WHAT DOES THE BRAIN TELL US ABOUT TRUST AND DISTRUST? EVIDENCE FROM A FUNCTIONAL NEUROIMAGING STUDY

By: Angelika Dimoka
Fox School of Business
Temple University
1801 Liacouras Walk
Philadelphia, PA 19122
U.S.A.
angelika@temple.edu

Abstract
Determining whom to trust and whom to distrust is a major decision in impersonal IT-enabled exchanges. Despite the potential role of both trust and distrust in impersonal exchanges, the information systems literature has primarily focused on trust, alas paying relatively little attention to distrust. Given the importance of studying both trust and distrust, this study aims to shed light on the nature, dimensionality, distinction, and relationship, and relative effects of trust and distrust on economic outcomes in the context of impersonal IT-enabled exchanges between buyers and sellers in online marketplaces.

This study uses functional neuroimaging (fMRI) tools to complement psychometric measures of trust and distrust by observing the location, timing, and level of brain activity that underlies trust and distrust and their underlying dimensions. The neural correlates of trust and distrust are identified when subjects interact with four experimentally manipulated seller profiles that differ on their level of trust and distrust. The results show that trust and distrust activate different brain areas and have different effects, helping explain why trust and distrust are distinct constructs associated with different neurological processes. Implications for the nature, distinction and relationship, dimensionality, and effects of trust and distrust are discussed.

Keywords: Trust, distrust, neuroIS, price premiums, functional neuroimaging, fMRI, cognitive neuroscience

Introduction
While information systems give rise to exchanges among people and organizations across the globe, these IT-enabled exchanges tend to be impersonal. Determining whom to trust and whom to distrust in such impersonal exchanges has been touted as an important element of modern society (Smith 2003), because both trust and distrust may have substantial effects on a person’s economic and social behavior.

While the importance and role of trust have been widely accepted by Information Systems scholars (e.g., Benbasat et al. 2008), there has been less work on distrust, despite the valid claim that wherever there is trust, there is a possibility of trust betrayal that may eventually lead to distrust (McKnight and Chervany 2000). The relative dearth of research on distrust might stem from the fact that earlier research has assumed trust and distrust to be the same construct that exists at opposite ends of a single continuum (Barber 1983; Deutsch 1958; Rotter 1980). The assumption that trust and distrust are opposite ends of a single continuum follows the classic view
that trust denotes cooperative conduct and distrust denotes opportunistic conduct (Arrow 1974). The simultaneous existence of trust and distrust has been viewed as unstable (Lewicki and Bunker 1995), drawing upon the view that trust and distrust are perfect substitutes that lie along the ends of a single continuum (Lewis and Weigert 1984). In contrast, recent theoretical work sees both trust and distrust coexisting simultaneously (Lewicki et al. 1998), while recent empirical work has shown that trust and distrust are distinct and only modestly correlated constructs (McKnight and Choudhury 2006). The distinction between trust and distrust is still an unresolved issue.

Besides, while there is much work on the nature and dimensionality of trust, with few exceptions (McKnight et al. 2003, 2004), there is little research on the nature and underlying dimensions of distrust. There is also little research on examining the relationship between trust and distrust and testing their relative effects on behavioral and economic outcomes (McKnight et al. 2003). Given the importance of studying both trust and distrust in IT-enabled impersonal exchanges, this study aims to shed light on the nature, distinction and relationship, dimensionality, and effects of both trust and distrust.

Virtually all empirical work on trust and distrust has been undertaken with psychometric measures (e.g., Benamati et al. 2006; Constantinople 1969; McKnight and Choudhury 2006; McKnight et al. 2003). While there is great value in psychometric measurement tools that have dramatically advanced the study of trust and distrust, an interesting question is whether the use of functional neuroimaging tools could complement the psychometric measurement of trust and distrust and offer additional findings.

Functional brain imaging has been proposed as a promising tool to measure brain activation directly and shed light on unanswered questions in economics, psychology, and marketing (e.g., Camerer 2003; Glimcher and Rustichini 2004; Lee et al. 2006; Zaltman 2003). This trend has also been extended to the IS literature (e.g., Dimoka and Davis 2008; Dimoka et al. 2010; Moore et al. 2005; Randolph et al. 2006). In this study, fMRI (functional Magnetic Resonance Imaging) is used to complement psychometric measures of trust and distrust by capturing the location, timing, and level of brain activity that underlies trust and distrust. fMRI captures the brain’s metabolic activity using the blood’s magnetic properties with superb spatial resolution. For a detailed explanation of fMRI experimental design, data analysis, and interpretation of fMRI results, please see Dimoka et al. (2007). This study also draws upon the cognitive neuroscience literature to link the brain activations of trust and distrust to existing neurological processes. Integrating the literature on trust and distrust in the social sciences with knowledge of the brain’s functionality from cognitive neuroscience with functional neuroimaging tools (fMRI), this study aims to answer the following research questions:

- **Where do trust and distrust reside in the brain, and what do their neural correlates reveal?**

Since there is a debate in the literature about the nature of trust and distrust and their reciprocal link, this study first aims to identify the neural correlates of these two constructs to shed light on their nature and relationship. For example, if trust and distrust activate the same areas of the brain, this may imply that they are merely opposite ends of a single continuum. Yet, if trust and distrust span different areas, this may imply that they are distinct constructs. Moreover, since the cognitive neuroscience literature has identified the functionality of many brain areas, identifying the neural correlates of trust and distrust may give valuable information about their nature, their underlying characteristics, and their relationship. Finally, since fMRI tools can potentially capture processes that are not easily obtained with self-reports, it may be possible to uncover brain areas associated with trust and distrust that have not been identified yet.

- **What do the neural correlates of trust and distrust tell us about their dimensionality?**

Since the dimensionality of trust is still a debated issue in the literature, and the dimensionality of distrust is still unclear, this study aims to examine the neural correlates of two key dimensions of trust (credibility and benevolence) and the corresponding dimensions of distrust (discredibility and malevolence). This examination may help determine whether both trust and distrust are multi-dimensional constructs, and whether their respective dimensions share similar or distinct brain areas. For example, even if the literature agrees that credibility and benevolence are two distinct trust dimensions (Ganesan 1994; Pavlou 2002), it is not clear whether the theoretical explanation and empirical testing with psychometric measures corresponds to how the brain actually activates in response to these two trust dimensions. Thus, the brain may offer useful knowledge about the dimensions of trust and distrust based on the brain areas they activate.

- **Can the neural correlates of trust and distrust predict economic outcomes?**
It is also important to examine whether the identified levels of brain activation associated with trust and distrust can predict economic outcomes. This would imply that the identified brain areas not only have a descriptive but also a predictive validity, plus test whether trust or distrust is the more predictive.

To shed light on these research questions, this study follows the recent guidelines for integrating neuroimaging and behavioral data (Huettel and Payne 2009; Yoon et al. 2009) to conduct two studies in the context of impersonal business-to-consumer (B2C) exchanges on eBay’s online auctions. In this context, where a buyer can transact with many sellers, it is possible for a buyer to simultaneously trust and distrust different sellers due to their feedback profile. The focus is on the buyer’s decision to trust or distrust a seller, and whether trust and distrust contribute to the buyer’s decision to offer a seller a price premium. The first study uses a classic behavioral experiment in which the subjects are given a set of four (2 × 2) feedback profiles of sellers who differ on their level of trust and distrust (high and low), and they are asked to report on a set of measurement items on trust, distrust, and price premiums about these sellers. The second study uses fMRI to capture brain activation when subjects in an fMRI scanner engage in the same activities as in the behavioral study. The fMRI study allows the simultaneous measurement of cerebral activation when subjects view the feedback profiles of these four sellers and respond to exactly the same psychometric measurement items. fMRI identifies and localizes the areas and levels of brain activation that correspond to trust and distrust and their proposed dimensions. The main purpose of these two studies is to assess the correspondence of the behavioral results across the fMRI and traditional studies to test the external validity of the fMRI study.

The results show that trust and distrust activate different brain areas, implying that they are distinct constructs that are associated with different neurological processes. Specifically, trust is associated with brain areas linked to anticipating rewards, predicting the behavior of others, and calculating uncertainty. Distrust is associated with brain areas linked to intense negative emotions and fear of loss. There results also show a clear distinction in the brain areas associated with the dimensions of trust and distrust with credibility and discredibility being mostly associated with the brain’s more cognitive areas, while benevolence and malevolence are mostly associated with the brain’s more emotional areas.

This paper contributes to the literature in three major ways: First, it identifies the neural correlates of trust and distrust and sheds light on their nature, distinction, and relationship. Second, it sheds light on the dimensionality of trust and distrust by identifying distinct neural correlates for their dimensions. Third, the brain activations associated with trust and distrust better predict economic outcomes (price premiums) than the corresponding self-reported psychometric measures of trust and distrust.

