CIFE CENTER FOR INTEGRATED FACILITY ENGINEERING

Rating Systems for AEC Subcontractors – The Impact of Rater Credibility

By

Martin A. Ekström
Hans C. Björnsson
Clifford I. Nass

CIFE Technical Report #134
July, 2002

STANFORD UNIVERSITY

If you would like to contact the authors, please write to:

c/o CIFE, Civil and Environmental Engineering
Stanford University
Terman Engineering Center
Mail Code: 4020
Stanford, CA 95305-4020

Abstract

In consumer electronic markets, rating mechanisms are important facilitators of trust between market participants. This paper investigates whether source credibility theory can support the evaluation of bidders in Architecture/Engineering/Construction (AEC) electronic market places. In contrast to consumer electronic marketplaces, the raters in AEC communities are skilled and connected, necessitating a reputation mechanism to account for the relationship between the user and the rater. To solve this problem, TrustBuilder, a prototype-rating tool, facilitates information between peer industry practioners by calculating a weighted rating based on source credibility theory. An experiment shows that AEC industry practitioners who evaluate bidding subcontractors trusted the information provided by TrustBuilder significantly more than the information from a standard unweighted rating tool.

Introduction

E-commerce, "the enabling technology that allows businesses to increase the accuracy and efficiency of business transaction processing," (Trepper, 2000) can decrease the costs of business-to-business (B2B) transactions (Malone and Laubacher, 1998) (Lucking-Reiley and Spulber, 2000). One example of an industry where entrepreneurs, as well as investors, envisioned reaping substantial benefits from electronic commerce is the large and fragmented Architecture Engineering Construction (AEC) industry. In April 2000 approximately $1 Billion was invested in around 200 AEC e-commerce start-ups. The problem is that if AEC industry practitioners are to use electronic market places to find new market partners, they must be certain that they can trust the participating organizations. In reality, the number of transactions conducted in these new AEC e-market places turned out to be very low, and only a handful of them were still in business by May 2002. One key problem was that if industry practitioners were to use electronic marketplaces to find new market partners, they must be certain that they can trust the participating organizations.

However, the Internet can also facilitate trust between market participants, as it enables information sharing and synthesizing (Hanson, 1999). The most significant manifestation of this concept is found in consumer-to-consumer electronic commerce. In virtually every transaction, the 40,000,000 members of eBay's online community make decisions about whether or not to trust an unknown seller or buyer. When buying and selling everything from sport's memorabilia to used cars, eBay's users can take advantage of a rating system. Based on an investigation of the eBay rating system, Reznick and Zeckhauser (2001) argue that the Internet provides a superior mechanism for distribution of the information which supports trust decisions. Rating systems are important features of major consumer-to-commerce (C2C) (e.g., eBay), as well as Business-to-Consumer (B2C) (e.g., Amazon.com) market places. Consumers shopping on eBay or Amazon can, with little effort and at no cost, obtain and contribute to the quality judgments made by peer consumers.

Online rating tools have so far been slower to catch on in B2B electronic markets, such as AEC, than in electronic commerce, which targets the consumer market. We argue, that B2B settings, are fundamentally different from consumer markets, and that a rating tool which targets, for example, AEC bidding, should reflect these differences. Ratnasingham and Kulmar's (2000) argue that in B2B electronic commerce, researchers and professionals must consider the "role of trust between human actors," a notion that goes back at least to Aristotle (the concept of *ethos*). While transactions in C2C marketplaces mostly take place between animus users, construction companies often conduct a substantial part of their business with a small number of recurring suppliers. Furthermore, the goods and services purchased by a general contractor are in general more complex than consumer goods. It does not take much experience and expertise to judge the performance of a seller of a used tennis racket on eBay. When evaluating a paving contractor's ability to maintain schedule, on the other hand, rater experience and competence become important factors.

In this research project, we present a reputation mechanism grounded in source credibility theory, an area of communication science that explicitly studies and formalizes trust between human actors. This paper also reports on an experiment which compared the added value of this reputation mechanism relative to a standard rating model in a context where industry practitioners evaluate bids in a business-to-business electronic marketplace. As a result, this paper investigates whether a consideration of source credibility can add value in AEC electronic commerce.

Background

Available Measures for the Evaluation of AEC Subcontractors

Based on my interviews with fourteen AEC practitioners, I determined a list of the most General contractors important criteria used by to evaluate bidding subcontractors. We have divided the criteria into three categories:

Objective Measures: Subcontractor performance can be measured quantitatively in a number of ways. Objective measures, such as the number of times an event occurs, enable us to define an unambiguous scale, which makes it less important who provides the measures/ratings.

Subjective measurements provided by a reputable third party: This type of criteria include, ratings from credit and insurance organizations. The third parties do apply elements of qualitative judgment when assessing these criteria, but the general assumption is that the measures are consistent for all evaluated contractors. Aggregation is therefore not a major issue for this type of criteria since they are generally provided by a single source perceived as both skilled and unbiased (e.g., Dun & Bradstreet for credit ratings in the US).

Subjective measurements provided by peer industry practitioners: The focus of this research project is criteria where the performance can, in a practical setting, only be measured subjectively by peer contractors. It comprises important performance measures such as maintenance of schedule, collaborativeness, quality of work, administrative skills, change orders, and payment of second tier subcontractors/ suppliers. As Table 1, illustrates the identity of the source of the information is essential when peer practioners provide qualitative information. Not all peer practioners posses the same expertise when it comes to assessing subcontractors on criteria which are difficult to evaluate. In addition, there exist incentives for dishonesty since a rater may gain a competitive advantage by not revealing a subcontractor's actual performance on a project to potential competitors. However, for this type of criteria, not only is the source of the information important, but there is also likely to be a high number of different ratings to be aggregated into an overall rating. In the current practice, this type of information is exchanged between industry practioners, who use gossip, interviews, reference checking, and, in some cases, internal rating systems. The problems with the existing methodologies are that they are very time consuming, and that information risks being lost or distorted. Construction management researchers have also identified subjective information provided by peer contractors as important determinants of bid decisions

Table 1 Classification of information available for the evaluation of AEC subcontractors

Type of Information	Objective	Subjective I (Third party)	Subjective II (peer practioners)
Example	Project Experience	Credit Ratings	Maintenance of Schedule
Number of Sources	Low	Low	High
Expertise of information provider	Not Applicable	High	Varying
Incentives for dishonesty	Low	Low	High
Difficulty in aggregating information	Low	Low	High

