs based on a particular concept requires less resource and efforts than generating the same number of designs using alternative and distinctive concepts. However, it is crucial that designers achieve a balanced exploration-exploitation (Fricke, 1996): Designers become fixated on the solutions too early when there is insufficient search in the solution space, but they have to spend time managing a diverse set of designs rather than improving on specific alternatives when there is excessive exploration.

We are interested in how designers' exploration in design contests is affected by examples that clients provide. <u>*Deviatory exploration*</u>, in our context, refers to the degree to which a designer considers design concepts that deviate from client-provided examples. By proactively searching for design concepts that differ from the examples, designers could tap into concepts that are being considered by few other designers. While these concepts may not necessarily be novel or original in the marketplace, they are likely to be distinctive from the concepts that other designers are thinking about in a contest.[35]

> *H1: Deviatory exploration is positively related to design distinctiveness in design contests.*

We are also keen in how client-provided examples influence <u>*design deviation*</u> of designers' work, or the extent to which a design submission deviates from the examples. Whereas exploration is an activity in the design ideation stage, design deviation pertains to outcomes in the design production stage. Looking at activities in these stages concurrently gives us a greater insight into the creative process. Theoretically, ideation and production activities are related in creative design projects – concepts from the exploration stage should be incorporated in the design output. However, due to different constraints in respective stages, there could be a disconnection between designers' exploration and their creations. For example, design specifications (such as dimensions and sizes) could cause designers to discard certain design concepts that they came across during the ideation stage. As we discuss below, designers could

[35] Designers could consider well-established concepts that differ from client-provided examples. In this case, while designers are engaged in deviatory exploration at the contest level, they are not exploring at the broader, marketplace level.

strategically submit work that conform to or differ from design examples. By using design concepts that are dissimilar to the client-provided examples (i.e., high design deviation), designers are more likely to submit relatively unique designs in the contests.

H2: Design deviation is positively related to design distinctiveness in design contests.

5.4 Impact of Design Examples on Designers' Behaviors

5.4.1 Design Quality of Examples

Advertisers like to receive high quality submissions that are attractive and appropriate for their needs in design projects. Therefore, they usually provide design examples that they think reflect such characteristics. Although advertisers would not intentionally provide low quality examples, the examples that they choose could be perceived as such by designers. This is because the evaluation of design concepts, problems, and solutions by designers and non-designers could and often differ. Due to their design training and experience, designers develop certain guiding principles or problem paradigms that affect how they approach design problems and develop solutions (Darke, 1979; Lloyd and Scott, 1995). Therefore, clients who lack such background may not evaluate design concepts the way designers do.

The quality of client-provided examples should affect the deviation in designers' exploration and work. When designers perceive the examples to be unattractive or inappropriate for the design projects, they are less likely to use concepts from these examples. Instead, they would search for other design concepts for ideas and inspirations. In contrast, designers would build upon client-provided examples and use relevant concepts when the examples appear to be of high quality. (As the saying goes, imitation is the sincerest form of flattery.) In this case, they would reduce their search for alternative design examples, which is a relatively costly process with uncertain payoff during design contests.[36]

H3: Quality of client-provided examples is negatively related to deviatory exploration.

---

[36] Designers cannot recover incurred costs in contests if they do not win. Furthermore, they face resource constraints because they often have to manage various design projects and deadlines concurrently. For example, designers in our sample had 6.8 project deadlines, on average, over a four-week period at the time of our experiment.

H4: Quality of client-provided examples is negatively related to design deviation.

Generally, designers look for certain numbers of design concepts in the ideation stage of a project. When a client provides few examples, designers need to conduct a baseline amount of search regardless of the example quality. We expect the level of deviatory exploration to be more sensitive to example quality as the number of client-provided examples increases. With more high quality examples in the contest, designers could undertake less deviatory exploration and instead use the provided examples in their designs. This approach lowers their search costs during the contest. Conversely, a large number of low quality design examples could drive designers to search more extensively for design concepts that differ from these examples.

H5: As the number of client-provided examples increases, deviatory exploration decreases (increases) when example quality is higher (lower).

5.4.2 Design Variability of Examples

In design projects, clients often provide multiple design examples that they like (see Figure 5.1). Design variability of examples describes the degree to which the examples differ in design concepts such as layout, colors, and/or content. Client-provided examples could be relatively similar (low design variability) in some cases, and highly varied (high design variability) in other cases. The impacts of design variability of examples on designers' exploration and work depend on how examples affect (i) designers' perception of clients' design preference, and (ii) designers' desire to differentiate from competition.

<u>*Design examples as signals of clients' design preference*</u>. In design contests, designers often fix their eyes on the prizes, and winning the contests is a key goal for them. As a designer commented after the experiment:

> *"As I began to grow more cognizant of this experiment, I realized the main reason behind my motivation – winning the prize."* – Designer A (4 years graphic design experience; participated in 4 contests and won 1.)