Literature Review

Trust

Trust is defined as a person’s (the trustor) willingness to be vulnerable to another person (the trustee) on the basis that the trustee will act according to the trustor’s confident expectations (Mayer et al. 1995). In buyer–seller relationships, trust is defined as the buyer’s willingness to be vulnerable to a seller based on the belief that the seller will transact in a manner consistent with the buyer’s confident expectations (Pavlou and Gefen 2004). While there are many ways to conceptualize the dimensionality of trust, a well-accepted view in buyer–seller relationships is a two-dimensional view with credibility and benevolence as the dimensions (e.g., Ba and Pavlou 2002; Doney and Cannon 1997; Singh and Sirdeeshmukh 2000). Credibility refers to the seller’s competence, honesty, and reliability (Sirdeshmukh et al. 2002); benevolence refers to a seller’s genuine interest in the buyer’s welfare (Garbarino and Lee 2003). Following Pavlou and Dimoka (2006), credibility is defined as the buyer’s belief that a seller is competent, honest, and reliable, and will fulfill contractual requirements. Benevolence is defined as the buyer’s belief that a seller has altruistic motives, is genuinely concerned with the buyer, and will act in a goodwill manner. The literature has shown credibility and benevolence to be distinct constructs with different relationships with other constructs (e.g., Ganesan 1994; Pavlou 2002; Pavlou and Dimoka 2006).

Distrust

While trust is viewed as an expectation of a partner’s beneficial conduct, distrust has been viewed in reciprocal terms as

2 Other views include a unidimensional view of trust, three dimensions (competence, honesty/integrity, benevolence) (Gefen 2002), and four dimensions (including predictability) (McKnight et al. 2002). In buyer–seller relationships, competence, honesty, and reliability collapse under the umbrella of credibility (e.g., Doney and Cannon 1997). While credibility often also includes integrity, since integrity is similar to benevolence, for clarity, integrity is omitted from the dimension of credibility, consistent with Barber (1983), Nooteboom (1996), and Pavlou and Dimoka (2006).
an “expectation of injurious action” (Luhmann 1979, p. 72) that the trustee will not act in the trustor’s best interests (Barber 1983). Distrust reflects the trustor’s expectation about the trustee’s poor capabilities, negative motives, and harmful behavior (Deutsch 1958; Ullmann-Margalit 2004), and it has been viewed as lack of confidence, fear of harm, harmful and hostile intentions, and lack of care about the trustor’s welfare (Grovier 1994). Thus, distrust has been negatively linked to effective social exchanges (Blau 1964), loyalty (Lewicki et al. 1998), communication (Arnulf et al. 2005), cooperative behavior (Vlaar et al. 2007), information sharing (Gillespie and Dietz 2009), responsiveness (Williams 2007), keeping promises (Robinson 1996), meeting obligations (Sitkin and Roth 1993), and online transactions (McKnight and Chervany 2000). Integrating the literature on distrust (Kramer 1999; Lewicki et al. 1998; Sitkin and Roth 1993), distrust is herein defined as the buyer’s unwillingness to be vulnerable to a seller on the basis that the seller will be inept, exhibit reckless behavior, violate obligations, not care about the buyer’s welfare, act against the buyer’s interests, and even intend to harm the buyer.

Distrust engenders when expectations of the trustee’s competence and benevolence are violated (McKnight and Chervany 2000). This is because distrust originates from either breaches due to technical incompetence (termed reliability-related distrust) or social violations (termed value-oriented distrust) (Hsia 2002). Thus, similar to the proposed dimensions of trust (credibility and benevolence), distrust is also proposed here as a two-dimensional construct. One proposed dimension—discredibility—deals with the trustor’s concerns about the trustee’s competence, honesty, and reliability, and the second—malevolence—relates to concerns about the trustee’s commitment to the trustor’s welfare. These two dimensions of distrust draw upon McKnight and Chervany (2000), who proposed three analogous dimensions of distrust (incompetence, dishonesty, and benevolence). However, for parsimony, their two distrust beliefs (incompetence and dishonesty) are integrated under the umbrella of discredibility, similar to how credibility includes competence, honesty, and reliability (Gefen 2002; Pavlou and Dimoka 2006). Malevolence draws directly upon McKnight and Chervany to refer to the trustor’s fear that the trustee will not act in her best interests.

**Trust Versus Distrust**

Whether trust and distrust are distinct constructs or the opposites of a trust–distrust continuum has been a debated topic in the literature. Some authors equated distrust with lack of trust (Ziegler and Lausen 2005). Rotter (1980) and Worchel (1979) contended that trust and distrust lie along the two ends of a continuum. Similarly, Omodei and McLennan (2000) proposed that trust and distrust are two ends of the same scale. Gans et al. (2001) also viewed distrust as merely the other side of trust, arguing for a symmetric scale with complete trust on one end and absolute distrust on the other. Luhmann (1979) posited that while trust and distrust are essentially the same construct, they are distinct functional equivalents that act in opposite ways.

Kahneman and Tversky (1979), who called for a distinction between positive and negative valence constructs, laid the groundwork for theorizing trust and distrust as distinct constructs. This view is consistent with that of Cho (2006, p. 26), who viewed distrust as “qualitatively distinct phenomena from trust: Distrust is not just the absence of trust.” This suggests that distrust does not mean lack of trust, but the expectation that the trustee will act in a way that violates the trustor’s best interests. While both trust and distrust deal with subjective expectations, trust deals with positive expectations about the trustee’s beneficial conduct, and distrust deals with negative expectations about the trustee’s harmful conduct. Deutsch (1973) called suspicion “distrust” and made a clear distinction between trust and distrust, a view supported by Kramer and Cook (2004). Some earlier empirical work had shown trust and distrust to be distinct (Constantinople 1969), and recent work has viewed distrust as a unique construct that is related to, yet it is distinct from, trust. For example, McKnight and Choudhury (2006) showed trust and distrust in Internet legal advice providers to be distinct, and Xiao and Benbasat (2003) and Lee and Huynh (2005) empirically reinforced this distinction. Thus, there is increasing empirical evidence to challenge the initial view that trust and distrust are the opposite ends of a single continuum.

In terms of whether trust and distrust have different relationships with their antecedents and effects, Lewicki et al. (1998) theorized that trust and distrust should have different relationships with other variables. Deutsch and Krauss (1962) theorized that distrust-based and trust-based relations have distinct antecedents. Sitkin and Roth (1993) showed empirically that trust and distrust operate differently and have different effects. Lee and Huynh (2005) validated that trust and distrust have different effects in IT outsourcing relationships. McEvily et al. (2006) showed different effects depending on whether trust was honored or violated (distrust). McKnight et al. (2004) illustrated that trust and distrust dispositions to have distinct effects. In e-commerce, Benamati et al. (2008) showed trust and distrust to have dissimilar effects on the intention to use online banks, and in a study of online sellers, Cho (2006) showed trust and distrust to have different antecedents. In sum, while there is ample support that trust and distrust often have different antecedents and effects, there is still an unresolved issue on whether they are distinct constructs or if they are the opposite ends of a single continuum.
Theory Development

The field of cognitive neuroscience focuses on the interaction between high-order mental beliefs, thoughts, and behaviors with their corresponding brain areas, termed neural correlates (Camerer 2003). While virtually all human processes are controlled by the brain, cognitive neuroscience is interested in cognitive, emotional, and social processes (Glimcher and Rustichini 2004). The field has created maps of these processes and linked them to associated brain areas (neural correlates). There are two major brain systems: the prefrontal cortex and the limbic system. The major areas of the prefrontal cortex are the dorsolateral (upper outer), ventromedial (lower middle), and orbitofrontal (above the eyes) cortices, and the limbic system consists of the brain’s interior areas (e.g., amygdala, caudate nucleus, putamen, and insular cortex). The prefrontal cortex has commonly been associated with cognitive processes (Ernst and Paulus 2005), while the limbic system has generally been linked to emotional and social processes (Sharot et al. 2004).

This article proposes a set of explicit hypotheses that link trust and distrust with specific brain areas. This follows the recent call for ex ante developing neuro-anatomical brain models by specifying the expected neural correlates (e.g., Hedgcock and Rao 2009; Huettel and Payne 2009; Yoon et al. 2009). However, given that use of fMRI in the study of trust and distrust is still in its infancy, besides a modest set of basic hypotheses about the neural correlates of trust and distrust and their dimensions, an exploratory, data-driven approach is followed here to allow the brain data to “speak for themselves.” For example, there is a dearth of research on the relationship between trust and distrust in the cognitive neuroscience literature to allow us to ex ante develop hypotheses about the relationship and relative effects of their neural correlates.

Neural Correlates of Trust

Following its definition, trust has three major attributes: First, confident expectations involves the anticipation of positive rewards. The cognitive neuroscience literature has identified the caudate nucleus as a key area associated with subjects anticipating positive rewards. The caudate nucleus is innervated by dopamine neurons that are activated when one receives a reward. The literature has also shown that the caudate nucleus is associated with the magnitude of an expected reward (e.g., Knutson et al. 2001). Also, King-Casas et al. (2005) showed caudate nucleus activation when subjects acted cooperatively to obtain an anticipated reward; Rilling et al. (2008) showed the caudate nucleus to predict the decision to reciprocate cooperation for mutual gain. Therefore, we expect trust to be associated with caudate nucleus activation.