Russell et al (1992) found that, in the process of pre-qualifying general contractors, owners place importance not only on objective criteria such as past experience and the number of completed projects, but also on criteria which can only be measured qualitatively by peer industry practitioners, such as "Change Order Frequency", "Schedule Performance", and "Willingness to Resolve Conflicts and Problems." Similarly, Holt et al (1995, 1996) found that the contractor's "actual quality", as well as cost and schedule overruns, influenced the owner's choice of contractors. However, the importance of the sources of this type of information has generally been neglected in earlier research. To support the evaluation of bidding AEC subcontractors, construction management researchers have proposed tools, which incorporate methodologies such as fuzzy logic (Okoroth and Torrance, 1998), analytical hierarchy process (Sik-Wah Fong and Kit-Yung Choi, 1999), mult-attribute utility theory (Wanous et al., 1999, Holt et al., 1995, Holt, 1996, Russel et al., 1992), and Multi Dimensional Scaling Techniques (Chinyio et al., 1998). However, also in this body of research, little attention has been given to the problem of evaluating subjective information provided by peer industry practitioners using qualitative measures. To avoid addressing this problem, the system can assume that all bidders are pre-qualified (Tseng and Lin, 2002), and that any differences between qualified subcontractors will only marginally impact decisions. Another approach is to limit the tool to internal use within an organization (Sik-Wah Fong and Kit-Yung Choi, 1999, Palaneeswaran and Kumaraswamy, 2000, Holt et al., 1995, Holt, 1996, Chinyio et al., 1998, Russel et al., 1992). Assuming that all raters within the organization are equally knowledgeable and trustworthy, all ratings should then be equally important. Moreover, Okoroth et al (1998) make the same assumption, even when raters from external organizations participate in the system. Finally, Wanous et al (1999) let the decision-maker herself input the ratings, and thus avoid any problem when qualifying subjective ratings. As the above analysis shows, there is an opportunity to contribute to the state of research in AEC bidding by investigating the applicability of a rating system that calculates rates depending on the source. More specifically, there is an opportunity to research how source credibility can support rating tools in AEC – bidding. Furthermore, little research has studied the added value of rating mechanisms in AEC bidding.

Theoretical Methodologies for Aggregating Information from Multiple Online Sources

Outside construction engineering and management, there is an emergent field of research which focuses on rating mechanisms in electronic commerce. This section discusses alternative approaches to construct reputation mechanisms which support B2B electronic commerce transactions. We give an overview of data analytical methods, network of trust models, and rule-based mechanisms and present the problems of implementing them in AEC electronic commerce.

Data Analytical Methods

Several researchers have presented methods which aggregates rating based on the outcome of past transactions.

Most well known in this category are collaborative filtering mechanisms which measure the extent to which users agree in terms of taste. The underlying reasoning can be exemplified as follows: If both Sam and Alice likes the movie *Titanic*, and Sam also likes the movie *Lord of The Rings*, then it is likely that Alice also will like *Lord of The Rings*. Pioneering work in using collaborative filtering for Internet-based recommender systems was done by Reznick et al in the GroupLens project (Reznick et al., 1994) and Shardanand and Maes (1995) automating "Word of Mouth". Collaborative filtering has been found to be most applicable when a large set of users rate items as difficult to quantify and describe (e.g., movie ratings, books).

As proposed by Avery et al. (1999), another approach is to apply statistical analysis of past ratings. Dellarocas (2000) has operationalized this concept by using clustering to differentiate between dishonest and honest raters.

Chen and Pal Singh (2001) propose a reputation hierarchy as a means to explicitly calculate rater reputation. Their model takes into account that a rater's expertise may vary depending on the domain that is being rated and also calculates the confidence level that can be associated with a rating. The weights are calculated through a propagation of endorsement (a function of discrepancy) across raters and groups organized in a hierarchy.

The major problem with data-analytical methods is that they require a substantial amount of data to obtain useful results. This may not be problem for C2C e-commerce merchants, such as Amazon.com, but could pose substantial difficulties in B2B e-commerce, especially during a start-up phase. Furthermore, it is not certain that the weights calculated by a collaborative filtering mechanism (or other data-analytic techniques) are consistent with user expectations. These models tend to over-emphasize the effect of strangers because they ignore personal trust.

Network of Trust

Building upon the assumption that people tend to trust the friend of a friend more than someone unknown, several researchers (e.g. (Zacharia et al., 1999)) have proposed formalizing people's "networks of trust" into rating applications. In practice, Epionion.com[i] has deployed a "Web of Trust" mechanism to a rating system for consumer reviews. The strength of such a solution is that it can build upon existing relationships, which could be important for a B2B community in the process of moving from an online to an offline presence. One important problem with this approach is measuring the trust that a user attributes to the members of his or her "Network of Trust". In interviews with industry decision-makers, we have found that the attitude towards the idea of a network of trust varies considerably. For some interviewees, the concept of trusting "a friend of a friend" seemed intuitive, while others did not consider it to be relevant whether they and an unknown rater turned had a common friend. A second problem is that the most common approach (e.g., Zacharia et al., 1999) is to use a single dimension to model trust, combining both a party's trustworthiness as a business partner (Will he cheat in business?) and credibility as an evaluator (Can I trust what he is saying?).

Rule-based mechanisms

Abdul-Rahman et al (2000) propose the deployment of rules to determine and update rater weights. These rules assess who to trust, based on outcomes of previous interactions. The problem with this approach is that the rules tend to be ad-hoc. For example, if a user A believes that rater B's rating of Supplier C is inaccurate, by how much would then A's trust in B decrease?

Source Credibility Theory

Trust and credibility are two fundamentally different concepts. Fogg and Tseng (1999) provide the following useful distinction:

trust <- "dependability"

credibility <- believability or "trust in information"

In the era of modern communication science, Hovland et al. (1953) performed pioneering work identifying perceived trustworthiness and expertise as the main dimensions of a source's credibility. The higher the trustworthiness and expertise a source is judged to have, the higher will be the importance given to information coming from that source.

Source credibility has been shown to be applicable in commercial settings (e.g., (Birnhaum and Stegner, 1979)), for the evaluation of organizations (Newhagen and Nass, 1989), as well as for the judgment of web content (Fogg and Tseng, 1999), but little research has investigated its applicability in electronic commerce.

We have presented three major criticisms of the applicability of existing rating mechanisms to AEC e-bidding: 1) reliance on input parameters that were difficult to measure 2) reliance on ad hoc operators, or 3) requiring large datasets of rating/transaction data for calibration. We will now show how a rating system based on source credibility has the potential to mitigate all three of these problems

First, source credibility theory provides tested frameworks (e.g.,(Birnhaum and Stegner, 1979)) for aggregating ratings from different sources. These frameworks decrease the dependence on ad-hoc operators. Second, there are validated scales for measuring a source's (rater's) credibility (McCroskey, 1966); these can serve as the key input parameter in a rating system based on source credibility. Finally, the weights in a rating based on source credibility theory depend on user preferences and not on rater behavior, which decreases the amount of data required to calibrate the rating application. The opportunity to measure the credibility of the rater's organization as well as of the person further decreases the amount of user input needed. Table 2 summarizes the opportunities for solving the three major problems of existing rating mechanisms, using a rating system based on source credibility.