Hence signals of clients' design preference are non-trivial in design contests as they indicate the types of designs that could be chosen as winning designs. These signals could influence and guide designers' exploration and work in the contest. When design variability of examples is low, the client's design preference could be perceived to be highly specific. Given this signal of the client's preference, designers may restrict their exploration of the design solution space, and mainly refer to design concepts in the examples. Doing so helps to align their designs with the client's preference and increase their chances of winning the contest. On the other hand, when design variability of examples is high, the client's preference may appear unclear and non-specific to designers. As such, designers have more room to consider and use a variety of design concepts. Hence, viewing examples as signals of clients' design preference in design contests, we hypothesize:

> *H6a: Design variability of client-provided examples is positively related to deviatory exploration.*

> *H7a: Design variability of client-provided examples is positively related to design deviation.*

<u>Design examples as benchmarks for creativity.</u> Many designers consider themselves as creative artists, and often set extremely high design standards and goals for themselves (Cross, 2003; Lawson, 1994). They also have personal design preferences and principles that guide their creative work (Darke, 1979; Lloyd and Scott, 1995). Hence, while winning design contests is important to them, they also strive to come up with novel and original work as part of the process. When a client provides examples with highly similar design concepts, designers may expect their competitors to submit designs that incorporate these concepts (see H6a and H7a). In order to stand out from the crowd, they would therefore consider alternative concepts (greater deviatory exploration) and/or submit work that differs from the examples (greater design deviation).[37] For example, a designer in our experiment remarked:

> *"I decided to choose a design style that did not go with what the examples conveyed because of my own perception and taste regarding design and typography. I also challenged myself to create something unique and conceptually driven that is aesthetically pleasing to my eye." – Designer B (5 years graphic design experience; participated in 2 contests and won 0.)*

In this case, the expected influences of design variability of client-provided are contrary to those when examples are perceived as signals of clients' design preference:

[37] Appreciate Mark Fichman for bringing up this line of thought.

H6b: Design variability of client-provided examples is negatively related to deviatory exploration.

H7b: Design variability of client-provided examples is negatively related to design deviation.

Because of the contests' incentive structures (winners take all) and designers' motivation in creative work (winning in style), the effects of design examples as signals of clients' preference and creative benchmark are related and exist simultaneously. One designer expressed the tension between these two effects during the contest:

> *"Seeing an example or a reference would have given me some more direction as to how I was going to design [the ad], but I also believe sometimes following an example [would lead to] a lot of close minded designs based off the example." – Designer C (3 years graphic design experience; participated in 1 contest and won 0.)*

In any case, we expect the signal and benchmark effects to be stronger when there are more client-provided examples. For instance, the indication of what a client likes becomes more salient when she provides more examples that are relatively similar. This could cause designers to deviate lesser (examples as signals of clients' preference) or more (examples as creative benchmark) from the examples than had she provided fewer of such examples. Hence while we hypothesize the deviations in designers' exploration and work to be more sensitive to design variability of examples when clients provide more examples, we leave the directions of the moderation effects as an empirical question.

H8: Deviatory exploration is more (less) sensitive to design variability of examples when the client provides more (fewer) examples.

H9: Design deviation is more (less) sensitive to design variability of examples when the client provides more (fewer) examples.

5.5 Design Contest Experiment

<u>*Design Contest*</u>. Data was collected from the design contest as described in Chapter 4.

Client-Provided Design Examples. The stimuli in this experiment were design examples that we showed to designers in the project brief. To manipulate design variability of examples, we used alternative categories of design examples, where examples within each category were relatively similar (low design variability), and examples across different categories were relatively dissimilar (high design variability). The four categories that we used in our experiment were (i) ads with collages, (ii) ads with wedding bouquet as focal point, (iii) ads with greenery background, and (iv) ads with top-and-bottom frames. Each category consisted of six banner ads promoting wedding photography services that we found on the Internet (Appendix 5.A).

Random Assignment. When a designer logged on to our contest website for the first time after they completed a pre-contest survey, the system randomly assigned design examples to her using a multi-step process. First, the system randomly chose the number of design examples to assign (0, 1, 2, or 4 examples)[38]. If the designer were assigned to see one example, the system randomly selected an example from the pool of 24 stimulus ads. If she were assigned to see two or more examples, the system randomly selected design variability of the examples (low or high). If the assigned variability were low, the system randomly selected an example category, and then randomly chose ads from that category. If the assigned variability were high, the system randomly chose ads from different example categories, but at most one ad from each category.

We showed the assigned examples in the project brief (as is typically done in design contests), and indicated that they were designs that we liked. Designers could access the project brief and see the examples any time during the contest. We scaled the examples such that the maximum width or height was 100 pixels in the project brief. Designers could click on individual examples to view them at the original dimensions. *All the designers gave attention to the assigned examples.* We tracked designers' clicks on the examples, and found all but two designers clicked on every assigned example at least once during the contest. Of the two designers who did not click on all assigned examples, one of them clicked on one example (of two assigned), and the other clicked on two examples (of four assigned).