Second, on the basis that the trustee will act according to the trustor’s confident expectations, trust involves predicting how the trustee will perform in the future. The cognitive neuroscience literature has identified the anterior paracingulate cortex as the key area activated when predicting how others will act (McCabe et al. 2001), an area in the limbic system responsible for calculating uncertainty (Rilling et al. 2004). Also, Krueger et al. (2007) showed the anterior paracingulate cortex to be activated when inferring another person’s intentions. Therefore, we expect trust to be associated with the anterior paracingulate cortex.

Third, the willingness to be vulnerable clause involves uncertainty about the potential effects of trust (Pavlou et al. 2007). A meta-analysis of 27 studies of uncertain decision-making showed the orbitofrontal cortex to be the primary area for calculating uncertainty (Krain et al. 2006). Huettel et al. (2005) showed that the activation in the orbitofrontal cortex increases in magnitude with higher levels of uncertainty, while Hsu et al. (2005) and Gonzalez et al. (2005) showed that the orbitofrontal cortex can distinguish among uncertainty levels. Hence, we expect a higher degree of trust to be associated with a lower activation in the orbitofrontal cortex.

Integrating these three attributes of trust, the following neural correlates of trust are proposed:

H1a: Trust is associated with higher brain activation in the caudate nucleus and the anterior paracingulate cortex, and lower brain activation in the orbitofrontal cortex.

Dimensions of Trust

The literature has viewed credibility to be associated with economic rationale (Williamson 1985) based on cognitive assessment of contracts, laws, and structural assurances, while benevolence is associated with caring intentions based on emotional assessment of a person’s goodwill (Barber 1983). Therefore, we expect credibility to be primarily associated with the prefrontal cortex, and benevolence to be primarily associated with the limbic system. Specifically, calculative decision making that activates the orbitofrontal cortex should be associated with assessment of credibility. Assessing the trustee’s characteristics (Deutch 1958) and motives (Fein and Hilton 1994) to predict the level of uncertainty are cognitive aspects of trust, which are associated with the orbitofrontal cortex. Also, Krueger et al. (2007) showed conditional (similar to calculative) trust to be associated with the calcu-
lation of rewards (associated with the caudate nucleus), whereas unconditional or goodwill trust (similar to benevolence) activates the septal area (which includes the anterior paracingulate cortex), which is associated with social attachment behavior. Accordingly, McCabe et al. (2001) and Walter et al. (2004) linked the anterior paracingulate cortex with predicting future social interactions. Thus, we propose

H1b: Credibility is associated with higher activation in the caudate nucleus and lower activation in the orbitofrontal cortex, while benevolence is associated with higher activation in the anterior paracingulate cortex.

Neural Correlates of Distrust

Building upon its definition, distrust has two major characteristics: First, it involves feelings of worry due to the fear of loss caused by the trustee’s harmful behavior. The insular cortex, a brain area associated with the fear of loss (Wicker et al. 2003) and risk prediction errors (Preuschoff et al. 2008), has been activated in studies of judgment of untrustworthy faces (e.g., Todorov 2008; Winston et al. 2002). Activation in the insular cortex helps people avoid negative interactions in the future (Rilling et al. 2008), which is consistent with the logic of distrust that helps shield people from harmful actions.

Second, distrust involves intense negative emotions of wariness, caution, defensiveness, vigilance, anger, hate, and betrayal (e.g., Lewicki et al. 1998). Kramer (1999) described distrust as having a paranoid character based on fear. The cognitive neuroscience literature has identified the amygdala as an area linked to intense negative emotions (LeDoux 2003). When rating faces as untrustworthy, Winston et al. (2002), and Todorov (2008) showed bilateral (both sides of the brain) amygdala activation because of intense harm avoidance. This is consistent with the conceptualization of distrust that involves readiness for danger, anticipation of harm, prevention, and vigilance against opportunism. In a lesion study, Adolphs et al. (1998) found impaired judgment of untrustworthiness in subjects with amygdala injury. Thus, we propose

H2a: Distrust will be associated with higher activation in the (1) insular cortex and (2) amygdala.

Dimensions of Distrust

In terms of the dimensions of distrust (discredibility and malevolence), discredibility deals with the trustor’s assessment of the potential of loss due to the trustee’s incompe-
tence, dishonesty, and unreliability. In contrast, malevolence is a highly emotional assessment that deals with concerns that the trustee will engage in a harmful behavior. While both the amygdala and the insular cortex are key sensory brain areas, the insular cortex is associated with the fear of loss, while the amygdala is associated with intense emotions. Since discredibility assesses the potential for loss due to poor seller characteristics, it will likely activate the insular cortex. On the other hand, malevolence is more likely to activate the amygdala as a result of intense emotions when predicting that the trustee will engage in harmful behavior. Therefore, we propose

H2b: Discredibility is associated with higher activation in the insular cortex; malevolence is associated with higher activation in the amygdala.

Trust Versus Distrust

Integrating H1 and H2, trust and distrust are distinct constructs that span different areas of the brain. This draws upon the theoretical foundations of Luhmann, that the articulation of trust and distrust as distinct, albeit related constructs. The primary reason that trust and distrust are distinct constructs is that distrust usually involves a strong emotional component while trust does not (Luhmann 1979, pp. 71-72). McKnight and Chervany (2000) explain that distrust reflects the emotion-charged human survival instinct, while trust is more calm and collected. For example, a seller’s violation of a buyer’s psychological contract (Robinson 1996)—a case of distrust—causes an emotional reaction to abruptly abandon the marketplace (Pavlou and Gefen 2005). While trust and distrust are formed by common processes (information sharing and assessment of one’s characteristics), Komiak et al. (2005) showed an asymmetry in the nature of trust and distrust because different processes contribute to trust versus distrust building. Also, McKnight et al. (2004) and McKnight and Choudhury (2006) offer empirical evidence showing that the antecedents and consequences of trust and distrust are distinct.

While both trust and distrust have emotional elements, trust is proposed to have emotions associated with hope and positive rewards (caudate nucleus) and cooperative intentions (anterior paracingulate cortex), while distrust is associated with the fear of loss (insular cortex) and intense negative emotions of worry, suspicion, and fear (amygdala). The proposed neural correlates of trust and distrust are consistent with McKnight and Chervany and with Sitkin and Roth (1993), who viewed trust as more of a cognitive belief and distrust as a more emotion-laden belief. Therefore, we propose
H3: Trust and distrust are distinct constructs that span different brain areas.

Effects of Trust and Distrust on Price Premiums

The effects of a buyer’s trust and distrust in a seller are proposed to be reflected on the price premium (or discount) the buyer is willing to give to the seller. A price premium is defined as the monetary amount above the average price received by multiple sellers that offer the same product (Ba and Pavlou 2002). Following Pavlou and Dimoka (2006), we expect the two dimensions of trust (credibility and benevolence) to have a positive effect on price premiums since buyers are willing to give higher prices to credible sellers who are likely to competently deliver the product and benevolently refrain from acting opportunistically. A buyer’s distrust in a seller is also proposed to have a negative effect on price premiums. Distrust makes buyers worry that the seller will engage in opportunistic conduct. Thus, buyers are likely to take defensive actions to shield themselves against these negative expectations (Luhmann 1979), thus giving a discount to sellers to compensate for the additional fear and uncertainty. Thus, we expect both dimensions of distrust (discredibility and malevolence) to negatively affect the buyer’s price premium to a seller.

In terms of the effects of trust and distrust on price premiums, distrust is expected to have a stronger effect than trust. Negative beliefs tend to be more powerful, credible, and weigh more on a decision than positive beliefs (Kahneman and Tversky 1979). Also, Taylor (1991) argued that distrust is more strongly coded in memory since it reflects a deviation from social norms, and Hardin (1993) noted that distrust can lead to dramatic effects on decisions compared to reduction in trust. McKnight et al. (2003) argue that the level of risk in the situation affects the relative effects of trust and distrust in the sense that distrust may embody higher levels of risk. Slovic (1993) also noted the asymmetry in trust and distrust, explaining that trust-building events are less visible while trust-destroying (distrust) events carry more weight. Therefore,

H4: Distrust is likely to have a greater effect on price premiums than trust.

The purpose of H4 is to examine whether the corresponding brain areas for trust and distrust can predict the level of price premium or discount that the buyer is willing to offer the seller, and particularly which brain area will be more influential. H4 also aims at examining the relative effect of the levels of brain activations associated with the dimensions of trust and distrust on price premiums. In doing so, we will explore whether the identified brain areas for trust and distrust and their dimensions can better predict price premiums compared to the corresponding psychometric measures of trust and distrust.

As shown in Figure 1, which integrates the proposed theoretical framework and research hypotheses, the neural correlates of trust and distrust are expected to precede their corresponding psychometric measures. This is because the measurement of brain activity when subjects read the measurement items of trust and distrust temporally precedes when the subjects are instructed to respond to the psychometric measurement items, as explained in the next section.

Research Methodology and Results

Two studies were conducted with the same stimuli and experimental design. The first study was a classic behavioral lab experiment with 177 students of a major U.S. university. The second study was an fMRI study in which the brains of 15 subjects were scanned while performing the same experimental tasks in an fMRI scanner. The measure operationalization and the two studies are described in detail below.