Table 2 Summary of how a rating system based on source credibility theory mitigates major problems of alternative rating mechanisms

Alternative Methodology	Key Problem of deploying Alternative Methodology in B2B e-commerce	Opportunity for solution using source credibility based reputation mechanism
Network of Trust	Difficult to measure input parameters	Scientifically validated scales
Rule based mechanisms	Rely on Ad hoc operators for aggregating ratings	Validated aggregation functions
Data Analytical Methods	Need large amounts of clean data for calibration	1.Relying on user preferences rather than rater behavior decreases the amount of data needed for calibration. 2. Measuring credibility of the organization further decreases amount of user input needed

As a result, it is interesting to investigate how source credibility theory can support a reputation mechanism in B2B electronic commerce. The remainder of this paper first presents TrustBuilder, a prototype-rating tool, which operationalizes source credibility theory, before discussing the results of an experiment where industry practitioners deployed this tool to evaluate bids from service providers.

TrustBuilder: A rating Model Based on Source Credibility Theory

Introduction

TrustBuilder is a prototype-rating tool, guided by source credibility theory, to calculate the weights of ratings of B2B service providers. In its first version, the tool supports the specific problem of evaluating subcontractors in the construction industry. TrustBuilder employs a three-step process to help the user transform a set of ratings provided by different raters into information which supports the evaluation of subcontractors: 1) Credibility input, 2) Calculation of rater weights, and 3) Display of ratings and rater information.

Step 1: Credibility Input

TrustBuilder applies the validated McCroskey (1966) twelve-item semantic differential seven-point Likert scale to measure rater credibility. The items on this scale measure two key dimensions of a source's credibility: Authoritativeness (which corresponds to Hovland's (1953) Expertise), and Character (Trustworthiness). Based on the results of an earlier experiment (Ekstrom and Bjornsson, 2002), we also included two additional factors in our model of rater credibility. First, TrustBuilder controls for whether the rater is known to the user. Second, TrustBuilder notes whether the rater is in the same organization as the user, even if the two do not know each other. These factors are of course independent: A procurement manager may regard a competitor as being a not very reputable company, but still trust those of the competitor's employees whom he knows on a personal basis. Conversely, a manager who does not know a rater may trust a rater more if they share the same organizational affiliation. In sum, TrustBuilder uses four different factors to model user i's estimate of rater j's credibility (C_{ij}):

- **Know Rater (KR_{ij}):** Does user i know rater j? This is a binary measure entered by the user.
- **Same Organization (SO_{ij})**: Do user i and rater j work for the same organization? The model calculates this binary measure based on the two's organizational affiliation.
- **Rater Expertise (X_{ij})**: What is the expertise of rater j in the opinion of the user i? The calculation of X_{ij} is shown in Table 3 below.
- **Rater Trustworthiness (TW_{ij}):** What is the trustworthiness of rater j in the opinion of user i? The calculation of TW_{ij} is shown in Table 3 below.

While TrustBuilder models "Know Rater" and "Same Organization" using binary variables, it applies the interval McCroskey (1966) scales of Rater Expertise and Trustworthiness. Table 3 shows the scale and its operationalization in TrustBuilder.

Table 3: The McCroskey scale and its operationalization in the TrustBuilder rating tool to model rater expertise and trustworthiness.

Factor	Scale items	Operationalization:
Expertise	Reliable-Unreliable Uninformed – Informed Unqualified – Qualified Intelligent – Unintelligent Valuable – Worthless Expert – Inexpert	$X_{ij} = \sum_{k=1}^{k=6} x_{ijk}$
Trustworthiness	Honest – Dishonest Unfriendly - Friendly Pleasant - Unpleasant Selfish - Unselfish Awful - Nice Virtuous -Sinful	$TW_{ij} = \sum_{k=1}^{k=6} tw_{ijk}$

As Table 3 shows, evaluating rater expertise and trustworthiness is straightforward when the user knows the rater. However, TrustBuilder also calculates rater expertise and trustworthiness in the event that the user does not know the rater. In this case, the system asks the user to rate two types of "typical" but unknown raters:

1. "***Typical Project Manager working for Contractor X***": The user rates the expertise and trustworthiness of typical project managers working for each of the contractors which 1) the user knows, and 2) has supplied ratings to the system. This allows the system to assign a value to raters who are unknown to the user but who works for a contractor the user is familiar with.
2. "***Typical Project Manager working for a typical California contractor***": This allows the system to assign expertise and trustworthiness values to raters when both the organization and the individual are unknown to the user.

The system calculates the overall scores for all raters on the four factors KR, SO, X, and TW, before converting them into z-scores[1]. The normalization ensures that for each user, all factors will have a mean of zero and a standard deviation of one.

The next step involves converting the credibility measures into an overall value, which reflects the user's assessment of the credibility (or weight) of the different raters.

[1] z-scores measures a scale reading's distance from the mean in terms of standard deviations. In this case the mean and standard deviation were calculated for each user and scale item.

As Equation 1 shows, TrustBuilder employs an exponential function to model rater j's credibility from user's i's perspective (C_{ij}):

$$C_{ij} = \exp(-1 + b_{KR} KR_{ij} + b_{SO} SO_{ij} + b_X X_{ij} + btwTW_{ij}) \quad (1)$$

where: KR, SO, X and TW are user i's z-scores for rater i on each of the four factors; and β_{KR}, $\beta_{SO,}$ β_X and β_{TW} are coefficients associated with each factor.

Step 2: Calculation of rater weights

The next step is to estimate the coefficients of Equation 1 to calculate rater weights. TrustBuilder uses a methodology of pair-wise comparisons. Figure 1 shows a user interface where a painting subcontractor ("PaintA") has been rated by two of the seven raters (see Figure 1).