[38] Because examples in the high design variability conditions have to be relatively different from each other, the number of alternative categories limited the maximum number of examples we could show to designers. In addition, we did not include conditions for 3-high variability and 3-low variability examples. We only needed the extreme values (in this case, 2 and 4 examples, and high and low variability) to test linear interactions between quantity and design variability of examples (McClelland, 1997).

Table 5.1 shows the assignment of 98 designers to the various experimental conditions.[39] These designers submitted at least one design and completed the post-contest survey in our experiment. As a variable of interest is the design variability of examples, we use conditions with multiple design examples in our main analyses. Nevertheless, we would include conditions with no example (where there are no perceived quality and variability measures) and one example (where there is no perceived variability measure) in our robustness analyses in Section 5.7.4.

Table 5.1 Number of Designers in Experimental Conditions

Number of examples	0	1	2		4	
Variability of Examples	-	-	Low	High	Low	High
No. of Designers	25	22	16	11	12	12

5.6 Measures

The data in this study is hierarchical, where designs (Level 1) are nested within designers (Level 2). We describe the measures in the respective levels below.

5.6.1 Level 1 (Design-Level) Measures

Design Distinctiveness. The measure of distinctiveness for each design is based on our procedure in Chapter 4. Designs with higher distinctiveness scores are, on average, more different from other designs in the contest.

Design Deviation. We recruited raters on Amazon Mechanical Turk (MTurk) to compare the similarity between each design submission and the examples assigned to a particular designer during the contest. Three raters evaluated the similarity of each pair of submission and assigned example on a 7-point scale. We reverse-coded and averaged the raters' scores to obtain the dissimilarity of each submission-example pair. To determine the deviation of each design from the relevant assigned examples, we averaged the dissimilarity score of the corresponding submission-example pairs.

Design Divergence. During the contest, we allowed designers to submit multiple designs. Design divergence refers to the extent to which a design differs from other designs submitted by

[39] There were 99 designers in the sample in Chapter 4. However, one of the designers in the 4-example condition did not complete the post-contest survey, in which we measured a key explanatory variable (deviatory exploration). Hence we dropped this designer from our sample in this part of the study.

the same designer (Dow et al., 2010). We used the estimated pairwise distances from Chapter 4 to compute the design divergence for each design.

5.6.2 Level 2 (Designer-Level) Measures

Deviatory Exploration. In the post-contest survey, designers who saw at least one example indicated the extent to which they considered using designs that are (i) similar to assigned examples, and (ii) different from those examples ("not at all… to a very great extent"). The sum of these items indicates a designer's total exploration for various design concepts. We used the degree to which a designer considered designs that differ from assigned examples relative to his total exploration effort as a proxy of his deviatory exploration:

$$\frac{consideration\ for\ design\ concepts\ that\ differ\ from\ assigned\ examples}{total\ exploration\ for\ design\ concepts}$$

Designer-Level Design Deviation. We computed the person-level design deviation by averaging the design deviation of all submissions by the respective designers. This variable captures the between-designer effects of deviating designs from examples. By including design deviation at both design- and designer-levels, we can also examine the difference in effects between the two levels (Raudenbush and Bryk, 2002).

As we pointed out earlier, deviatory exploration and design deviation relate to designers' behaviors at different points in the creative process. The empirical relationship between these two measures in this study supports this assumption. The Cronbach's alpha for these two variables is .0003, which means the two variables do not reliably measures the same construct. The correlation between them is also very weak at -0.05 ($p > .10$).

4-Example Condition. We dummy-coded the number of examples that was assigned to designers. The value of this variable equals 1 for designers who saw four examples in the project brief, and equals 0 for those who saw two examples.

Quality of Examples. In the post-contest survey, we asked designers to rate the appropriateness and attractiveness of each assigned example on a 7-point scale (Cronbach's alpha = .89). We averaged all designers' ratings for each example to improve the accuracy of our quality measure. To determine the quality of examples for individual designers, we took the average ratings of the examples that were assigned to them.

Design Variability of Examples. Designers who saw multiple examples rated the similarity of those examples in the post-contest survey. They evaluated each pair of examples in terms of their overall similarity, layout, and images on a 7-point scale (extremely dissimilar to extremely similar) (Cronbach's alpha = .83). We reverse-coded and averaged the ratings to compute the dissimilarity between each pair of examples. To measure design variability of examples for individual designers, we averaged the ratings of relevant pairs of examples.

Contest Attitude. We asked designers the extent to which participating in design contests impact their (i) design experience, (ii) design skills, and (iii) design portfolio ("extremely negative impact… extremely positive impact"; Cronbach alpha = .94). Designers with positive attitude towards design contests might have exerted more efforts in creating designs and paid more attention to information in the project brief.

Design Experience. Because we recruited designers in the field, it is necessary to control for their different background in graphic designing. In the pre-contest survey, we asked designers to state their graphic design experience in years.

5.7 Results and Analyses

We analyzed our data using Stata (version 12.1). Given the multi-level structure of our data, we used Hierarchical Linear Modeling (HLM) with restricted maximum likelihood estimation method. HLM takes into account dependence among observations within designers, and provides efficient estimates in unbalanced, nested designs (Raudenbush and Bryk, 2002).