Measure Operationalization

Seller Profiles: Four seller profiles were created to reflect the $2 \times 2$ high and low trust and distrust matrix (Figure 1). The stimuli for trust and distrust are seller feedback profiles manipulated to create substantial variation in the level of trust and distrust. To incorporate the dimensions of credibility/benevolence and discredibility/malevolence, the four seller profiles were created by manipulating each seller’s feedback text comments. According to Pavlou and Dimoka (2006), feedback text comments offer information about a seller’s past transactions, thus helping buyers build trust in a seller’s credibility and benevolence, or distrust due to the seller’s discredibility and malevolence. The seller profiles were created by manipulating the number of extraordinary text comments that were shown to build trust in sellers by spawning an element of surprise. Following Pavlou and Dimoka, a set of feedback text comments that convey evidence of extraordinary (outstanding or abysmal) past seller behaviors of credibility and benevolence were identified. Outstanding credibility comments denote evidence of a seller excelling in transaction fulfillment, such as exceptional delivery and faithful adherence to contractual promises. Outstanding benevolence comments give evidence of a seller’s extraordinary goodwill behavior,
such as empathy, genuine interest in the buyer’s interest, and going beyond the call of duty. Abysmal discredibility comments give evidence of major incompetence, dishonesty, and unreliability, such as extreme delay or unusual shipping problems. Abysmal malevolence comments give evidence of a seller’s intentional past opportunism, such as fraud. Finally, most feedback text comments are ordinary and do not offer evidence of extraordinary behavior. These five categories of text comments were manipulated to create the four seller profiles to resemble real-life eBay sellers (Appendix A) by including feedback text comments and numerical ratings based on Ba and Pavlou (2002) and Pavlou and Dimoka (2006).

Since each of the two dimensions of trust and distrust were manipulated to “move together,” three four seller profiles were created following Lewicki et al.’s (1998) $2 \times 2$ (high versus low trust/distrust) matrix (Figure 1). First, the low trust/low distrust (LL) seller does not contain any extraordinary text comments. Second, the high trust/low distrust (HL) seller contains both outstanding credibility and benevolence comments. Third, the low trust/high distrust (LH) seller contains only discredibility and malevolence comments. Fourth, the high trust/high distrust (HH) seller contains both outstanding credibility and benevolence comments and also outstanding discredibility and malevolence text comments.

McKnight et al. (2004, pp. 39-40) illustrate these four seller profiles with the aid of structural assurances:

1. for the LL seller, the buyer perceives some fear about the seller because of weak structural assurances
2. for the HL seller, the buyer feels little or no fear due to the presence of effective structural assurances
3. for the LH seller, the buyer fears or worries about the seller due to the lack of structural assurances
4. for the HH seller, while the buyer feels fear about the seller, structural assurances reduce these fears

The manipulation between high and low trust and distrust was one standard deviation, above and below, respectively, the average number of text comments and percentage of numerical ratings. Table 1 shows the number of outstanding and abysmal credibility and benevolence text comments in each seller’s profile of the last 25 text comments (first webpage), and total (lifetime) number of positive and negative ratings.

The exact wording of the text comments was obtained from actual quotes of eBay buyers, as reported by Pavlou and Dimoka. After selecting the number of text comments under each category (Table 1), the text comments were randomly

---

Footnotes:
3 High trust sellers were high on both credibility and benevolence, and high distrust sellers were high on discredibility and malevolence. There were no seller profiles with high benevolence and low credibility, high malevolence and low discredibility, or any profile combinations with high and low benevolence/malevolence and credibility/discredibility.

4 The existence of a seller that is both high trust and high distrust (HH) is often questioned. However, Lewicki et al. (1998) and Benamati et al. (2008) explain that trust and distrust can operate together. This is because buyers often get conflicting information about the seller (such as from feedback text comments), thus generating both high trust and high distrust.
because the ultimate purpose of the seller profiles and measurement items was to serve as the stimuli in the fMRI study to induce activation in the brain’s trust and distrust areas, special care was taken to ensure that there were controls for the complexity and length of the measurement items for trust and distrust (Appendix B). This is necessary because differences in the stimuli (such as extra visual stimuli due to longer and more complex measurement items) can cause spurious activations in the brain that are unrelated to the focal stimulus, such as the visual cortex (occipital lobe). Such spurious activations are problematic because they may raise the overall level of brain activity and may suppress statistically significant “true” activations in focal brain areas. Thus, fMRI experiments must design appropriate controls to cancel out or “subtract” all other brain activations besides the desired activation due to the focal stimulus (trust and distrust). The complexity and length of the measurement items were then accounted for by adapting them to be similar across sellers, and by creating control measurement items that have the same complexity and length (Appendix B). The brain activation when the subjects responded to the control measurement items was subtracted from the activation when the subjects responded to the measurement items for trust and distrust and their dimensions.

Behavioral Study

The behavioral study was conducted for three reasons. First, to calibrate the seller manipulations and other experimental procedures for the fMRI study. Second, to collect behavioral data from a classic lab experiment and compare them with the corresponding behavioral data from the subsequent fMRI study. Thus, the behavioral study was used to overcome concerns that the setting and artificiality of the fMRI environment could bias the behavioral data. A similarity between the fMRI and the behavioral data can alleviate concerns for lack of external or ecological validity. Third, the behavioral data were used to refine the stimuli and tasks of the fMRI study, such as allowing subjects sufficient time to read and process the items.
the seller profiles before the fMRI study and the measurement items presented to them during the study. However, no problems were identified in the pilot study, and the behavioral and fMRI tasks were identical.

Behavioral Design and Manipulations

The context of the study was eBay’s auction marketplace, and the subjects were given the task to buy an MP3 player (iPod Nano) from four made-up electronics sellers whose profiles differed on trust and distrust by manipulating their feedback text comments and ratings, as shown in Figure 1 and prescribed in Table 1. After receiving the instructions to purchase an MP3 player from four manipulated eBay sellers, the subjects were asked to browse the four seller profiles as potential candidates from whom to buy the iPod Nano. Then, the subjects responded to the study’s measurement items (Appendix B) for the four seller profiles and standard demographics in a within-subject design on an online survey form (hosted by SurveyMoney, www.surveymonkey.com).

Behavioral Data Analysis and Results

Descriptive statistics are shown in Table 2. The manipulations of the four seller profiles were successful because the two low distrust sellers had significantly lower levels of distrust and higher levels of trust (p < .001) than the high distrust sellers. Also, the HL seller had higher levels of trust (p < .01) than the LL seller. The HH seller had higher levels of trust than the LH seller and lower levels of distrust (p < .01) than the LL sellers (Table 2).

Because the behavioral study did not capture brain activity, the analysis only focused on whether trust and distrust and their dimensions are distinct (H3), and whether distrust has a stronger effect than trust (H4).

To test H3, multiple tests for discriminant validity were run. For exploratory factor analysis (EFA), the principal components factor analysis in SPSS with promax rotation was used; for confirmatory factor analysis (CFA), the CFA procedure in PLS Graph and LISREL (Gefen and Straub 2005) was used. In terms of EFA with eigenvalue greater than 1, the analysis only revealed one construct, failing to distinguish between trust and distrust and their dimensions. When the number of extracted variables was forced to two or four (Appendix C), it was possible to see some distinction between trust and distrust or among the four dimensions of trust and distrust, respectively, yet without clear separation (following Gefen and Straub’s guidelines). Other EFA and CFA methods and rotations were also tried with similar results. Since these discriminant validity tests did not clearly distinguish between trust and distrust, H3 is not supported by the behavioral data.

To test the relative effect of trust and distrust and their two respective dimensions on price premiums (H4), a regression model was run with the four dimensions being used to predict price premiums (Table 3).

As shown in Table 3, given the lack of discriminant validity among the predictor variables and the low tolerance values (less than .70), the two dimensions of trust and distrust modestly predict price premiums (R² = 31%). Still, both dimensions of distrust (discredibility and malevolence) have stronger effects and explain a higher variance in price premiums (20%) than the corresponding two dimensions of trust (11%), thus supporting H4.

fMRI Study

The fMRI study focused on capturing cerebral activation when subjects were undertaking the same tasks as in the behavioral experiment while their brains were being scanned. Fifteen right-handed subjects (9 males and 6 females) who were pre-screened for fMRI safety (no medical implants, no metal piercings, no physiological problems) participated in the fMRI study for $35 compensation. The number of subjects was chosen to ensure adequate power of analysis for obtaining statistically significant brain activations (Desmond and Glover 2002). For a threshold of p < .05 with 80 percent power of analysis, n = 12 is needed, and 15 subjects exceeded the threshold. The subjects were recruited from the population of the metropolitan area of a major U.S. university using an open flyer for an fMRI study. The fMRI protocol and ad were reviewed and approved by the university’s institutional review board. The technical details associated with the fMRI scanner, experimental procedures, and detailed data analysis procedures are reported in Appendix E.

fMRI Design and Manipulations

The design used the four (2 × 2) seller profiles in a within-subject design. Before the fMRI session, similar to the behavioral study, the subjects were asked to buy an iPod Nano

While a between-subject design might be preferred, it would require four times as many subjects (about 60 subjects) and would be impossible to conduct given the cost and time constraints of fMRI studies.
Table 2. Descriptive Statistics of Dimensions of Trust and Distrust for Behavioral Study

