

## RESEARCH

## Open Access

## Güven: estimating trust from communications

Anup K. Kalia<sup>1\*</sup>, Zhe Zhang<sup>2</sup> and Munindar P. Singh<sup>1</sup>

\*Correspondence:

akkalia@ncsu.edu

<sup>1</sup>Department of Computer Science,  
North Carolina State University,  
27695 Raleigh, USAFull list of author information is  
available at the end of the article**Abstract**

The extent to which an agent trusts another naturally depends on the outcomes of their interactions. Previous computational approaches have treated the outcomes in a domain-specific way. Specifically, these approaches focus on the mathematical aspect and assume that a positive or negative experience can be identified without showing how to ground the experiences in real-world interactions, such as emails and chats. We propose *Güven*, an approach that relates trust to the domain-independent notion of commitments. We consider commitments since commitment outcomes can be associated with experiences and a large body of works exist on commitments that include commitment representation and semantics. Also, recent research shows that commitments can be extracted from interactions, such as emails and chats. Thus, we posit *Güven* can provide an useful basis to infer trust between agents from their interactions.

To evaluate *Güven*, we conducted empirical studies of two decision contexts. First, subjects read emails extracted from the Enron dataset (and augmented with some synthetic emails for completeness), and estimated trust between each pair of communicating agents. Second, the subjects played the Colored Trails game, estimating trust in their opponents. *Güven* incorporates a probabilistic model for trust based on commitment outcomes; we show how to train its parameters for each subject based on the subject's assessments. The results are promising, though imperfect. Our main contribution is to launch a research program into computing trust based on a semantically well-founded account of interpersonal interactions.

**Keywords:** Multiagent systems, Commitments, Trust**Introduction**

We consider multiagent system settings where agents interact with each other. A multi-agent system is an open system consisting of autonomous and heterogeneous parties or agents. By autonomy, we mean that agents can act independently. And, by heterogeneity, we mean agents have diverse internal representations, including goals and internal policies. We consider a multiagent system to be open: agents may potentially enter such a system interact with others, and leave the system. Real-world examples of such systems arise in the corporate and military sectors where agents collaborate with each other in teams. In such systems, based on their mutual interactions, an agent as a *truster* estimates (and continually revises) its trust for another agent as a *trustee*. For example, in a corporate setting, an employer (truster) can assign a task to an employee (trustee). If the employee performs the task, the employer's trust increases for its employee. Similarly, in

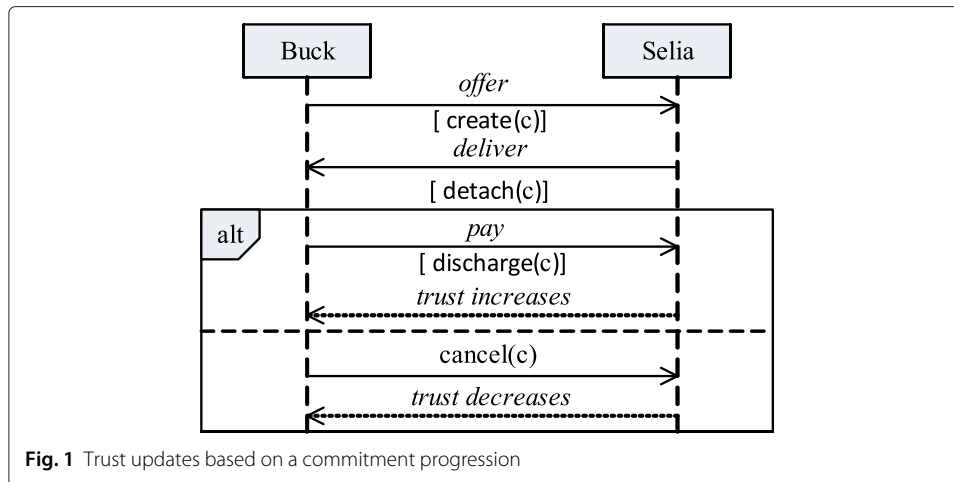
the military, a commander (truster) can ask a subordinate (trustee) to destroy a particular target. If the subordinate success, the trust of the commander toward the subordinate presumably increases.