5.7.1 Drivers of Design Distinctiveness

First, we examined the impact of deviatory exploration and design deviation on design distinctiveness (Table 5.2). Model 1 is the baseline model that includes designers' attitude towards design contests, their design experience, and the design divergence of each design. We added a random slope effect for design divergence in the model. The likelihood-ratio (LR) test that compared this model with a model that only has a random intercept supports the inclusion of the random slope (LR = 25.64, p < .001). We find a positive impact of design divergence on design distinctiveness (β = .30, p < .01). Designers who varied the design concepts in their work submitted designs that were more distinctive in the contest. The estimated variance component for the intercept is 27.17, with a 95% confidence interval from 9.26 to 79.67. These estimates suggest that there are variations in design distinctiveness across designers.

Table 5.2 Impacts of Deviatory Exploration and Design Deviation on Design Distinctiveness

	Model 1	**Model 2**
DV: Design Distinctiveness	**Coef.**	**Coef.**
Constant	386.92^{**} (1.15)	360.62^{**} (11.00)
Contest Attitude	0.78 (1.14)	0.86 (1.09)
Design Experience	-0.17 (0.16)	-0.23 (0.15)
Design Divergence	0.30^{**} (0.11)	0.31^{**} (0.11)
Deviatory Exploration		11.49^{*} (5.76)
Design Deviation		1.32 (1.37)
Within-Subject Design Deviation		4.83^{*} (2.02)
Random Effect	**Var. Component**	**Var. Component**
var(Design Divergence)	0.208 (0.115)	0.203 (0.112)
var(Constant)	27.168 (14.913)	18.432 (12.791)
var(Residual)	129.54 (17.092)	129.604 (16.984)
Wald statistics	9.91^{*}	21.00^{**}
Log restricted-likelihood	-802.11	-791.56
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$		

In Model 2, we added deviatory exploration and design deviation (at the designer- and submission-level). A LR test that compared Models 1 and 2 is significant (LR = 21.10, p < .001), which supports the inclusion of these variables in our model. The change in the variance component of the intercept between the models indicates that deviatory exploration and design deviation explained 32% of the variation in design distinctiveness. We find that design distinctiveness is positively associated with deviatory exploration (β = 11.49, p < .05), supporting H1. The more designers considered concepts that differ from the assigned examples, the more distinctive their designs were in the contest. Design distinctiveness and design deviation at the designer-level are also positively related (β = 4.83, p < .05), supporting H2 Designs that deviated more from the assigned examples were also more distinctive.

5.7.2 Impact of Design Examples on Design Deviation

The results above indicate that the degree to which designers' work deviated from assigned examples affected designs distinctiveness. We followed up on this finding by examining how client-provided examples influenced design deviation (Table 5.3).

Table 5.3 Impact of Design Examples on Design Deviation

DV: Design Deviation	**Model 1**	**Model 2**	**Model 3**	**Model 4**
	Coef.	**Coef.**	**Coef.**	**Coef.**
Constant	5.42^{**} (0.08)	5.38^{**} (0.10)	5.38^{**} (0.10)	5.41^{**} (0.07)
Contest Attitude	0.06 (0.07)	0.18^{*} (0.08)	0.18^{*} (0.09)	0.18^{*} (0.08)
Design Experience	0.00 (0.01)	0. 00 (0.01)	0. 00 (0.01)	0. 00 (0.01)
4-Example		0.06 (0.15)	0.07 (0.16)	
Quality of Examples		-0.46^{*} (0.20)	-0.43^{*} (0.21)	-0.48^{*} (0.19)
Design Variability of Examples		-0.15^{*} (0.07)	-0.17^{*} (0.08)	-0.14^{*} (0.07)
4-Example X Design Variability of Examples			0.08 (0.15)	
Random Effect	**Var. Component**	**Var. Component**	**Var. Component**	**Var. Component**
var(Constant)	0.124 (0.062)	0.118 (0.056)	0.123 (0.058)	0.111 (0.054)
var(Residual)	0.531 (0.061)	0.516 (0.058)	0.515 (0.058)	0.517 (0.058)
Wald statistics	0.71	10.04^{+}	10.19	10.02*
Log restricted-likelihood	-242.56	-241.37	-242.24	-240.50
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$				

We included designers' attitude and experience in the baseline model (Model 1). The estimates for these variables and the Wald statistics are not significant at .10 levels. Next, we included the number, quality, and design variability of examples in Model 2. The Wald test is marginally significant ($W = 10.04$, $p < .10$), and the three variables explained 5% of the variation in design deviation. We find that design deviation is negatively affected by the examples' quality ($\beta = -.46$, $p < .05$) and design variability ($\beta = -.15$, $p < .05$). Designers deviated more from assigned examples when the examples' quality was low, and when the examples were relatively similar. These results support H4 and H7b, respectively. Furthermore, after we controlled for the example characteristics, we find that designers with more favorable attitude towards design contests engaged in greater deviation in designs ($\beta = .18$, $p < .05$).