<table>
<thead>
<tr>
<th></th>
<th>Credibility</th>
<th>Benevolence</th>
<th>Discredibility</th>
<th>Malevolence</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seller (LL)</td>
<td>5.65 (1.18)</td>
<td>5.42 (1.24)</td>
<td>2.23 (1.28)</td>
<td>2.27 (1.31)</td>
<td>$93</td>
</tr>
<tr>
<td>Seller (HL)</td>
<td>6.34 (1.15)</td>
<td>6.12 (1.23)</td>
<td>1.86 (1.28)</td>
<td>1.74 (1.14)</td>
<td>$98</td>
</tr>
<tr>
<td>Seller (LH)</td>
<td>2.94 (1.23)</td>
<td>2.66 (1.27)</td>
<td>5.23 (1.23)</td>
<td>5.05 (1.24)</td>
<td>$77</td>
</tr>
<tr>
<td>Seller (HH)</td>
<td>3.28 (1.29)</td>
<td>3.26 (1.28)</td>
<td>4.98 (1.24)</td>
<td>4.72 (1.34)</td>
<td>$84</td>
</tr>
<tr>
<td>All Sellers</td>
<td>4.62 (1.84)</td>
<td>4.41 (1.82)</td>
<td>3.68 (1.91)</td>
<td>3.51 (1.90)</td>
<td>$88</td>
</tr>
<tr>
<td>Reliability</td>
<td>0.93</td>
<td>0.92</td>
<td>0.91</td>
<td>0.95</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 3. Regression Results for Predicting Price Premiums with Trust and Distrust (Behavioral Data)

<table>
<thead>
<tr>
<th>Psychometric Measures</th>
<th>Standardized Beta Coefficient</th>
<th>ΔR²</th>
<th>Tolerance</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credibility</td>
<td>.10*</td>
<td>.04</td>
<td>.42</td>
<td>.31</td>
</tr>
<tr>
<td>Benevolence</td>
<td>.19*</td>
<td>.07</td>
<td>.57</td>
<td></td>
</tr>
<tr>
<td>Discredibility</td>
<td>-.24**</td>
<td>.09</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>Malevolence</td>
<td>-.28**</td>
<td>.11</td>
<td>.55</td>
<td></td>
</tr>
</tbody>
</table>

**Significant at p < .01; *Significant at p < .05; +Significant at p < .10

From four eBay sellers, and they were given ample time to review the four manipulated seller profiles (Appendix A). Subjects were instructed to pay particular attention to the seller profiles because those were to be used in the fMRI experiment, and they typically spent about 20 minutes reading the feedback text comments in the four seller profiles. Further clarification about the fMRI study was available from a research assistant.

The subjects then entered the fMRI scanner, lying comfortably on their back. Visual stimuli (Figure 2) were projected through fiber-optic goggles connected to a computer. First, one randomly selected seller profile was presented for 3 seconds, which served to remind the subjects about each seller without having to read the individual text comments. This was followed by a randomly selected measurement item for a randomly selected construct for the focal seller. Each item was shown for 5 seconds without the scale, which was shown to be ample time for subjects to read and process (Dimoka and Davis 2008). Then, the seven-point Likert-type scale appeared, and the subjects selected their choice by depressing one of the seven buttons using a fiber-optic mouse they held with their right hand. Subjects had unlimited time to make their choice, but the actual time they took in practice was about 2 seconds. After clicking on their choice, they were shown a new randomly selected seller profile followed by a randomly selected item. This procedure was repeated for all sellers, measurement items, and control items, as shown in Figure 2.

Since fMRI studies need some repetition to obtain statistically significant brain activations, similar to multi-item scales, subjects answered a set of nine similar, but not identical measurement items (Appendix B). The total time subjects spent in the fMRI scanner was $6 \times 9 \times 4 \times 10 = 2,160$ seconds (5 constructs plus 1 control times 9 items for each construct times 4 sellers times about 10 seconds for each cycle) or 36 minutes (plus 4 minutes for obtaining the anatomical brain images in the beginning of the fMRI experiment). Finally, upon completion of the fMRI experiment, the subjects were thanked, debriefed, and dismissed.

---

6fMRI studies require appropriate control variables to “cancel out” spurious brain activation due to visual stimuli, movement, and other sources of noise, and thus isolate brain activation only associated with the experimental stimuli. For the control items, the subjects were randomly shown each seller profile, and they were instructed to press a certain button on the fiber-optic mouse using a statement that resembled the study’s measurement items in terms of format type and length (Appendix B). Similar to the other focal constructs, nine control items were used for each seller.
fMRI Data Analyses and Results

Before analyzing the fMRI data, the corresponding behavioral data during the fMRI study was analyzed to ensure that the manipulation of the four seller profiles was also evident to the fMRI subjects, and also that their psychometric responses in the fMRI scanner were similar to the behavioral results (shown in Table 2).

Table 4 summarizes the descriptive statistics for the fMRI study’s principal constructs, which closely correspond to the behavioral results (Table 2). Seller HL is associated with high trust and very low distrust, while seller LH is associated with high distrust and very low levels of trust. Also, the HH seller has lower trust and higher distrust than the LH seller, consistent with McKnight et al. (2003) who argued that distrust stimuli are more salient than trust stimuli. Since the self-reported psychometric data of trust and distrust are similar across the behavioral and fMRI studies (p < .05), this implies that the fMRI context did not bias the subjects’ responses, and the behavioral data are similar across the fMRI and the behavioral studies.

The analysis of the fMRI data was performed with SPM5 freeware. Whole-brain 3Tesla fMRI data were acquired in a time-series of approximately 36 minutes to provide 28 contiguous 5 mm thick brain slices allowing the subjects to respond to the measurement items with a gap of approximately 10 seconds between two measurement items. All brain activations were obtained during the latter part of the 5-second period where the subjects were reading each measurement item (before posting their response) to prevent hand movement when responding to the measurement item and to assure temporal separation between the brain activity while reading the measurement item and the self-reported response on the seven-point Likert-type scale.

The first stage of the analysis aimed at identifying the neural correlates of trust and distrust and of their dimensions. The analysis was undertaken by contrasting the brain activations of each construct for each seller profile relative to the control items, aiming to localize the neural correlates of each construct while minimizing confounds (e.g., visual stimuli and movement). The difference between the experimental and the control image reflects the brain activation only due to the measurement item for each seller profile. The brain images were first analyzed and localized for each subject. Then, second-level one-sample t-tests were performed on the aggregate data to create random-effect group analyses for two constructs (e.g., trust) and two dimensions (credibility/benevolence) for each of the four sellers, totaling 16 different conditions. For each condition at the aggregate level, statistical parametric maps were generated with the z-value of each voxel (3D pixel) that exceeded a p < .05 threshold. The first analysis included all measurement items of trust (all 18 items

---

The z-values in SPM correspond to the unit normal distribution that renders the same p-values as the t-statistic.
of credibility and benevolence) and distrust (all 18 items of discredibility and malevolence) as stimuli, aiming to identify the brain areas activated by trust or by distrust. Region of interest (ROI) analysis was used as the standard method to define an area of interest in the brain within which to observe fMRI data (King-Casas et al. 2005). Figure 3 shows all brain activations (ROI analysis) associated with trust and distrust for the four focal seller profiles (Lewicki et al. 1998) that span the continuum of trust and distrust.

**fMRI Results: Neural Correlates of Trust**

As Figure 3 shows, the HL seller spawns activations in the caudate nucleus ($z$-value = 3.36, $p < .01$), as expected. However, there was also activation in the putamen ($z = 2.92$, $p < .01$), another area associated with the prediction of rewards. The HL seller also spawned activation in the paracingulate cortex (anterior PCC) ($z = 3.67$, $p < .001$), and moderate activation in the orbitofrontal cortex ($z = 2.12$, $p < .05$). The LL seller has weaker activations in the caudate nucleus, putamen, and anterior paracingulate cortex ($p < .05$), yet stronger activations in the orbitofrontal cortex ($z = 2.77$, $p < .01$). Since the HH and LH sellers have stronger activations in the orbitofrontal cortex ($z = 3.51$ and $z = 3.78$, respectively, $p < .001$), while neither the LH or HH seller spawns any significant activation in the reward or prediction areas of the brain, these findings suggest that trust is associated with lower activation in the orbitofrontal cortex and with higher activation in the anterior paracingulate cortex, caudate nucleus, and putamen. These findings jointly support H1a.

---

8. The level of brain activation is determined by the value of the t-test. The peak $z$-value in a certain brain area is a proxy for the overall strength of all adjacent voxels of the given brain area (Delgado et al. 2005; Rilling et al. 2002).

9. Direct comparisons between the HL and LL sellers confirmed these differences in brain activation ($p < .001$).

**fMRI Results: Neural Correlates of Distrust**

Figure 3 also shows the neural correlates of distrust, which are primarily evident in the LH seller that shows strong activation in the bilateral amygdala and insular cortex ($p < .001$). Similar, albeit weaker, activations are observed for the HH seller (direct comparisons between the LH and HH sellers confirmed the difference at $p < .01$). However, neither of the two low distrust sellers rendered a significant activation in the amygdala and insular cortex. Since strong activations in the bilateral amygdala and insular cortex are only associated with the two high distrust sellers and not the low distrust sellers, there is support for H2a.