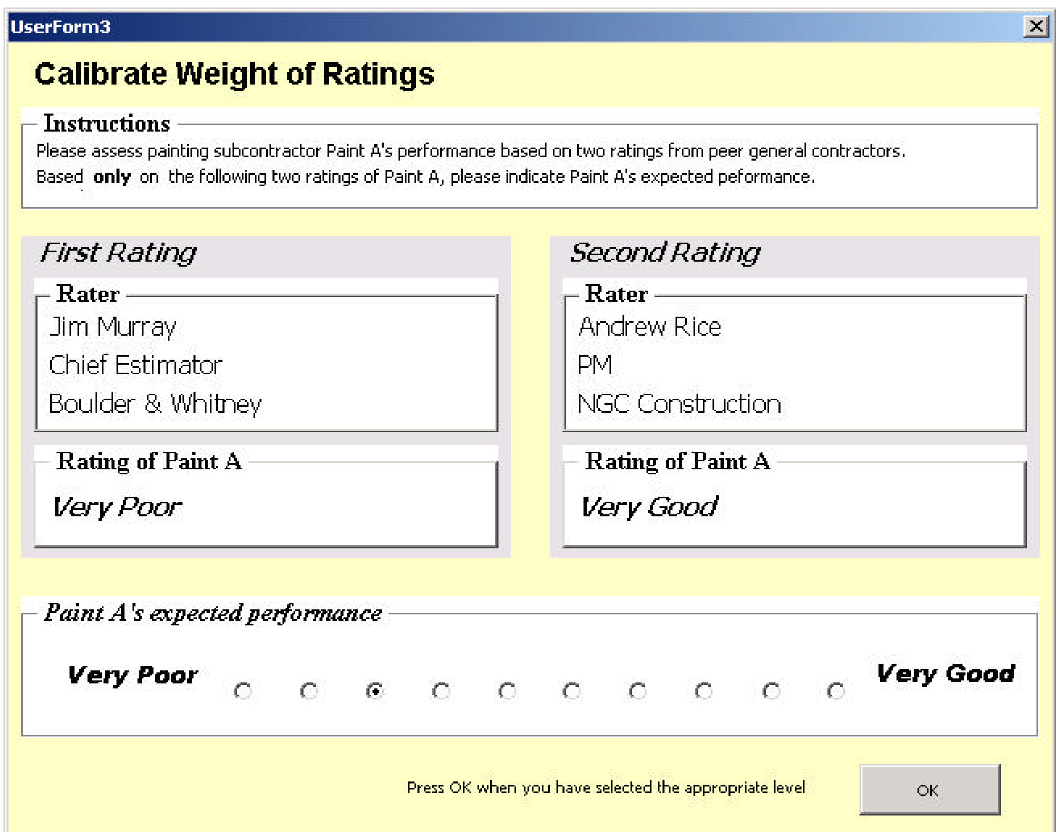

Figure 1: User interface to calibrate weight of ratings through pair-wise comparisons.

Rater 1 rated PaintA's performance as "Good" and Rater 2 rated it as "Poor". Participants submit their evaluations by clicking a 10-point Likert scale between the values "Very Poor" and "Very Good". The value (w_{12}) corresponds to the weight that the user attributes to Rater 1's ratings vis-à-vis Rater 2's. By modeling the credibility of each rater as an exponential function, we obtain the following model for w_{12}:

$$\hat{w}_{1,2} = \frac{C_1}{C_1 + C_2} \tag{2}$$

where:

$$C_{ij} = \exp(a + b_{KR} KR_{ij} + b_{SO} SO_{ij} + b_X X_{ij} + btwTW_{ij})$$

TrustBuilder can then estimate β_{KR}, $\beta_{SO,}$ β_X and β_{TW} by minimizing the sum of squares of the errors associated with all pairs (k,l) of raters included in the pair-wise comparisons.

The overall rating (R_{im}) of a subcontractor m from the user i's perspective will equal the ratings provided by each rater (j) multiplied by i's estimate of j's credibility. The result is the following straightforward formula:

$$R_{im} = \sum_j R_{jm} * C_{ij} / \sum_j C_{ij} \tag{3}$$

$$R_{jm} \neq 0$$

Step 3: Subcontractor Evaluation

TrustBuilder also displays ratings and rater information. Figure 2 shows an example of the TrustBuilder user interface, which provides the overall ratings for one subcontractor (service provider).

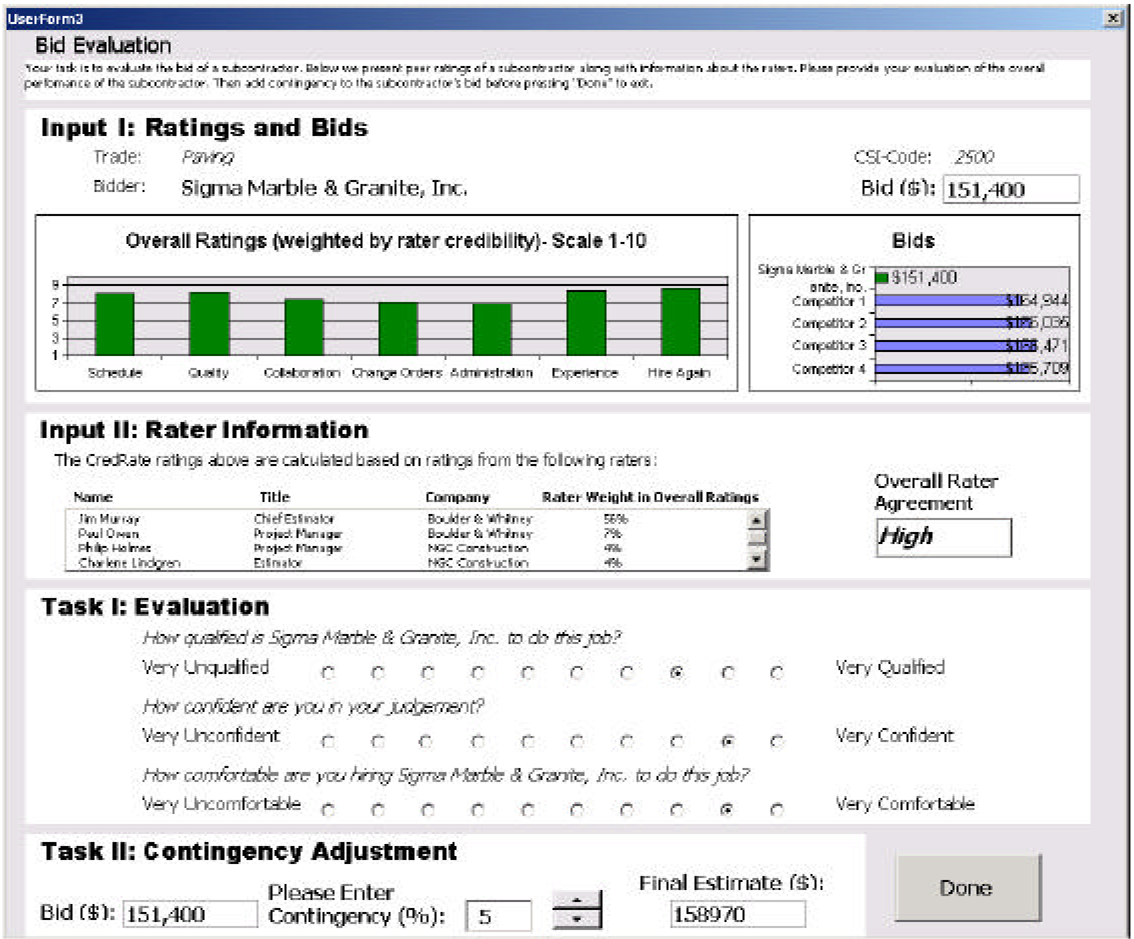

Figure 2: User interface show the ratings of a subcontractor on seven criteria.

TrustBuilder displays ratings for seven different criteria which all involve qualitative judgment: schedule, quality, collaboration, change orders, administration, experience, and hire again. Peer industry practitioners provide the ratings by indicating on ten point Likert scales the extent to which they agree with statements such as: “I would be willing to hire SubA to work for me again.” The TrustBuilder tool also shows the identity of each rater along with his/her relative weight in the overall ratings, as well as measures of rater agreement and total rater credibility.