In Model 3, we added the interactions between the number and design variability of examples. The estimate of this interaction is not significant ($\beta = .08$, $p > .10$). Hence H9 is not supported.

Since the number of examples does not have significant main and interaction effects, we excluded this variable in Model 4. The Wald statistics is now significant ($W = 10.02$, $p < .05$). As in Model 2, we find that design deviation is negatively impacted by the examples' quality ($\beta = -.48$, $p < .05$) and design variability ($\beta = -.14$, $p < .05$). In addition, the quality and design variability of examples accounted for 10% of variance in design deviation in Model 4. Thus, the number of examples does not improve our model sufficiently. This finding suggests that during the design production stage, the degree which designers deviated their submissions from assigned examples was independent of the number of examples.

5.7.3 Impact of Design Examples on Deviatory Exploration

Lastly, we looked at how design examples impacted designers' consideration for design concepts during the ideation stage. Because deviatory exploration is a designer-level characteristic, we used OLS regression for this analysis (Table 5.4). In Model 1, we included designers' attitude and experience, as well as the number, quality and design variability of examples. The F-test and example characteristics are all not significant at .05 levels.

Table 5.4 Impact of Design Examples on Deviatory Exploration

DV: Deviatory Exploration	**Model 1**	**Model 2**
	Coef.	**Coef.**
Constant	0.91^{**} (0.16)	0.89^{**} (0.14)
Contest Attitude	-0.07^{*} (0.03)	-0.06^{*} (0.03)
Design Experience	0.00 (0.00)	0.00 (0.00)
4-Example	0.00 (0.06)	-0.02 (0.05)
Quality of Examples	-0.05 (0.07)	0.05 (0.08)
Design Variability of Examples	0.04 (0.03)	0.04 (0.03)
4-Example X Quality of Examples		-0.55^{**} (0.14)
4-Example X Design Variability of Examples		-0.12^{*} (0.05)
F statistics	1.75	4.21^{***}
Adjusted R-squared	0.07	0.31
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$		

In Model 2, we added the interactions between (i) the number and quality of examples, and (ii) the number and design variability of examples. This model explained 31% of the variance, which is significantly higher than the 7% in Model 1. We find that the number and quality of examples jointly affect designers' exploration of the solution space ($\beta = -.55$, $p < .01$), supporting H5 (Figure 5.1). As the number of high quality examples increased, designers lessened their search for concepts that differ from these examples.

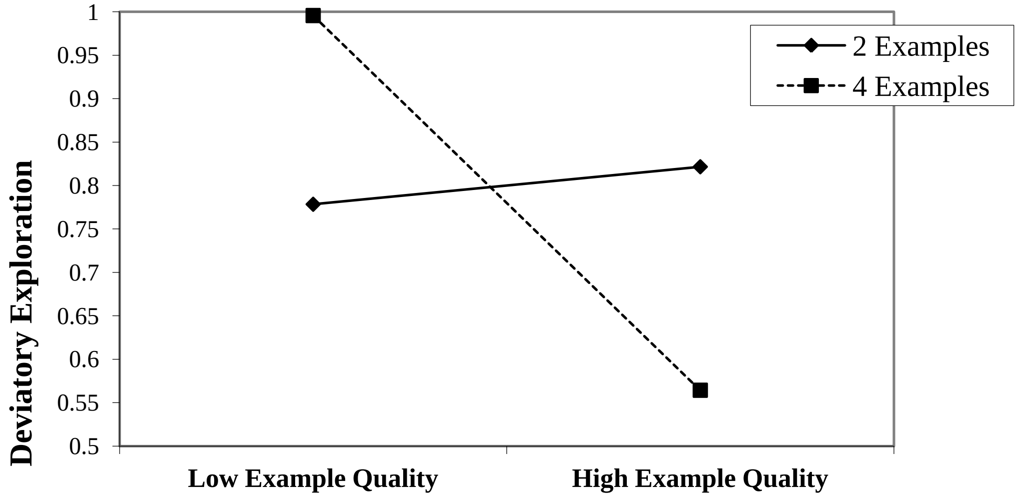


Figure 5.1 Interaction Between Number and Quality of Examples

The interaction between the number and design variability of examples is also significant (β = -.12, p < .05). The impact of design variability of examples on deviatory exploration is .04 when there were 2 examples, and .04-.12 = -.08 when there were 4 examples. The differences in the absolute values of these estimates imply that deviatory exploration was more sensitive to design variability of examples when there were more examples, supporting H8 (Figure 5.2).