Understanding such interactions between agents and estimating trust from them is an interesting and challenging topic. Scissors et al. [1] exploit linguistic similarity in chat messages to estimate trust between message senders and receivers. Adalı et al. [2] calculate relationship strength between two users in Twitter based on social and behavioral aspects such as their numbers of friends and followers, the number of messages they exchange, and the time delay between the messages exchanged. DuBois et al. [3] provide an algorithm to compute trust and distrust in a social network. Wang et al. [4] combine positive and negative evidence to estimate trust. Teacy et al. [5] formulate trust as the count of fulfilling or violating obligations. Jøsang [6] represents trust as the belief measure of a truster that the trustee will cooperate. The above approaches are promising but they are limited to numerical heuristics. Such approaches have been justifiably criticized by richer approaches [7–9] for missing the essential intuitive considerations of trust, e.g., regarding the autonomy of the participants and the vulnerability of the truster to decisions by the trustee. The truster's vulnerability refers to his or her willingness to take a risk on the trustee with the expectation that the trustee will perform the task promised to the truster [10]. However, the richer approaches have limitations since although they are formally represented, they do not lend themselves to computational techniques that could be applied in practice.

We seek to bridge the above gap between theory and practice. Specifically, we propose *Güven*,<sup>1</sup> a computational model of trust founded on commitments that supports how agents determine trust in others based on their interactions. Commitments are important for trust because they can be identified from interpersonal interactions and can help us characterize the outcomes of such interactions in high-level terms. We limit *Güven* to commitments, although it can be potentially extended to related concepts such as prohibitions and authorizations.

A commitment  $C(\text{debtor}, \text{creditor}, \text{antecedent}, \text{consequent})$  means that the debtor commits to bringing about the consequent for the creditor provided the antecedent holds. For example,  $C(\text{Buck}, \text{Selia}, \text{deliver}, \text{pay})$  means that Buck (buyer) commits to Selia (seller) to paying a specified amount provided Selia delivers the goods. When Selia delivers, the commitment is detached. When Buck pays, the commitment is discharged or satisfied. If Selia delivers but Buck does not pay, the commitment is violated. In essence, a commitment describes a social relationship between two persons giving a high-level description of what one expects of the other. As a result, it is natural that commitments (and their satisfaction or violation) be useful as a basis for trust. In the above example, if Buck discharges the commitment, he brings a positive experience to Selia and Selia's trust for Buck may increase; if Buck violates the commitment, he brings a negative experience to Selia and Selia's trust for Buck may decrease. Figure 1 illustrates the intuition graphically.

Despite the apparent match, few approaches relate trust and commitments. Singh [9] and Chopra et al. [8] relate trust and commitments in terms of logical postulates and from an architectural perspective. In contrast, we understand trust and commitment in probabilistic terms, considering the outcomes of specific commitments and their effect on the trust relationships between the concerned parties.



We conduct two empirical evaluations, respectively, on emails (automatically analyzed using Kalia et al.'s approach [11]) and via the Colored Trails cooperation game [12] (analyzed manually). We show how to train the model parameters so as to capture a user model indicating each user's propensity to trust given commitment outcomes. Our evaluations yield promising, though imperfect, results on the viability of inferring trust from the commitments arising in interactions, suggesting the need for better extraction techniques. Our main contribution is to show how trust can be computed via the domain-independent concept of commitments. The contribution takes a step further to bring existing theories into practice.

This paper is organized as follows. We begin with a review of the related work. Next, we provide essential background on commitments and intuitions regarding how commitments affect trust. Then, we describe Güven as an evidence-based approach for updating trust based on commitments along with the requisite computational methods. We present our hypotheses in informal terms along with a strategy for evaluating them, followed by our evaluation and results. We conclude with a discussion and future directions.