An Experiment Comparing a Credibility-weighted Rating Model to an Unweighted Model

Introduction

We performed an experiment to test the applicability of a rating system based on source credibility in the context of industry practitioners making procurement decisions. The experiment involved actual ratings of construction industry subcontractors; all participants were industry professionals specialized in the procurement of subcontractors. The experiment compared the performance of two different rating models:

- *Credibility-weighted tool*: The TrustBuilder model
- *Unweighted Rating tool*

Hypotheses

The experiment evaluated the two rating models in terms of a set of research hypotheses. The first two hypotheses investigate the applicability of source credibility theory for calculating the weight of ratings; the other hypotheses relate to the added value of a source credibility-rating tool in an industry setting.

Hypothesis 1: The factors used in the credibility-weighted model influence rater weight.

Insignificant (i.e., close to zero) coefficients for a factor in the credibility-weighted model (Equation 1) indicate that it is not relevant to the model or is heavily correlated with other factors.

Hypothesis 2: A credibility-weighted model will better model the rater weights expressed by users in pair-wise comparisons than does an unweighted model.

These weights can be seen as the users' subjective opinion of what weights are appropriate when aggregating ratings from two different raters.

The experiment also investigated user behavior. The underlying assumption is that the more a user trusts the ratings provided by a rating tool, the more she will let them affect her decisions. This leads to:

Hypothesis 3: Users will vary their evaluations of the bidder's overall quality more when using the credibility-weighted tool than when using the unweighted tool.

A direct attitudinal measure of user trust is the users' confidence in their judgment when evaluating overall subcontractor performance. The corresponding hypothesis is:

Hypothesis 4: The use of a credibility-weighted relative to an unweighted tool results in increased user confidence in the user's judgments of overall performance.

Finally, we were interested in user's assessments of the credibility-weighted tool, viz.,

Hypothesis 5: Users will assess a credibility-weighted tool to be more useful than an unweighted tool.

Method

Fifteen construction industry professionals who worked for three California general contractors participated in the experiment. All of the participants were actively involved in evaluating bidding subcontractors, and each participant knew at least two of the other participants.

The users evaluated a set of real bids from the subcontractors that had been hired to construct a San Francisco office building in 2001. In total, the experiment involved twenty-six subcontractors bidding to perform the sixteen different trades that were subcontracted on the $3M office building.

To provide a set of ratings of the subcontractors' performance, the participants had, prior to the experiment, rated the twenty-six subcontractors on seven different criteria using ten-item Likert scales. The experiment was a within-participant design that was carried out on an individual basis using a personal computer. The participants calibrated the TrustBuilder tool by first rating a their peer raters on the McCroskey credibility scale (see Table 3), before making the pair-wise comparisons to allow the tool to calculate rater weights.

In the next step, the participant used the two tools to evaluate the subcontractors. Half of the participants used TrustBuilder (see Figure 2) to evaluate thirteen of the subcontractors before using the unweighted tool to evaluate the remaining half; the other participants used the tools in reverse order. The unweighted rating tool was a simplified version, which showed the average ratings and rater agreement along with the number of raters. In both tools, the subcontractors' low bids were roughly equal to those of the original project. The two tools also displayed the bids of four competing subcontractors. It is important to note that the name of the low bidding subcontractor had been changed to prevent the participants from recognizing the subcontractors and thus evaluate them based on previous experience. In both tools, the participants evaluated overall subcontractor quality, stated how confident they were in their judgment as well as how comfortable they were hiring the subcontractor, and adjusted the subcontractor's low bid by adding a risk buffer.

Results

Validity of Credibility-weighted Model

The results showed, consistently with Hypothesis 1, that the four factors (Know Rater, Same Organization, Trustworthiness, and Expertise) proposed in the TrustBuilder model were all significant. This conclusion was the outcome of a bootstrap analysis in which a set of fifteen users were randomly sampled (with replacement). The program running the analysis then performed the exponential regression (to estimate the coefficients in Equation 1) based on the 315 comparisons provided by the fifteen users in the sample. The program performed this procedure 2000 times to obtain statistically significant estimates. Figure 3 shows that all four factors were positive within a 95% confidence interval in the bootstrap analysis.

The current results provide evidence that the two classical factors in source credibility theory, perceived expertise and trustworthiness, contribute to the prediction of rater weights in an AEC rating application. The new factors included in this study, whether the rater knows the user and whether the user and rater are in the same organization, also influenced user assessments of credibility. These results are particularly striking given that the different factors are by nature correlated. For example, the fact that the user knows a rater increases the likelihood that the two will work for the same organization, and makes it more probable that the user will find the rater trustworthy and competent.

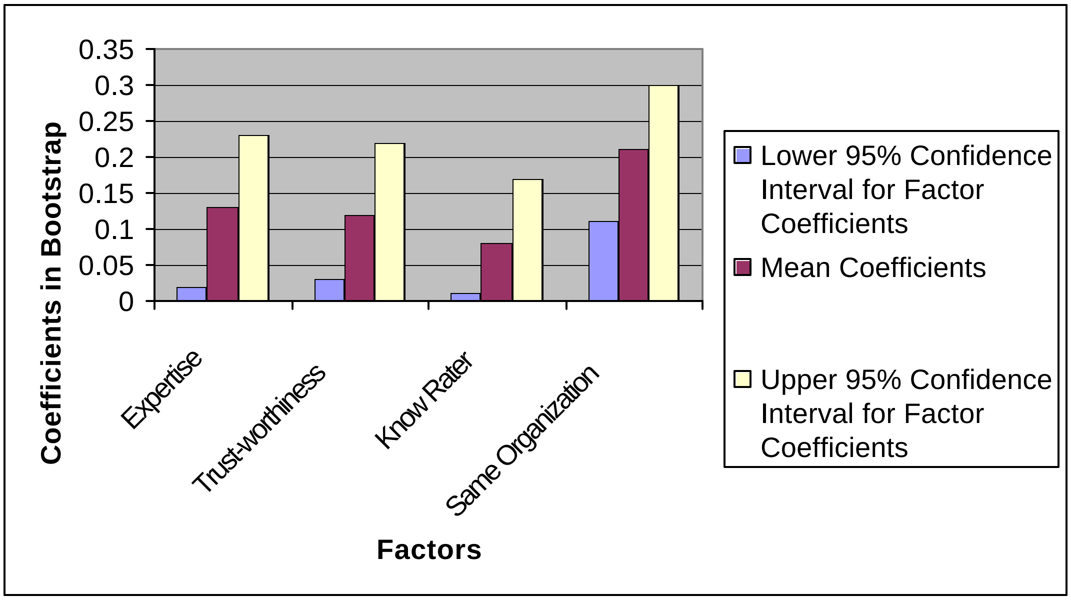

Figure 3 Results from a bootstrap analysis show coefficients of factors in exponential regression of rater weights. As shown, all coefficients are positive in the 95% confidence interval. The results show that all factors in the model (including perceived expertise and trustworthiness from source credibility theory) are significant predictors of rater weight.