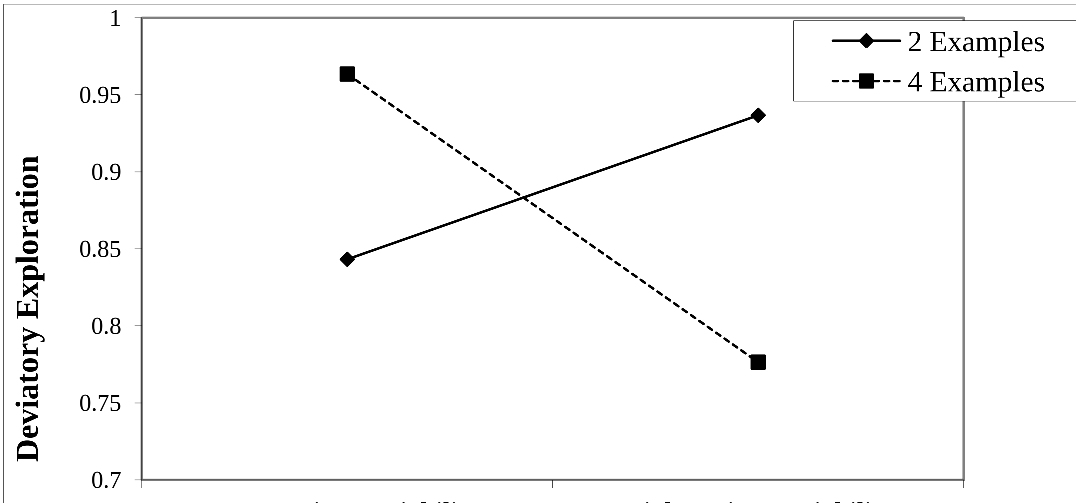


Figure 5.2 Interaction Between Number and Design Variability of Examples

Our results suggest that the number of examples affected the relative effects of design variability's roles as signals and creative benchmarks. With two examples, designers' deviatory exploration increased with design variability of the examples (β = .04). This is consistent with our hypothesis that design variability serves as signals of clients' preference (H6a). The more varied the client-provided examples, the weaker is the indication of the clients' preferences. Designers would therefore explore more widely in the design solution space.

However, with four examples, we see a different relationship between deviatory exploration and design variability. We find that deviatory exploration was higher when the client-provided examples were less varied (β = -.08). While signals of clients' preferences should still be relevant, the desire to submit unique work to differentiate from competitors appeared to have a stronger effect on designers' behavior. Hence designers engaged in greater exploration for concepts that differ from the assigned examples, consistent with H6b (creative benchmark effects). When we consider these results together with that in Section 5.7.2, where design deviation is higher when design variability of examples is low (β = -.14, $p < .05$), we see that designers placed a great emphasis on being creative. Even though winning the monetary prizes was important (as reflected in Designer A's comments), they were willing to risk their chances of winning by deviating their exploration and work from the assigned examples.

5.7.4 Robustness Analyses

Due to our interest in design variability of examples, we only used conditions with multiple assigned examples in our main analyses. Here, we used the full sample to check whether design distinctiveness varied systematically across the experimental conditions, including those with zero and one example. We created dummy variables for each of the conditions. Using HLM, we regressed design distinctiveness on different experimental conditions, designers' attitude, experience, and design divergence. We also included a random effect for design divergence in the model. We ran this model twice, rotating the base group between 0-example and 1-example conditions. None of the estimates for the experimental conditions are significant at .05 levels.

Using the full sample, we also examined whether the number of examples in each example category that a participant saw affected design distinctiveness. Although we randomly assigned examples in various categories to designers, it is possible that the different categories might have systemically affected designers' work. For example, some categories could be more inspirational than others, and examples in these categories might have influenced the creative process. To check this possibility, we regressed design distinctiveness on the numbers of examples in each category that were assigned to the respective designer. We also controlled for designers' attitude, experience, and design divergence in the HLM model. None of the estimates for example categories are statistically significant.

Finally, we excluded the 0-example condition from our sample to check the robustness of our results for example quality. In the main analyses, the variable for the number of examples variable was dummy-coded (equals 1 (0) if designer was assigned four (two) examples). Here, the variable, before being mean-centered, takes the value of 1, 2, and 4. We first estimated the impact of example quality on design deviation using HLM. Although the effect of example quality is negative as hypothesize, it is not significant ($\beta = -.17$, $p > .10$). Hence H4 is not supported here. This result differs from that in Section 5.7.2, where the negative effect of example quality is stronger and significant ($\beta = -.46$, $p < .05$).

We then looked at the effects of example quality on deviatory exploration. The results are shown in Table 5.5. In Model 1, we find that the quality of examples has a negative main effect on deviatory exploration ($\beta = -.09$, $p < .05$), consistent with H3. Next, we included an interaction term between the number and quality of examples in Model 2. Although the estimated sign is consistent with H5, the estimated effect is only weakly significant ($\beta = -.08$, $p < .10$). In contrast, the estimate of this interaction term in Section 5.7.3 is much stronger ($\beta = -.55$, $p < .01$).