**Related work**

Teacy et al. [5] provide a trust model based on fulfilling or violating obligations. In their model, the trust of a truster ( $a_{tr}$ ) toward a trustee ( $a_{te}$ ) is represented as the expected probability that the trustee will fulfill its obligations ( $B_{a_{tr},a_{te}}$ ) given the outcome of all interactions ( $O_{a_{tr},a_{te}}^{1:t}$ ) between the truster and the trustee. In contrast to their work, we consider representing trust as the basis of commitments (instead of obligations). We consider commitments since there has been an extensive work on capturing and formalizing the semantics of commitments in multiagent interactions [13, 14]. In addition, the lifecycle of commitments [15, 16] describe how commitments are created and progressed in multiagent interactions. Further, we evaluate our models on an email and a game dataset proving that such models are applicable to real-world settings.

Wang et al. [4] define trust ( $\alpha$ ) as the ratio of positive outcomes experienced ( $r$ ) by the truster from the trustee to the total number of positive and negative outcomes( $s$ ), i.e.,  $\alpha = \frac{r}{r+s}$ . Further, to denote the certainty of the truster toward a trustee, Wang et al. [4] define a certainty function ( $c(r, s)$ ) that employs a beta distribution and takes  $r$  and  $s$  as its input parameters. There are important limitations with Wang et al. [4] contributions:

(1) the model is purely mathematical and hence, does not clearly describe how it can be applied to real-life interactions such as emails and chats. Specifically, the definitions of positive and negative outcomes are not formally captured with respect to agents' interactions; (2) the model needs the initial trust to be manually set. That is, it assumes fixed values. Specifically, Wang et al. mention that the initial trust is set based on the truster's prior experience with the trustee. Compared to their model, we clearly define positive and negative evidence as outcomes of commitments thereby extending their model to be applicable to real-world settings. Further, our model learns the initial trust from users' data instead of assuming fixed values.

Osman et al. [17] describe a model that estimates the trust of a truster based on the trustee's capability and willingness to execute a commitment. The willingness of a trustee to execute the commitment is computed using the trustee's past behavior in executing similar commitments. The capability of the trustee is computed by matching the capability needed for the current commitment with the capabilities of the trustee observed in the past. Osman et al.'s model suffers from two important limitations: First, the model considers commitments as a set of actions to be performed by the trustee, thus omitting a formal representation as well as an operationalization (lifecycle) of commitments. This makes the model less intuitive with respect to its applicability on real-world data such as emails. Second, similar to Wang et al.'s [4] model, Osman et al.'s trust model needs the initial trust to be manually set. For example, the model considers the initial trust as zero which may not hold in different settings.

Kastidou et al. [18] describe a trust model based on promises made and delivered by a trustee toward a truster. Similar to Osman et al. [17] model, Kastidou et al.'s model does not provide the semantics of promises, requires manual setting of the initial trust values, and does not consider real-world datasets for the evaluation.

Burnett and Oren [19] examine the effects of delegation using a probabilistic trust model [20] and propose an approach for weighting trust updates based on shared responsibility or delegation. Burnett and Oren do not restrict the delegation chain length. In contrast, we restrict our trust update to delegation chains of length three (the debtor, the new debtor, and the creditor). For example, if the new debtor delegates the commitment to another debtor (debtor'), trust between debtor and debtor' remains unaffected. This means if debtor' satisfies or violates the commitment, the trust of the debtor toward debtor' neither increases or decreases, respectively, since the debtor still depends on the new debtor to satisfy the commitment and may be ignorant of debtor'. We observe that such longer chains are rare in a real-world text corpus.