**fMRI Results: Neural Correlates of the Dimensions of Trust**

In terms of the neural correlates of the two dimensions of trust, Appendix D shows the fMRI results with the respective nine measurement items for credibility and benevolence separately (Figures D1 and D3). For credibility, there is a similar pattern of activation in the orbitofrontal cortex in the HL (no activation), LL ($z = 2.01$, $p < .05$), HH ($z = 3.01$, $p < .001$), and LH ($z = 3.77$, $p < .001$) sellers as in overall trust. There is also significant activation ($p < .01$) in the caudate nucleus and putamen for the HL seller. Also, for benevolence, as proposed, there is significant activation ($z = 2.75$, $p < .01$) in the anterior paracingulate cortex for the HL and LL sellers, but no activation for the high distrust sellers (LH and HH sellers). However, benevolence also triggers activation in the caudate nucleus and putamen for the HL and LL sellers ($p < .01$). The results suggest that there is no perfect separation between credibility and benevolence since both trust dimensions activate the caudate nucleus and putamen, implying that benevolence is also linked to higher rewards. Thus, there is partial support for H1b in terms of the orbitofrontal cortex (only credibility) and anterior paracingulate cortex (only benevolence), but not in terms of the brain’s reward areas (both credibility and benevolence).
Figure 3. Brain Activation Associated with Trust and Distrust for the Four Seller Profiles
fMRI Results: Neural Correlates of the Dimensions of Distrust

In terms of the neural correlates of distrust, Appendix D shows the fMRI results for the dimensions of credibility and malevolence separately (Figures D2 and D4). For credibility, there is brain activation in the insular cortex for both the LH and HH sellers (higher levels of activation in LH than in HH) (p < .01), and no significant activation in the low-distrust sellers (LL and HL). For malevolence, both distrust sellers activated the amygdala and the insular cortex (p < .01), while the low-distrust sellers (LL and HL) rendered no significant activation. While the insular cortex is activated by both credibility and malevolence (perhaps because the potential for loss has both a calculative and an emotional element), the results show a clear separation in terms of the amygdala (only activated by malevolence), thus partially supporting H2b.

fMRI Results: Trust Versus Distrust

Integrating the neural correlates of trust and distrust shows a stark contrast in the activated brain areas across sellers. Trust is associated with the brain’s reward, prediction, and uncertainty areas, while distrust is associated with the brain’s intense emotions and fear for loss areas. Since trust and distrust span distinct brain areas, there is support for H3 from the brain activations that trust and distrust are distinct constructs.

fMRI Results: Predicting Price Premiums

H4 proposes that the level of brain activation in the areas associated with trust and distrust can predict what price premium the subject will offer to each seller. Peak level activations were obtained for all areas activated for trust and distrust (Figure 3) for each subject and seller profile and these levels were used as independent variables in a regression model to predict price premiums across sellers (Table 5). Using the peak level of brain activation is a standard practice in fMRI studies (e.g., Delgado et al. 2005; Rilling et al. 2002).

As Table 5 attests, the brain areas associated with trust and distrust adequately predict price premiums (R² = .49%). The areas associated with distrust (amygdala and insular cortex) have the strongest (negative) effects, explaining 29 percent of the variance, while only 20 percent of the variance in price premiums is explained by the neural correlates of trust (Cohen’s F = .18, which denotes a medium-large difference in R²). Besides the caudate nucleus and putamen whose tolerance values are slightly below the recommended 0.70 value, all brain areas are clearly discriminant from each other, denoting no problems of collinearity. This suggests that the neural correlates of distrust are stronger predictors of price premiums than those of trust, supporting H4.

Figure 1 shows that the effects of brain activations on price premiums is mediated by the self-reported responses on trust and distrust since the brain activations temporally precede the subjects’ self-reporting. To examine whether the self-reported data fully mediate the effect of the corresponding brain activations, the psychometric measures of trust and distrust were used to predict the level of price premiums (Table 6).

As shown in Table 6,10 the adjusted variance explained by the dimensions of trust and distrust (R² = 31%) is much lower (ΔR² = 18%) than that of corresponding brain activations (R² = 49%) (Table 5). Moreover, the formal test of mediation (Baron and Kenny 1986) shows that when including all independent variables in a stepwise regression model (omitted for brevity), the significant predictors were only the brain areas identified in Table 5. Hence, the self-reported data do not mediate the effect of the corresponding brain data, albeit they presumably reflect the observed brain activations. This is perhaps because the self-reported data are more subject to common method bias that inflates their correlations and creates collinearity concerns (which are reflected by the relatively low tolerance values in Table 3), or because the psychometric measures of trust and distrust may not fully capture the information captured by the corresponding brain activations.

fMRI Results: Activation in Other Brain Areas

In addition to the hypothesized brain areas, other brain areas activated during the fMRI study included the dorsolateral prefrontal cortex (DLPFC). The DLPFC is a unique and highly evolved part of the brain that is associated with cognitive information processing. While all seller profiles activate the DLPFC, the LL seller has the lowest activation (z = 1.90, p < .10), the HH seller spawns the highest activation (z = 3.8, p < .001), while the HL and LH sellers spawn moderate activation (z = 2.37 and z = 2.68, respectively) (p < .05). These findings may suggest that the subjects engaged in a cognitive effort to process the feedback profiles of the four sellers, and the LL seller profile is the simplest to process (no

10Please note that the results of the fMRI study in Table 6 are very similar to those of the behavioral study (Table 3), both in terms of variance explained and the standardized regression coefficients. This further supports that the fMRI setting did not bias the self-reported responses, which are very similar across both the behavioral and fMRI studies.
Table 5. Regression Results for Predicting Price Premiums with Trust and Distrust (fMRI Study)

<table>
<thead>
<tr>
<th>Brain Area</th>
<th>Standardized Beta Coefficient</th>
<th>ΔR²</th>
<th>Tolerance</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caudate Nucleus</td>
<td>.05**</td>
<td>.00</td>
<td>.51</td>
<td>.49</td>
</tr>
<tr>
<td>Putamen</td>
<td>.21*</td>
<td>.04</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>Orbitofrontal Cortex</td>
<td>-.22**</td>
<td>.05</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Anterior Paracingulate Cortex</td>
<td>-.31**</td>
<td>.11</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Bilateral Amygdala</td>
<td>-.42**</td>
<td>.16</td>
<td>.81</td>
<td></td>
</tr>
<tr>
<td>Bilateral Insular Cortex</td>
<td>-.35**</td>
<td>.13</td>
<td>.76</td>
<td></td>
</tr>
</tbody>
</table>

**Significant at p < .01; *Significant at p < .05; N/S Nonsignificant

Table 6. Regression Results for Predicting Price Premiums with Trust and Distrust (fMRI Study)

<table>
<thead>
<tr>
<th>Psychometric Measures</th>
<th>Standardized Beta Coefficient</th>
<th>ΔR²</th>
<th>Tolerance</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credibility</td>
<td>.14*</td>
<td>.04</td>
<td>.51</td>
<td>.31</td>
</tr>
<tr>
<td>Benevolence</td>
<td>.17*</td>
<td>.06</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>Discredibility</td>
<td>-.24**</td>
<td>.08</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>Malevolence</td>
<td>-.35**</td>
<td>.12</td>
<td>.55</td>
<td></td>
</tr>
</tbody>
</table>

**Significant at p < .01; *Significant at p < .05

extraordinary text comments) while the HI seller profile is the most complex with both outstanding and abysmal text comments. The other profiles are moderately complex with either outstanding or abysmal text comments. These findings imply that the subjects did process the different seller profiles and exerted differential cognitive effort in their evaluation. The activation level in the DLPFC did not have any significant effects on trust, distrust, or price premiums.

fMRI Results: Gender Differences

Since the results are largely driven by the brain activations in the more emotional areas of the brain, we redid the entire analysis by gender. While the small sample size for women (n = 6) precludes us from inferring definite conclusions, there were several interesting results. First, while the overall activation pattern (Figure 3) was consistent across men and women, the brain activations in the limbic system (e.g., amygdala, insular cortex, putamen, caudate nucleus) were generally higher for women, while the level of brain activation in the prefrontal area (orbitofrontal cortex) was higher for men. Second, while benevolence was not hypothesized to spawn activation in the brain’s reward areas, the observed activation was due almost exclusively to women. Finally, the non-hypothesized activation in the insular cortex for malevolence was also caused, almost exclusively, by women. In sum, these results imply that emotional responses in the brain are more salient for women, implying that there are differences between genders, consistent with the IS literature (e.g., Gefen and Straub 1997; Venkatesh and Morris 2000).

Discussion

Key Findings

This study has several interesting findings, opening the “black box” of the brain. First, it identifies the neural correlates of trust and distrust, showing that trust is associated with the brain’s reward, prediction, and uncertainty areas, while distrust is associated with the brain’s intense emotions and fear of loss areas. While the corresponding psychometric data could not clearly distinguish between trust and distrust, the brain data suggest that trust and distrust and their dimensions are distinct and activate different brain areas. Second, the study separates the dimensions of trust and distrust by showing that credibility and discredibility are mostly associated with the brain’s cognitive areas (prefrontal cortex), while benevolence and malevolence are mostly associated with the emotional areas (limbic system). Third, the identified brain areas adequately predict price premiums, and the levels of brain activation have a stronger predictive power than the corresponding self-reported psychometric measures. Finally, there is a close relationship between the behavioral
results taken from the fMRI and the classic lab experiment, implying that the unique fMRI context did not cause any biases in the subjects’ behavioral responses.