Table 5.5: Impact of Design Examples on Deviatory Exploration; Including 1-Example Condition

DV: Deviatory Exploration	**Model 1**	**Model 2**
	Coef.	**Coef.**
Constant	0.70^{**} (0.11)	0.71^{**} (0.11)
Contest Attitude	-0.02 (0.02)	-0.02 (0.02)
Design Experience	0.00 (0.00)	0.00 (0.00)
Number of Examples	0.00 (0.02)	0.00 (0.02)
Quality of Examples	-0.09^{*} (0.04)	-0.15^{**} (0.05)
Number X Quality of Examples		-0.08^{+} (0.04)
F statistics	2.43^{+}	2.63^{*}
Adjusted R-squared	0.07	0.10
Standard errors in parentheses. $^{+}p < .10$ $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$		

We find that the estimated impacts of example quality on designers' behaviors in the robustness analyses are generally weaker than those in the main analyses. A possible explanation is that we did not include design variability in the model. (Designers in the 1-example condition did not rate on the similarity among the examples.) As design variability affected the behaviors of designers in the multiple-examples conditions, omitting this variable from our models could bias the results.

5.8 Discussion

Obtaining a diverse set of solutions is important in creative projects because it increase the likelihood of identifying an outstanding solution (Girotra et al., 2010; Terwiesch and Xu, 2008). To this end, by providing access to a large and diverse group of designers, crowd-based design contest platforms help advertisers to obtain a variety of ad designs. However, the information that advertisers provide in contests may shape designers' creative process and behavior in unexpected ways, and limit the range of diverse ideas from the crowd. In this study, we find that high quality client-assigned design examples caused designers to engage in less deviatory exploration and design deviation, which in turn reduce distinctiveness of their designs. Although advertisers may be happy with the quality of the resulting work, such designs with low distinctiveness may not perform well in advertising campaigns (as we showed in Chapter 4).

Does this mean advertisers should provide low quality examples during design contests? The short answer to this question is no. Instead, to encourage diversified exploration by designers, advertisers can be more strategic about the timing and types of high quality design examples that they show. We provide two recommendations about showing high quality examples in design contests.

First, if and when possible, advertisers should *delay showing high quality design examples in the contests*, instead of providing examples at the start of the contests as it is typically done. Showing design examples later allows designers to explore the solution spaces at the onset of the contests without them being fixated on those examples. This would encourage designers to explore more widely for potential design solutions. When the client-provided examples are subsequently shown, designers could then integrate various design concepts and perhaps develop more distinctive work.

Such two-phase approach is similar to the process in hybrid brainstorming teams, which has been found to produce favorable idea generation outcomes (e.g. Girotra et al., 2010). (One can view the client and individual designers in a design contest as a collaborative team that is working together to generate design solutions.) In a hybrid brainstorming team, individuals first work independently to generate ideas, and then they come together to share findings and perform additional exploration together. This process differs from that in a typical brainstorming team, where everyone works together to come up with ideas from the beginning. Girotra et al. (2010) found that hybrid teams generated more and better quality ideas than typical brainstorming teams do. More importantly for our context, they also found that ideas were more diverse in hybrid teams than in the typical brainstorming teams. This is because individuals in a typical brainstorming team are likely to build upon each other's ideas right from the start, which leads to the ideas being similar to one another. This phenomenon is similar to designers building upon and not deviating from client-provided (high quality) examples in our case. Therefore, giving designers time to engage in individual exploration at the beginning of the contests should help them generate more diverse design concepts.

Second, advertisers can *provide examples that are not highly related to the design problem domains*. In our experiment, we asked designers to create banner ads for an online wedding photography directory, and showed them design examples of such ads (see Appendix A). Because of the high relevancy of the examples to the project, designers could pattern their designs after these examples without much restriction (Ward, 1994), and reduced their search for and use of other design concepts. To overcome this, advertisers can consider using design examples from other domains. Because such examples may not be directly applicable in the specific design problems, they add to the constraints in the creative task. These constraints can cause designers to divert from the path of least resistance and engage in more creative cognitive processing (e.g., Moreau and Dahl, 2005). For example, designers have to find novel ways to frame and incorporate design concepts across different domains, which could result in more distinctive work.

Our results also show that by giving strong indications of their design preference (through providing highly similar examples), advertisers may end up getting designs that differ from their preference. However, this is not necessarily a bad outcome: greater deviation from client-provided examples could lead to more distinctive ad designs, which tend to perform better in online advertising campaigns (see Chapter 4).

Limitations. There are two main limitations in this study. First, while designers faced a tension between wanting to win the contest and wanting to win it in style (as reflected in our results in Section 5.7.3 and in Designer C's comment), we are not able to dissect this tension in our study. It would be useful to understand how designers manage and work through these different objectives in design contests. Such insights can help design contest platforms implement appropriate incentive schemes and contest structures.

Second, because designers participated in our study remotely, we could not observe their exploration of the design solution space.[40] We therefore relied on designers' responses in our post-contest survey to measure the degree of their deviatory exploration. Had we been able to see what they searched for during the contest, we could have measured the deviatory exploration construct more objectively. We could also have compared the total solution space covered by the various experimental conditions. This would allow us to examine how characteristics of client-provided examples affect the total exploration space at the crowd-level.

Notwithstanding these limitations, the experimental design and sample in this study add to the external validity of our results. We recruited experienced graphic designers to participate in our randomized experiment. Many of them also actively participated in real-world design contests. Furthermore, the monetary reward structure in our experiment was reflective of that in design contest platforms. In general, the designers agreed that the US$250 to US$600 prizes that we offered to winners were attractive (7-point scale; mean = 6.1, std. dev. = 1.2). Future studies can adopt a similar experimental setup to explore other phenomenon in design contests and creative design tasks.