Adali et al. [2] correlate textual features (linguistics, processes, and psychological processes) with social and behavioral features (reciprocity, assortativity, attention, and latency). Textual features are derived based on the content of messages exchanged between users whereas social and behavioral features are computed based on user's social network (nodes, edges). Both the textual and behavioral features are indicator of trust behavior between users. However, such measures are considered based on their frequency of occurrence and they do not capture the vulnerability of a truster toward a trustee. For example, textual features indicate number of verbs, pronouns, affective processes (emotions), cognitive processes (causation, certainty), perceptual processes (see, hear, feel), and so on whereas behavioral features indicate number of number of friends, followers, messages sent, degree similarity between users, reciprocity of conversation and

propagation messages, in-degree, out-degree and so on. Thus, we emphasize capturing semantic meanings of messages in terms of commitments (i.e., how they are created, satisfied, and violated) from the text content exchanged between users.

Scissors et al. [1] performed an empirical evaluation with 62 students and found that different forms of linguistic similarity, such as content (e.g., positive emotion words, task-related words), structure (e.g., verb tense, phrasal entrainment), and style (e.g., chat abbreviations, emoticons) reflect different level of trust between participants. In contrast, we consider commitments between participants to estimate trust between them.

**Background on commitments**

Figure 2 presents the commitment lifecycle we adopt. A commitment  $C(\text{debtor}, \text{creditor}, \text{antecedent}, \text{consequent})$  is *created* when a debtor either voluntarily creates it (*commissive creation*) or is directed to do a certain task (*directive creation*). Given the debtor’s autonomy, the latter presumes a prior commitment on part of the debtor. A commitment is *detached* if a condition or an antecedent present for a commitment holds true; *discharged* when a debtor executes a committed task. A commitment is *terminated* when a debtor cancels the commitment before it is detached or a creditor releases the commitment. A commitment is *violated* when a debtor cancels the commitment after it is detached or when a consequent timeout occurs. Additionally, a commitment can be delegated and assigned. A commitment is *delegated* when the debtor of a commitment is replaced by a new debtor and *assigned* when the creditor of a commitment is replaced by a new creditor. We map interactions between persons to commitment operations.

Consider a commitment  $C(\text{Alice}, \text{Bob}, \text{pay}, \text{ship goods})$  where Alice (buyer) commits to Bob (seller) to ship goods provided Bob pays. When Alice offers Bob to ship the goods, C gets created from Alice toward Bob and becomes conditional. When Bob makes the payments, C gets detached. If Alice ships the goods, C gets discharged. If Alice does not ship the goods despite the payment, C gets violated. If C is conditional and Alice cancels C, C gets terminated. Alice can delegate C to a new agent Charlie to create commitment  $C(\text{Charlie}, \text{Bob}, \text{pay}, \text{ship goods})$  where Charlie becomes the new debtor. Similarly, Bob can assign commitment C to a new creditor John to create  $C(\text{Charlie}, \text{John}, \text{pay}, \text{ship goods})$ . In case of delegation and assignment, when a new commitment are created, the older commitment remains suspended and its outcome depend upon the outcome of the new commitment. If the new commitment is satisfied, so