Predictive Ability

The results also demonstrate that the credibility-weighted model is better at predicting rater weights (Hypothesis 2) than the unweighted model. As Figure 4 shows, the average squared error in the credibility-weighted model (0.017) is considerably smaller than in the unweighted model (0.071). A maximum likelihood ratio test, which takes account of the different degrees of freedom (60 vs. 0) of the two models, shows that this difference is significant ($p < .001$).

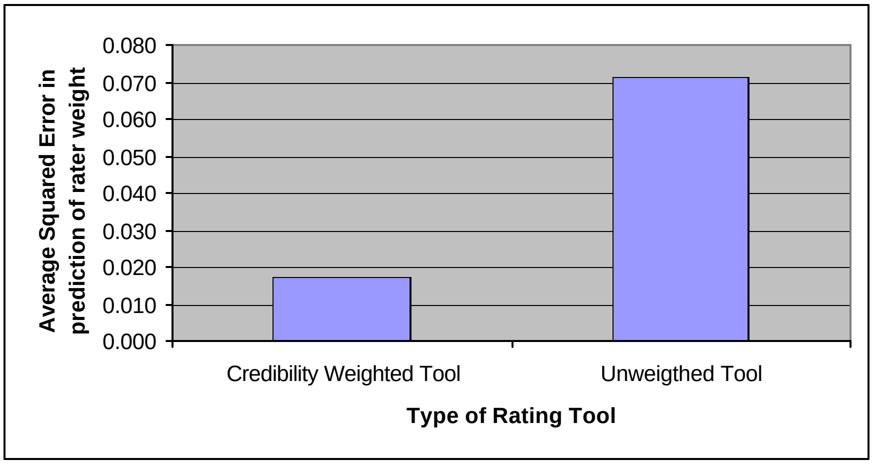

Figure 4 When predicting the users' assignments of rater weights the average errors squared errors were considerable smaller in the credibility-weighted model (0.017) than in the unweighted model (0.071).

To confirm these results, we performed a cross validation in which we first trained the model on a sample of 12 users before testing it on the remaining three users. In all of the 2000 simulated cases, the mean error was smaller for the credibility-weighted model than for the unweighted model. As a result, we conclude that the experiment provides evidence that a credibility-weighted tool is better than an unweighted (constant) model at predicting the relative weights users attribute to different raters. Furthermore, the fact that the results were also consistent with Hypothesis 1 leads us to conclude that it is possible to operationalize concepts from source credibility for ratings in a B2B electronic commerce setting.

Variance of Bid Evaluations

We also found that users varied their evaluations of overall subcontractor quality more when using the credibility-weighted tool than when using the unweighted tool (Hypothesis 3), based on a Wilcoxon matched-pairs signed-ranks test (W+ = 94.50, W- = 25.50, N = 15, p <= 0.05). This outcome suggests that the users trust the data supplied by the credibility-weighted tool more than the information supplied by the unweighted tool. This is especially the case, when it is credible peer raters who are supplying the ratings as this will lead to increased user trust. Or to quote a typical statement made by a participant during the experiment: *"The overall rating of Trojan Electric is 9 out of 10 and Chief Estimator X is among the raters. I go with the 9."*

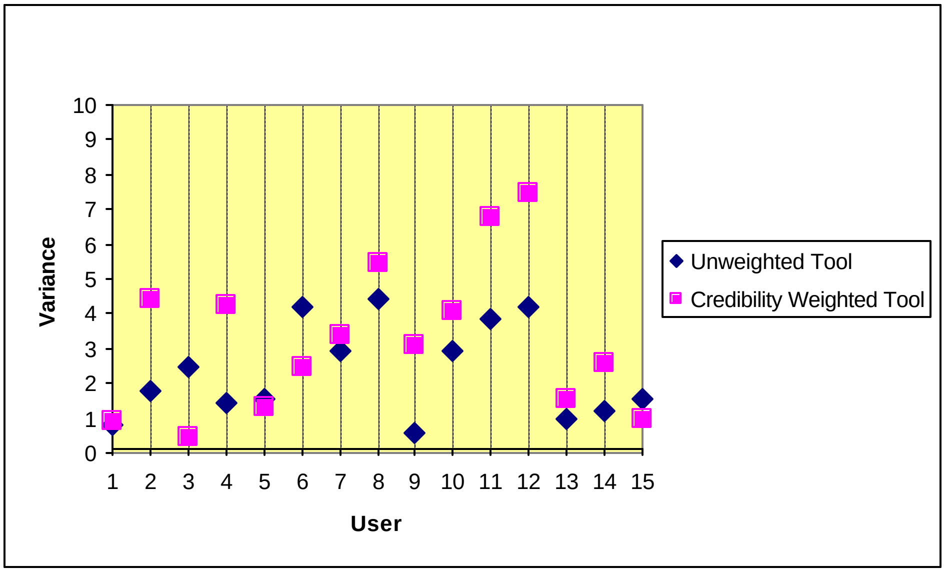

Figure 5: Variance of each user's evaluations of overall subcontractor qualification using the unweighted and credibility-weighted tools. The variance is higher for the credibility-weighted tool than for the unweighted tool ($p < .05$).

The results provide evidence that credibility information adds value when AEC practitioners evaluate subcontractors. More generally, we argue that the outcome shows that, in B2B settings, the type of rating tool (or reputation mechanism) can affect decision-maker behavior.

User Confidence in Evaluations

We also found that the use of a credibility weighted rating tool positively influenced user confidence (Hypothesis 3). This finding was the result of a linear regression of confidence in ratings, in which the factors had been normalized to z-scores. The model included the following predictors of user confidence:

- *Type of rating tool* is a binary measure, which is coded 0 for the credibility-weighted tool and 1 for the unweighted tool.
- *Overall Qualification*: The user's estimate of the overall ratings of the subcontractor.
- *Agreement* is the rater agreement displayed in the user interface.

- *Number of Raters* who had rated the subcontractor.
- *Bid Amount:* The logarithm of the bid in dollars.

The regression model generated three significant predictors of user confidence. Most notably, Figure 6 shows that the type of rating tool is a significant predictor of rater confidence. The significant negative coefficient (-0.10, p<.05) indicates that users are more confident when using the credibility weighted tool (0) than when using the unweighted tool (1).