[40] We did provide a Google Image search box in the project brief during the experiment. Designers could search for images using this search box, and we would record the keyword of the searches. However, this feature was under-utilized during our experiment. Only three designers used this feature to search for images, and the keywords were "wedding", "wedding memories", and "damask".

Conclusion

Design contest platforms help firms to source for various types of designs relatively easily and affordably. These two chapters shed some light on how firms should conduct the contests and select the right (ad) designs so as to get the most out of using these platforms. Yet, we only touched on some aspects of design contests. Going forward, there are other pertinent issues that need to be addressed. For example, how does the competitive dynamic among designers impact their creative processes in design contests? How do the various contest options (e.g., allowing designers to view each others' submissions during the contests) influence designers' behaviors and work? Insights into these questions can further improve the structures of these platforms, and help firms and designers better benefit from their participation in the contests.

References

Audia, P. G. and Goncalo, J. A. "Past Success and Creativity over Time: A Study of Inventors in the Hard Disk Drive Industry," *Management Science* (53:1), 2007, pp. 1-15.

Ball, L. J., Evans, J. S. T. and DENNIS, I. "Cognitive Processes in Engineering Design: a longitudinal study," *Ergonomics* (37:11), 1994, pp. 1753-1786.

Cross, N. "Expertise in Design", *in* Cross, N. and E, E., ed., Creativity and Cognition Press, University of Technology, Sydney, Australia, 2003.

Darke, J. "The primary generator and the design process," *Design Studies* (1:1), 1979, pp. 36-44.

Diao, F. and Sundar, S. S. "Orienting Response and Memory for Web Advertisements: Exploring Effects of Pop-Up Windowand Animation," *Communication Research* (31:5), 2004, pp. 537-567.

Dorst, K. and Cross, N. "Creativity in the design process: co-evolution of problem–solution," *Design Studies* (22:5), 2001, pp. 425-437.

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L. and Klemmer, S. R. "Parallel prototyping leads to better design results, more divergence, and increased self-efficacy," *ACM Trans. Comput.-Hum. Interact.* (17:4), 2010, pp. 18:1-18:24.

Fricke, G. "Successful individual approaches in engineering design," *Research in Engineering Design* (8), 1996, pp. 151-165.

Girotra, K., Terwiesch, C. and Ulrich, K. T. "Idea Generation and the Quality of the Best Idea," *Management Science* (56:4), 2010, pp. 591-605.

Heiser, R. S., Sierra, J. J. and Torres, I. M. "Creativity Via Cartoon Spokespeople in Print Ads," *Journal of Advertising* (37:4), 2008, pp. 75-84.

Herring, S. R., Chang, C.-C., Krantzler, J. and Bailey, B. P. "Getting inspired!: understanding how and why examples are used in creative design practice,"'Proceedings of the 27th international conference on Human factors in computing systems', ACM, New York, NY, USA, 2009, pp. 87-96.

Keller, K. L. "Memory and Evaluation Effects in Competitive Advertising Environments," *Journal of Consumer Research* (17:4), 1991, pp. 463-476.

Lawson, B. *Design In Mind*, Butterworth-Heinemann, Oxford, UK, 1994.

Li, H. and Bukovac, J. L. "Cognitive Impact of Banner Ad Characteristics: An Experimental Study," *Journalism & Mass Communication Quarterly* (76:2), 1999, pp. 341-353.

Lloyd, P. and Scott, P. "Discovering the design problem," *Design Studies* (15:2), 1994, pp. 125-140.

McClelland, G. H. "Optimal Design in Psychological Research," *Psychological Methods* (2:1), 1997, pp. 3-19.

Moreau, C. P. and Dahl, D. W. "Designing the Solution: The Impact of Constraints on Consumers' Creativity," *Journal of Consumer Research* (32:1), 2005, pp. 13-22.

Raudenbush, S. W. and Bryk, A. S. *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)*, Sage Publications, Inc, 2002.

Sundar, S. S. and Kim, J. "Interactivity and Persuasion: Influencing Attitudes with Information and Involvement," *Journal of Interactive Advertising* (5:2), 2005, pp. 5-18.

Terwiesch, C. and Xu, Y. "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science* (54:9), 2008, pp. 1529-1543.

Ullmana, D. G., Dietterich a, T. G. and Stauffera, L. A. "A Model of the Mechanical Design Process Based on Empirical Data," *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing* (2:1), 1988, pp. 33-52.

Ward, T. "Structured Imagination: the Role of Category Structure in Exemplar Generation," *Cognitive Psychology* (27:1), 1994, pp. 1-40.

Appendices

Appendix 5.A Experiment Stimulus

Example Category	Design Examples		
1. Collages			
			
2. Wedding bouquet as focal point			
			
3. Greenery background			

			
4. Top-and-bottom frames