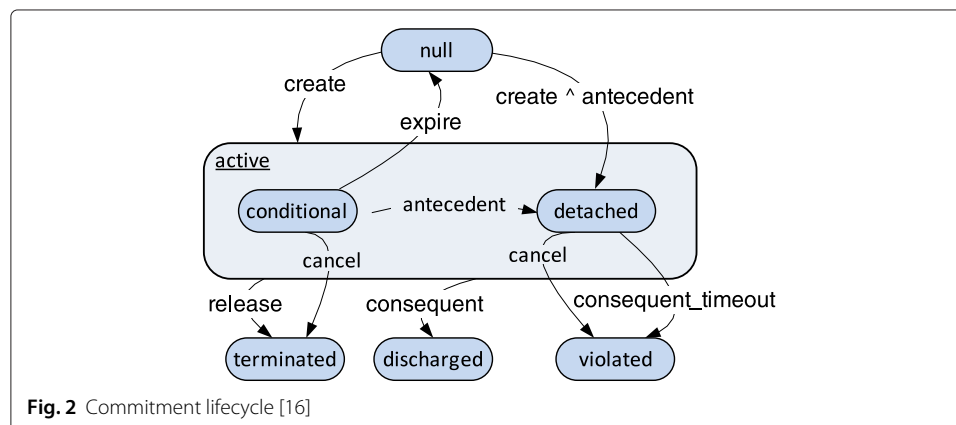


Fig. 2 Commitment lifecycle [16]

is the older commitment. If the new commitment is violated, the older commitment becomes active (conditional or detached, depending upon the truth of its antecedent) again.

### **Intuitions on trust and commitments**

We describe some criteria for how trust values may be updated based on operations on commitments.

#### **Effects of commitment operations on trust**

We describe the effect of commitment operations on trust. Before we describe the effects, let us consider some situations wherein a commitment exists from a debtor toward a creditor.

**Effect of discharge.** When a commitment is discharged, the creditor's trust for the debtor increases.

**Effect of violation.** When a commitment is violated, the creditor's trust in the debtor decreases.

**Effect of delegation and discharge.** When a commitment is delegated by the original debtor to a new debtor, and the new debtor satisfies it, the creditor's trust in both the original and the new debtor increases.

**Effect of delegation and violation.** When a commitment is delegated from the original debtor to a new debtor and the new debtor violates it, the creditor's trust in both the original and the new debtor decreases.

**Effect of assignment and discharge.** When the commitment is assigned from the original creditor to a new creditor and the original debtor discharges it, the trust of the original and new creditor for the original debtor increases.

**Effect of assignment and violation.** When the commitment is assigned from the original creditor to the new creditor and the debtor violates it, the trust of the original and the new creditor for the original debtor decreases.

We make the following assumptions regarding the increase or decrease of trust. These are simplified assumptions and could be relaxed in some settings.

- In our basic approach (baseline), the change in trust is the same for all commitment outcomes. We additionally provide an approach in which the change in trust depends upon the strength of a commitment. For example, a commitment with a strict deadline when satisfied may produce a different level of trust compared to a commitment with a flexible deadline.
- We assume commitment discharge and violation to be all or none; in our scenarios, partial success is not easy to infer.
- In case of violation, the trust of the creditor for the debtor decreases irrespective of whether the debtor was truly responsible. An agent's beliefs and goals are private and cannot be identified directly from his or her interactions.
- In case of delegation, the original creditor's trust in the original and the new debtor changes equally, reflecting the idea that the creditor has a positive experience thanks to the two debtors.

In case of a situation where the new debtor violates a commitment and the original debtor has nothing to do with it, it was still the responsibility of the original debtor to satisfy the commitment. (This is one of the patterns of delegation identified by Singh

et al. [21]). Thus, the creditor’s trust will decrease equally for both the new and the original debtor.

- In case of assignment, the new creditor’s trust in the debtor changes as much as the original creditor’s trust in the debtor, reflecting the intuition that both creditors’ expectations are met.

**Subjectivity, memory, and strength**

Trust is modulated by features that affect how trusters judge outcomes, such as the satisfaction or violation of a commitment. First, trust assessment is subjective. Trusters differ in how they reward or penalize a trustee when a commitment is discharged or violated, respectively. Second, trust assessment depends on the truster’s memory: trusters with limited memory would tend to forget all but (some varying number of) recent experiences. Recent experiences may turn out to be more predictive of future experiences (that trust is about) than past experiences. Third, the effect on trust of a commitment’s outcome would be greater when the commitment is more important.