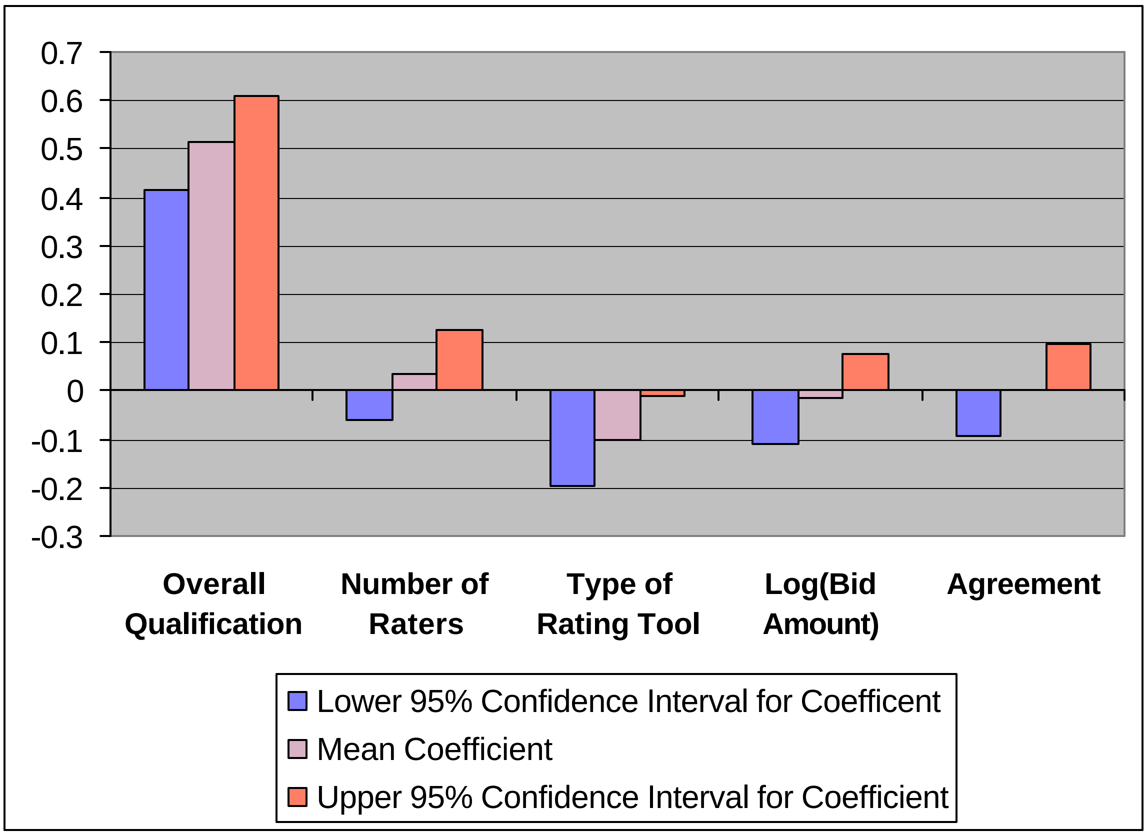

Figure 6 Coefficients in regression of user confidence. The negative coefficient (-0.10, p<.05) indicates that users are more confident when using the credibility-weighted tool (0) than when using the unweighted tool (1).

Tool Usefulness

The results for usefulness were consistent with those for the two other measures of added value. If users are confident in the ratings, and thus vary their evaluations more, they can also be expected to find the tool more useful. (Hypothesis 6.)

Originally, the credibility-weighted tool was designed to support an e-marketplace, but interviews indicated that it could also be useful in an internal rating application; we therefore investigated its usefulness in both settings.

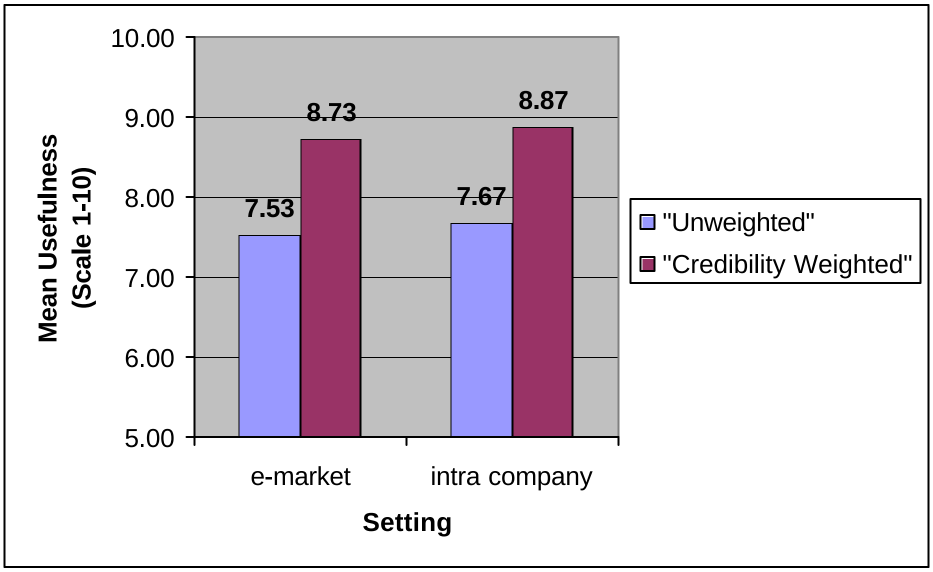

Figure 7 User's estimated the credibility-weighted tool to be more useful both in an e-market place and an intra company setting.

Figure 7 illustrates that the users results were similar for the e-marketplace and the intra-company setting. For the e-marketplace, users found the credibility-weighted tool more useful (M=8.73, SD=1.03) than the unweighted tool (M=7.53, SD=2.45, paired t(14)=2.24, $p<.05$). Also for the intra company settings, the participants estimated the usefulness to be higher for the credibility-weighted tool (M=8.87, SD=1.13) than for the unweighted tool (M=7.67, SD=2.19, paired t(14)=2.10, $p<.05$).

These results illustrate the potential use of credibility-weighted rating tools in two different settings. In an e-marketplace, knowledge is exchanged across organizations. As we would expect, users appreciate the opportunity to give different weights to users within their own and other organizations. A credibility-weighted rating system can also be deployed internally, i.e., within the organization of a large contractor. The results show that the participants also found it useful to differentiate between different types of users within their own organization, for whom both skill level and trustworthiness can vary.

Finally, the consistent results across Hypotheses 4, 5, and 6 provide evidence that source credibility information can add value in a B2B electronic commerce setting such as the evaluation of construction industry subcontractors.

Discussion

The findings of this study have important implications for the construction industry, as well as the research community. In contrast to previous construction management research, we have focused enabling the information sharing of qualitative information about subcontractor performance. We have showed that the presented methodology enables the integration of subjective information from multiple AEC practitioners of varying reliability. More specifically, the experiments provided evidence that it is possible to operationalize dimensions of source credibility when calculating rater weights.

This study also provides evidence that a credibility-weighted rating mechanism adds value in a practical AEC electronic commerce setting. We argue that the strength of a rating tool operationalizing source credibility theory is that it 1) incorporates tested frameworks for aggregating information, 2) applies validated scales for measuring the input parameters, and 3) does not require large amounts of data for calibration.