**Implications for Theory and Practice**

The use of physiological tools in the social sciences is starting to gain prominence, and the use of fMRI tools to study brain activity is attracting unparalleled attention in the academic and practitioner literature because of its potential to uncover interesting implications that cannot be derived easily with existing tools. This study’s implications for trust and distrust in IT-enabled impersonal online exchanges follow.

**Implications on the Nature of Trust**

First, trust is associated with the caudate nucleus and putamen that jointly constitute the dorsal striatum, an area associated with reward prediction (Haruno and Kawato 2006). While the caudate nucleus has already been associated with trust (e.g., King-Casas et al. 2005), to our knowledge, the putamen has not been linked to trust, thus constituting a new hidden area that is activated by trust. While both areas are part of the reward brain that receives dopaminergic input from the midbrain (Knutson et al. 2001), they have distinct functions. The caudate nucleus is involved in comparing actual and expected rewards and is activated proportionately to the magnitude of the expected reward (Hsu et al. 2005), while the putamen is involved in evaluating rewards to make predictions (Haruno and Kawato 2006). Thus, these two areas work together to compare and evaluate the potential gains of trust to guide behavior. This finding is perhaps why only the putamen has a significant effect on price premiums, while the caudate nucleus does not (Table 3). While only credibility was expected to be associated with cognitively evaluating rewards, benevolence also activates these two reward areas. This result implies that the expected rewards of trust are not only associated with economic rewards, but also with social rewards linked to the seller’s expected goodwill.

Second, neuroimaging studies have identified the anterior paracingulate cortex as a key area of the “social brain” for predicting the behavior of others (Gallagher and Frith 2003). To build goodwill trust, partners must infer each other’s intentions to assess whether to trust someone and whether the trustee will reciprocate their trust. The anterior paracingulate cortex is designated to comprehend the mental states of others, and it is shown to be a unique area associated only with benevolence (and not credibility). Since the anterior paracingulate cortex is only activated when a person engages in a social and not any other type of interaction (Walter et al. 2004), this explains why there is brain activation only for the HL and LL sellers (but not for LH and HH sellers), and only for benevolence (but not credibility). This has implications for the value of fMRI to uncover hidden areas that have a strong predictive role on economic outcomes.

Third, the orbitofrontal cortex that has been linked to uncertainty (Hsu et al. 2005) has a negative link to trust, and it varies across seller profiles. While the HL and LL sellers trigger modest levels of activation, the HH and LH sellers trigger stronger activation. This result is explained by the high degree of uncertainty associated with these two types of sellers. Negative information alerts the brain to assess the level of uncertainty, and the orbitofrontal cortex is used to perform this function. These findings have implications for the nature of trust since uncertainty is an inherent component of the trustor’s decision whether to assume vulnerability.

Taken together, the brain areas imply that trust is associated with the trustor’s assuming the uncertainty (orbitofrontal cortex) of engaging in a social interaction with a trustee (anterior paracingulate cortex) based on expectations of future economic and social gain (caudate nucleus and putamen). These findings have implications for the trust literature in terms of better understanding the nature and characteristics of trust.

**Implications on the Nature of Distrust**

The findings from this study also shed light and have implications on the nature of distrust. First, distrust is associated with the amygdala, a critical area triggered by intense and sudden emotional states (Kenning et al. 2005), particularly negative ones (LeDoux 2003). This finding corresponds to the IS literature that has viewed distrust as a “hot” and emotion-laded construct (McKnight and Chervany 2000), which helps “buffer oneself from the effects of another’s conduct” (Lewicki et al. 1998, p. 439). Amygdala activation tends to be abrupt and short-lived to enable quick retreat to prevent harm from a potentially injurious situation, consistent with the view of distrust that stresses vigilance against danger. This is perhaps the reason why distrust is only associated with malevolence (not discredibility), which specifically stresses concerns for harm (Luhmann 1979). Besides, while the amygdala is associated with emotional (versus cognitive) aspects, it has also been linked to the interaction between emotion and cognition in decision making (Phelps 2006). This is perhaps why the amygdala has strong effects by translating intense negative emotions to cognitive outcomes. This finding is also consistent with Murphy and Zajonc...
Second, distrust is associated with the insular cortex, another primary sensory cortex for visceral information and autonomic arousal. While the insular cortex is also triggered by intense negative emotions (Wicker et al. 2003), its primary focus is the fear and anticipation of losses (Preuschoff et al. 2008). The insular cortex captures another aspect of distrust. Similar to the positive rewards of trust that are associated with both dimensions of distrust. While the amygdala and insular cortex are distinct brain areas, Winston et al. (2002) argued that the amygdala generates changes in bodily states that are remapped to the insular cortex. Coupled with the finding that insular cortex activation predicts aversive behavior (Rilling et al. 2008; Sanfey et al. 2003), the abrupt and intense negative emotions in the amygdala are remapped as fear of loss in the insular cortex, which drives the decision not to assume the vulnerability of a potentially risky and/or harmful situation.

**Implications for the Distinction and Relationship between Trust and Distrust**

Using functional neuroimaging data, this study shows that trust and distrust may be distinct constructs that span different brain areas. This finding is in contrast to the behavioral data that could not distinguish between trust and distrust and their dimensions in terms of discriminant validity (Appendix C), even though the brain activations were caused by the exact same IT stimuli. The distinction is notable since the areas associated with distrust do not trigger activation in the high trust sellers (HL or LL), and the brain activations associated with trust do not trigger activity in the high distrust sellers (LH or HH) (besides the orbitofrontal cortex). The orbitofrontal cortex, which distinguishes among levels of uncertainty, was the only area activated by all four seller profiles. Since trust and distrust reflect the trustor’s decision to assume vulnerability based on the level of seller uncertainty, trust and distrust can be viewed as the opposite ends of a continuum only in terms of assessing uncertainty. The orbitofrontal cortex as the brain area activated by all sellers depending on their level of uncertainty may help resolve the debate in the literature whether trust and distrust are distinct constructs or not, by supporting the view that the only continuum revolves around calculating uncertainty.

Trust and distrust are also associated with expected positive rewards and negative losses, respectively. While one may argue that rewards and losses are perfect substitutes, the literature has distinguished between gains and losses and attributed a stronger weight on negative outcomes (Kahneman and Tversky 1979). This finding confirms Lewicki et al.’s (1998) view that using the average of trust and distrust is misleading, and that low distrust is not the same as high trust, nor is high distrust is the same as low trust. This also explains why positive versus negative rewards trigger different brain areas, and why the emotionally laded areas for negative rewards have stronger effects on price premiums than those that relate to positive rewards. This finding also supports the literature that sees distrust as having a distinct, stronger emotional component than trust (Luhmann 1979), and that distrust arises emotionally versus cognitively (Sitkin and Roth 1993).

Finally, trust is associated with the anterior paracingulate cortex, which predicts intentional engagement in a social relationship (Winston et al. 2002). This finding is in contrast to the autonomic engagement in the amygdala and insular cortex that seeks to prevent a harmful event. These brain areas stress the distinction between trust and distrust in terms of type of response (intentional versus autonomic), nature of response (collected versus frenzied) (McKnight and Chervany 2000), and time horizon (long-term versus short-term). These findings are consistent with the literature that views trust as developing slowly over time through careful deliberation, while distrust is quick and episodic, based on emotional cues (Wiethoff and Lewicki 2005).

**Implications for the Dimensions of Trust and Distrust**

The areas of brain activation associated with trust and distrust also shed light on their dimensionality. While there is no perfect distinction in the activated brain areas since both dimensions of trust and distrust activate the reward and loss areas, respectively, there is clear distinction in terms of the prediction (only benevolence), uncertainty (only credibility), and intense negative emotions (only malevolence). The previous findings support the multidimensionality of both trust and distrust by identifying at least two dimensions, respectively, and they also support the notion that the two proposed dimensions of trust (Gefen et al. 2003) and distrust...
Implications for the Relative Effects of Trust and Distrust and Their Dimensions

The fMRI study showed that the brain areas associated with distrust (amygdala and insular cortex) are the strongest predictors of price premiums. These findings correspond to the behavioral data from both studies that show that the two dimensions of distrust have a stronger effect on price premiums than the two dimensions of trust. These findings confirm the warning by McKnight et al. (2003) that ignoring distrust may yield weak predictions due to a missing variable problem, and that distrust stimuli are more salient than trust stimuli. The stronger effect of distrust relative to trust can be partly explained by the nature of distrust, which is more emotional than trust (McKnight and Choudhury 2006), and emotions are more salient and are easier to access and draw upon (Luhmann 1979; McKnight et al. 2003). In contrast, less emotion-laden variables, such as trust, are likely to erode over time. Drawing from prospect theory that showed losses to be weighed more heavily than gains (Kahneman and Tversky 1979), empirical work in the word-of-mouth (WOM) communication and satisfaction literatures (Anderson 1998) confirmed that buyers place more weight on negative than on positive WOM. This finding is consistent with Cho (2006), who found that distrust information “tends to be more noticeable, perceived more credible, and weighed more on a behavioral decision than the positive information of a similar magnitude” (p. 27). It is also consistent with the IS literature on end-user training (Galletta et al. 1995) that unfavorable WOM statements were more potent than favorable WOM statements. The fMRI study not only verifies the propositions from this literature with brain data, but it also explains that the stronger effect of distrust is due to the strong emotional component of its neural correlates (mainly the amygdala) due to the secretion of the chemicals epinephrine and norepinephrine, the “fight or flight” hormones secreted when danger is present. The fMRI study thus offers a neurological and a physiological justification to the literature on the stronger effects of negative information. Finally, it supports the proposition by McKnight et al. (2003) that academics and practitioners should also focus their efforts on reducing distrust (a distinct and more influential construct) as opposed to only building trust.