**Güven: model of trust based on commitments**

We adopt Wang and Singh’s [22] trust model, which represents trust as evidence  $\langle r, s \rangle$ . Here,  $r \geq 0$  and  $s \geq 0$  respectively represent the positive and negative experiences the truster has with the trustee. Both  $r$  and  $s$  are real numbers. Wang and Singh calculate trust as the probability of a positive outcome as  $\alpha = \frac{r}{r+s}$ . Suppose Buck and Selia transact 10 times and exactly eight transactions succeed from Selia’s perspective. Then Selia’s trust in Buck would be 0.8.

The basic idea is for each truster to maintain evidence  $\langle r, s \rangle$  about each trustee. The initial evidence,  $\langle r_{in}, s_{in} \rangle$ , represents the truster’s bias. An interaction may yield a positive, negative, or a neutral experience. In these cases, the evidence is updated by respectively adding  $\langle i_r, 0 \rangle$ ,  $\langle 0, i_s \rangle$ , and  $\langle \lambda i_r, (1 - \lambda) i_s \rangle$ , where  $\lambda \in [0, 1]$ . In essence, we characterize each truster via five parameters  $(i_r, i_s, r_{in}, s_{in}, \lambda)$ .

**Considering subjectivity**

To evaluate  $H_1$ , we learn a specific truster’s parameters based on positive, negative, and neutral experiences it acquires from trustees and the truster’s actual trust in various trustees. For the  $k^{th}$  trustee, let  $\alpha_k$  represent the truster’s actual trust (as revealed) and  $\hat{\alpha}_k$  the truster’s predicted trust in  $k$ . Let  $E_k^+$ ,  $E_k^-$ , and  $E_k$  represent the numbers of positive, negative, and neutral experiences, respectively. Then,

$$\hat{\alpha}_k = \frac{r_{in} + E_k^+ i_r + \lambda \cdot E_k i_r}{r_{in} + s_{in} + E_k^+ i_r + E_k^- i_s + E_k (\lambda i_r + (1 - \lambda) i_s)} \tag{1}$$

Via nonlinear least-squares regression technique that uses trust region reflective algorithm [23], we estimate the truster’s parameters to minimize the mean absolute error (MAE) of prediction,  $\sum_{k=1}^n |\hat{\alpha}_k - \alpha_k|$ .

**Considering memory**

We capture the effect of memory by considering a *discount window* [24], defined simply as the most recent  $W$  experiences. Let  $n$  be the total number of experiences the truster

acquires from the trustee. Let  $t = \min(n, W)$ . Let  $E_t^+$ ,  $E_t$ , and  $E_t^-$  be the positive, neutral, and negative experiences inferred from the  $t$  transactions. The trust of a truster in the trustee depends on whether  $t$  is less than  $W$ . When  $t < W$ , the truster's trust is  $(E_t^+ i_r + \lambda E_t i_r, E_t^- i_s + (1 - \lambda) E_t i_s)$ ; otherwise, it is  $(r_{in} + E_t^+ i_r + \lambda E_t i_r, s_{in} + E_t^- i_s + (1 - \lambda) E_t i_s)$ . When  $t < W$  we ignore the initial bias as the truster's trust is based on recent  $W$  experiences, which simply means that the truster has already forgotten its initial bias.

### Considering strength

We posit that a truster acquires experiences of varying *weights* based on commitment outcomes (satisfied or violated). To calculate the weight, we identify the following features in a sentence indicating a commitment creation.

**Commissive over directive.** *A commissive (e.g., "I will ...") may carry a greater weight than a directive (e.g., "Could you please ...") because it holds even without the presumption of another commitment.*

**Debtor's type.** *A single debtor may carry a greater weight than multiple debtors "we" ("We will follow up"). A single debtor, as in "I will follow up," has clearer responsibility than multiple debtors.*

**Creditor's type.** *Multiple creditors may carry a greater weight than a single creditor.* Multiple creditors arise when a debtor commits to a set, e.g., when a product manager commits to his employees to review a product. The intuition is that having multiple creditors makes the debtor accountable to more parties.