As we have shown, the TrustBuilder model shares common characteristics with both rating tools in the general electronic commerce literature, as well as decision support tools which focus on the construction industry processes. While theoretical models of both types abound, researchers have given little attention to how these can support the end user's decisions. This research project shows that it is possible to investigate the added value of rating mechanisms from a user perspective. The need for further studies of this type is emphasized in AEC electronic commerce settings, which involve skilled users, as well as high stakes. The experiment provides evidence that decision-makers' evaluations will vary depending on the rating tool they are using, but did not consider the extent to which rating mechanisms can affect the final purchasing decision. A complicating factor is that in practice, information from outside a rating mechanism, such as credit ratings and experience, will be available to support the decisions. At what point will the output of a rating system cause the construction industry estimator to determine that the low bidding subcontractor is not the best bidder? Designers of reputation mechanisms and other decision support tools, should keep in mind that the output of a model will only add value if it actually changes the user's decisions.

Another important implication is that in AEC electronic commerce, mechanisms which support trust should not be developed in isolation of the existing business and intra-personal relationships. In highly connected communities, a reputation mechanism should take into account the relationship between the user and the rater when aggregating information. Grounding a reputation mechanism in source credibility theory is a technological solution which builds upon, rather than replaces, the existing networks of trust between human actors. A credibility-weighted rating tool enables the industry practitioner to utilize her experience, judgment and relationships to enhance the information sharing, which the Internet facilitates. The user can take advantage of knowledge, which is pooled across the industry, while being confident that the opinions of persons that he or she knows and trusts will be given added weight.

The experiment also illustrates the need for customized rating tools, showing that each user aggregates information in a unique way. For example, some users regarded rater expertise as imperative, while others considered the identity of the rater to be of minor importance. Furthermore, a rating tool could also integrate additional factors, such as time and risk averseness, which further increases the necessity to account for the decision-maker's individual rationale.

An interesting avenue for further research would be to combine a rating filter based on source credibility with data analytical methods. A rating mechanism could evaluate raters who the user knows based on source credibility, while applying collaborative filtering or statistical methods to differentiate between unknown raters. Our intention is also to make researchers and practitioners aware of the opportunity for incorporating frameworks developed in social science in technologies which support human interaction across the Internet. There is no reason that the substantial body of research studying areas, such as trust, credibility, and reputation should not be valid also in online situations.

Another interesting avenue for future research is the integration of the aggregate ratings provided by the TrustBuilder credibility model with other decision support tools (e.g., (Russell et al., 1992, Sik-Wah Fong and Kit-Yung Choi, 1999, Tseng and Lin, 2002, Wanous et al., 1999)) which mainly account for more quantative information. The end user, who is deciding which subcontractor to hire, will want to know, not only a bidder's weighted ratings on criteria such as maintenance of schedule and quality of work, but also its credit rating, project experience, and bond rates. There is a risk that presenting that presenting this amount of information at detailed level would inundating the user, but at the same time she would be unlikely to let computerized tool make the final decision. As a result, an interesting research issue is determining at what level should information be aggregated to best support decision-making in AEC bidding.

The fact that the users found the credibility-weighted rating tool useful, both in an e-market place setting and as an internal rating application, has practical as well as theoretical implications. The capability of integrating ratings from external and internal raters, opens for a potential trajectory towards the adoption of rating mechanisms in AEC electronic commerce. Large contractors could initially deploy the rating tool internally, before, at a later stage, increasing the underlying knowledge base by opening for information sharing with other contractors. By rating the rater, industry practitioners can move one step closer towards leveraging the power Internet technologies to increase the efficiency of AEC business transactions.

Bibliography

1. Abdul-Rahman, A. and Hailes, S. (2000) In *Hawaii's International Conference on Systems Sciences* IEEE, Maui, Hawaii.
2. Avery, C., Reznick, P., Zeckhauser, R. (1999) *American Economic Review,* **89,** 564-584.
3. Birnhaum, M. H. and Stegner, S. E. (1979) *Journal of Personality and Social Psychology,* **37,** 48-74.
4. Chen, M. and Singh, J. P. (2001) In *ACM Conference on Electronic Commerce (EC'01)*(Ed, Shoham, Y.) Tampa.
5. Chinyio, E. A., Olomolaiye, P. O., Kometa, S. T. and Harris, F. C. (1998) *Construction Management and Economics,* **16**.

6. Dellarocas, C. (2000) In *Proceedings of the 2nd ACM Conference on Electronic Commerce* Minneapolis, Mn.
7. Ekstrom, M. and Bjornsson, H. (2002) In *CIB W65 Symposium September 2002*Cincinnati, Ohio, USA.
8. Fogg, B. and Tseng, H. (1999) In *CHI 99*ACM, Pittsburgh.
9. Hanson, W. A. (1999) *Principles of Internet Marketing,* South-Western College Publishing, Cincinnati.
10. Holt, G. D. (1996) *Building Research and Information,* **25**.
11. Holt, G. D., Olomolaiye, P. O. and Harris, F. C. (1995) *Building and Environment,* **29**.
12. Hovland, C. I., Janis, I. L. and Kelley, H. H. (1953) *Communication and Persuasion,* Yale University Press, New Haven.
13. Lucking-Reiley, D. and Spulber, D. F. (2000) .
14. Malone, T. and Laubacher, R. (1998) *Harvard Business Review,* **,** 145-152.
15. McCroskey, J. C. (1966) *Speech Monographs,* **33,** 65-72.
16. Newhagen, J. and Nass, C. (1989) *Journalisms Quarterly,* **66,** 277-284.
17. Okoroth, M. I. and Torrance, W. B. (1998) *Construction Management and Economics,* **17,** 315-327.
18. Palaneeswaran, E. and Kumaraswamy, M. M. (2000) *Construction Engineering and Management,* **126**.
19. Ratnasingham, P. and Kulmar, K. (2000) In *International Conference on Information Systems*
20. *(ICIS-2000)*Brisbane, Australia.
21. Reznick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994) In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*.
22. Reznick, P. and Zeckhauser, R. (2001) *Journal of Industrial Economics,* Fourthcoming.
23. Russel, J. S., Hancher, D. E. and Skibniewski, M. J. (1992) *Construction Management and Economics,* **10,** 117-135.

24. Russell, J. S., Hancher, D. E. and Skibniewski, M. J. (1992) *Construction Management and Economics*, **10**.
25. Shardanand, U. and Maes, P. (1995) In *ACM Special Interest Group on Computer-Human Interaction*, Vol. 37 ACM Press/Addison-Wesley Publishing Co, .
26. Sik-Wah Fong, P. and Kit-Yung Choi, S. (1999) *Construction Management and Economics*, **18,** 547-557.
27. Trepper (2000) *E-commerce Strategies*, Microsoft Press.
28. Tseng, H. P. and Lin, P. H. (2002) *Automation in Construction*, **11,** 105-125.
29. Wanous, M., Boussabaine, A. H. and Lewis, J. (1999) *Construction Management and Economics*, **18**.
30. Zacharia, G., Moukas, A. and Maes, P. (1999) In *Thirty-second Annual Hawaii International Conference on*
31. *System Sciences (HICSS-32)* Wailea, Hawaii.

[i] www.epinions.com