The fMRI results are also consistent with the behavioral results from studies that both benevolence and malevolence have stronger effects on price premiums than credibility and discredibility, respectively. While the salience of benevolence relative to credibility has been empirically shown (Pavlou and Dimoka 2006), to our knowledge, the relative importance of the dimensions of distrust has not been examined. This study offers a justification for these relative differences by identifying the neural correlates of the dimensions of trust and distrust. As noted above, the amygdala, which largely underlies malevolence, has an important role by engaging fight or flight hormones (epinephrine and norepinephrine) to protect against danger. The anterior paracingu late cortex, which only underlies benevolence and not credibility, has a strong role in building social interactions. While credibility and discredibility were expected to be associated with rewards and losses, respectively, the goodwill aspect of trust and distrust (benevolence and malevolence) is also associated with the reward and loss brain areas, thus also attesting to their role in price premiums. While the caudate nucleus and putamen are both associated with the positive rewards of trust, the putamen is the only significant predictor. This finding is consistent with the literature that the putamen is linked to reward prediction, and the caudate nucleus with the evaluation of the reward (Haruno and Kawato 2006).

In sum, the fMRI study not only helps to predict theoretical propositions and explain empirical findings from the behavioral sciences with direct brain data, it also renders neurological and chemical justifications.

Limitations and Suggestions for Future Research

This study has certain limitations that call for interesting opportunities for future research. The four seller profiles aimed at creating stark differences across sellers to uncover the neural correlates of high versus low trust and distrust. However, they may not correspond to real-life seller differences in online auctions. Since it is beyond the scope of this study to offer practical recommendations for online sellers, future research could create more realistic seller profiles that correspond to actual sellers. Besides, the dimensions of trust and distrust moved in the same direction by either having outstanding or abysmal credibility and benevolence text comments (or both), thereby not having seller profiles with high levels of credibility/discredibility and low levels of benevolence/malevolence. While this setup made it more difficult to distinguish the respective dimensions of trust and distrust from each other, the brain activations were starkly different across the dimensions of trust and distrust. Future research could allow the dimensions of trust and distrust to vary within a construct,
potentially creating more interesting and realistic seller profiles.

The proposed two-dimensional view of trust and distrust used to trigger the respective brain areas implies at least two dimensions, but it does not preclude a higher-order dimensionality for trust and distrust. For example, credibility can be split further into competence, honesty, and integrity (Gefen et al. 2003). Distrust can also be broken down into more dimensions (McKnight and Chervany 2000). The dimension of credibility triggers activation in three distinct brain areas, perhaps implying the need for further research. Future fMRI studies can thus examine a higher level dimensionality for trust and distrust.

An important question is the temporal ordering of trust and distrust and of their respective dimensions. While not examined in our study, the timing of brain activations associated with certain constructs can shed light on their temporal ordering and even their causal identification (Yoon et al. 2009). Future fMRI studies are well positioned to examine the temporal ordering of trust and distrust and their underlying dimensions.

In terms of the antecedents of trust and distrust, this study has already generated the seller profiles aimed at spawning trust and distrust. However, having identified the neural correlates of trust and distrust, future research can test how IT design artifacts can affect trust and distrust using the identified brain areas as the dependent variables, and IT artifacts can be tested on their ability to build trust and reduce distrust.

Kramer (1999, p. 593) states that distrust is “very difficult to invalidate through experiences” and that “presumptive distrust tends to be perpetual distrust” because trusting people avoid new experiences and do not learn from them. This logic is consistent with our fMRI findings that intense negative emotions, such as the insular cortex, can have a long-lasting impact on a person’s behavior. On the other hand, familiarity is shown to build trust gradually (Gefen 2000), a finding consistent with the brain activations for trust in our fMRI study that have a more enduring nature, such as the anterior paracingulate cortex and dorsal striatum. Future research can study how familiarity may shape the neural correlates of trust versus distrust over time.

Our study used peak level activation to infer the level of brain activation in a certain area to predict price premiums. Peak activation is a standard technique (e.g., Delgado et al. 2005; Rilling et al. 2002), and it is assumed to reflect the level of activation in the entire set of adjacent voxels in the focal brain activity. Since it only captures the level and not the localiza-

tion of the brain activity, peak activation cannot explain the ability of brain data to distinguish between trust and distrust. However, in terms of its ability to predict price premiums relative to the behavioral data, peak activation can either inflate or deflate the “true” level of brain activation and accordingly be more or less predictive. While there is no systematic bias in favor of the brain data since it can go either way, there is active research on how to best estimate the true level of brain activation, and future research can replicate our findings with different measures of brain activity.

fMRI is a noninvasive method that does not rely on chemical or radioactive material. However, other neurobiological methods can shed light on trust (and perhaps distrust). For example, neurobiological studies show that the chemical oxytocin raises people’s propensity to trust (Kosfeld et al. 2005; Zak et al. 2004). Oxytocin also dampens activity in the amygdala (Kirsch et al. 2005), thus reducing the impact of betrayal (similar to distrust). Zak et al. (2005) showed distrust to be linked to higher levels of dihydrotestosterone, an effect present in men but not women. While such invasive methods to induce behavior are rare in the IS literature, future research may examine gender to identify differences.

fMRI studies are difficult in terms of the physical conduct of the experimental design, the cost of the fMRI scanner, and the complexity of the data analysis due to the large amounts of imaging data. Also, there are concerns due to the idiosyncratic nature of the fMRI setting. While our fMRI and behavioral data give similar results (which is common in fMRI studies), future research could examine how the unique fMRI setting creates concerns about external validity versus traditional laboratory studies. Despite these issues, neuroimaging studies are prevalent in the social sciences. Hopefully, future fMRI studies will be conducted in IS to build a broad foundation for identifying the neural correlates of IS constructs and building superior IS theories.

Final Remarks

Opening the black box of the brain has the potential to enhance our understanding of trust and distrust, as this study attests. Several new insights on the nature of trust and distrust have emerged, such as the neural underpinning of their dimensions, their distinction and relationship, and their relative effects on behavior. Besides the predictive validity of the identified brain areas on certain behavioral and economic outcomes (price premiums), perhaps the greatest potential of functional neuroimaging is to justify theoretical propositions (such as the distinction and relationship between trust and
distrust), explain empirical findings (such as the relative effects of trust and distrust), shed light on the nature of theoretical constructs (such as trust and distrust), and complement existing measurement sources (such as uncovering hidden processes that cannot be easily captured with self-reported methods). This study seeks to encourage neuroimaging studies to explore the nature, dimensionality, antecedents, and consequences of trust and distrust even further, and also to explore other IS constructs that could benefit by identifying their neural correlates.

Acknowledgments

I would like to thank Senior Editor David Gefen for his excellent guidance that has tremendously improved the quality of this manuscript. I would also like to thank associate editor Dennis Galletta for his detailed and insightful comments that enhanced the manuscript’s presentation, structure, and contribution. Moreover, I thank the three anonymous reviewers for their developmental feedback and excellent suggestions. I would also like to thank Richard Bagozzi for his constructive comments on an earlier version of this manuscript. Finally, I would like to thank Feroze Mohamed, Vasilis Megalookonomou, and Erickson Miranda for their assistance and support in the data collection and analysis.

This work was supported by the Institute for Business and Information Technology, and by the Center for Neural Decision Making, Fox School of Business, Temple University.

References


### About the Author

Angelika Dimoka is an assistant professor of Marketing and Management of Information Systems at the Fox School of Business, Temple University. She is also director of the Center for Neural Decision Making. She holds a Ph.D. from the Viterbi School of Engineering (specialization in Neuroscience and Brain Imaging) with a minor in MIS from the Marshall School of Business at the University of Southern California. Angelika’s research interests are in the areas of cognitive neuroscience and functional neuroimaging in marketing and MIS (Neuromarketing and NeuroIS), quantitative analysis of uncertainty in online marketplaces, and modeling of information pathways in the brain. Her research has appeared (or is scheduled to appear) in *Information Systems Research, MIS Quarterly, IEEE Transactions in Biomedical Engineering, Annals of Biomedical Engineering, IEEE in Biology and Medicine, Neuroscience Methods*, and the proceedings of the International Conference on Information Systems and the Workshop on Information Systems and Economics.