**Modal verbs.** *Some modal verbs (e.g., will or shall) may convey high confidence over others (e.g., can, could, may, would) [25].* The intuition is that "will" indicates that a commitment will be surely satisfied whereas "can" indicates that the commitment may not be satisfied. We learn the weights of different modal verbs based on data obtained from human subjects.

**Action verbs.** *Some action verbs convey a greater level of importance than others. For example, "resolving an issue" may be more important than "reviewing a proposal."* We compute the weights of verbs using Burchardt et al.'s [26] *FrameNet* tool, which provides weights for words used in different senses, e.g., 1 for *resolve* and 0.383 for *review*.

**Deadlines.** *Noun phrases with deadlines [27] may convey more importance than noun phrases without deadlines.* For example, an explicit deadline, as in "I will repair the car by Monday," enhances the importance of the commitment. We assume that merely mentioning a deadline increases the seriousness of a commitment. We defer to future research additional subtleties, such as the duration or urgency of a deadline and the extent to which it may be broken, since in our empirical settings durations and urgency do not arise.

Except the feature *action verb*, we evaluate rest of the features empirically from a subject evaluation. We provide the outcome of our evaluation in Table 6. Our evaluation ranks feature values as discussed above. We map the ranks to weights (cardinal numbers, higher for higher ranked features) and sum the weights to compute a value. For example, consider two commitment creations from a trustee toward a truster: (1) "I will repair the car by Monday" and (2) "I can repair the car". These examples have four common features: the commitment type (commissive), the debtor type ("I"), the creditor type (single), and the action verb ("repair") and two different features: modal verbs ("will"



versus “can”) and deadlines (“Monday” versus no deadline). Thus, based on the features from these two examples and the rank of those features provided in Table 6, we compute their weights as shown in Table 1. The weights computed (11 versus 8) indicate that the truster might have a different experience from the outcome of the first example than the second example.

**Evaluation strategy**

We consider two decision contexts for evaluations. The first context involves subjects (as bystanders) reading emails exchanged between other agents and assessing the levels of trust between these agents. The second context involves the subjects playing a game with each other. The game has some cooperative and some competitive elements. The subjects (as interested participants) assess the trustworthiness of their opponents.

We now present our research hypotheses in informal terms. These hypotheses are based on subjectivity, memory, and strength, as proposed in Section ‘Güven: model of trust based on commitments’ to compute trust. Thus, these hypotheses motivate our evaluation strategy and study design. The following section on evaluation refines these hypotheses into technical claims.

*H<sub>1</sub> (Subjectivity)* Predicting trust values by *learning trust parameters* for each subject yields more accurate results than using *fixed trust parameters* for all subjects.

The details of the trust parameters are given in Section ‘Güven: model of trust based on commitments’. Assuming *H<sub>1</sub>* holds, we consider learned parameters as the *baseline* approach. We check if other approaches improve accuracy beyond the baseline.

We consider *H<sub>1</sub>* since trust models [4, 5, 17, 18] described in Section ‘Related work’ compute trust considering fixed parameters. That is, they consider fixed values for  $\langle r, s, r_{in}, s_{in}, \lambda \rangle$ . In contrast, we compute trust by learning these parameters from subjects’ assigned level of trust. Thus, we posit that learning the trust parameters would improve the trust prediction accuracy.

*H<sub>2</sub> (Memory)* Predicting trust values by learning a specific *discount window size* for each subject yields more accurate results than the baseline.

We consider *H<sub>2</sub>* since a subject might assign a trust level based on his or her most recent experiences [24]. Thus, we posit that learning a specific discount window size in addition to learning trust parameters for each subject can improve the trust prediction accuracy more than the baseline.

*H<sub>3</sub> (Strength)* Predicting trust values by inferring *strengths* of positive and negative experiences yields more accurate results than the baseline.

**Table 1** Computing commitment weights for two examples of commitment creations

Features	Example 1	Example 2
1. Commissive, Directive	Commissive (2)	Commissive (2)
2. Debtor’s type	“I” (2)	“I” (2)
3. Creditor’s type	single (1)	single (1)
4. Modal Verbs	“will” (4)	can (2)
5. Action Verbs	“repair” (1)	“repair” (1)
5. Deadlines	“Monday” (1)	no deadline (0)
Total Weight	11	8

We consider  $H_3$  since a subject might assign a trust level based on his or her varying experiences from different commitment outcomes. Thus, we posit that considering weights of such commitments in addition to learning trust parameters for each subject can improve the trust prediction accuracy more than the baseline.

$H_4$  Subjects' trust assessment behavior as bystanders differs from their trust assessment behaviors as players.  $H_4$  is measured in terms of the following subhypotheses, which posit that subjects' trust assessment behaviors vary across decision contexts.

$H_{41}$  The correlation coefficients (R) between subjects' trust values and positive experiences in emails and the game, respectively, are different.

$H_{42}$  The correlation coefficients (R) between subjects' trust values and neutral experiences in emails and the game, respectively, are different.

**Evaluation process**

Figure 3 summarizes our evaluation process. Our evaluation strategy is to gather data from subjects in the two decision contexts and proceed as follows.

**Step 1.** Build a dataset of interpersonal interactions with trust values. For emails, subjects provide third-party assessments; for games, subjects provide their own trust assessments. Table 2 shows the examples of email and chat interactions. Based on the emails exchanged between Kim and Dorothy, the subjects assign trust values in the range of 0 and 1, from Kim toward Dorothy as well as from Dorothy toward Kim. Similarly, for chats  $P_1$  assigns a trust value for  $P_4$  and  $P_4$  assigns a trust value for  $P_1$ .

**Step 2.** Identify commitment operations from the interactions. For emails, using Kalia et al.'s [11] trained classifier; for games, we find these in the chat interface and analyze them manually. Table 2 shows examples of commitment operations identified in emails and chats, respectively.

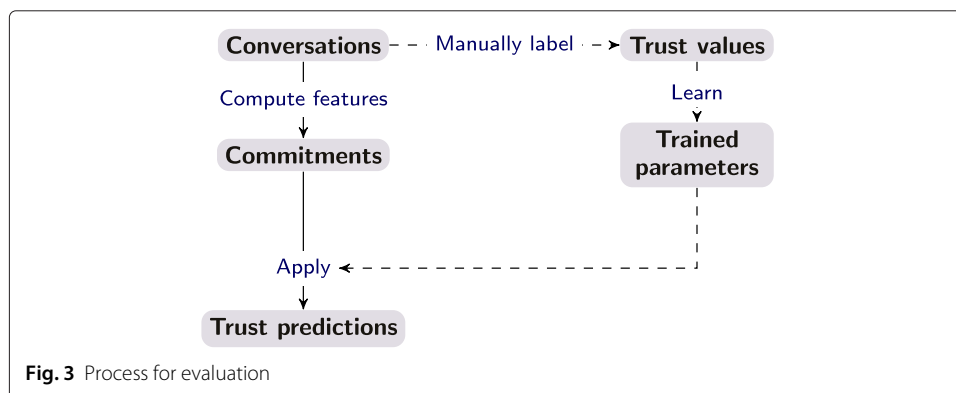
**Step 3.** Partition the dataset into training and test datasets. Learn model parameters for each subject from the training data.

**Step 4.** Apply the learned model to predict trust in the test data and compute the model's accuracy.

We repeat the process for all subjects and present our results.

**Evaluation**

We evaluated Güven via an empirical study with 30 subjects (graduate and undergraduate students from various academic departments at our university). We conducted the



**Table 2** Examples of email and chat interactions

Sender	Receiver	Email content	Operation
Kim	Dorothy	<i>I will also check with Alliance Travel Agency...</i>	create(C <sub>1</sub> )
Kim	Dorothy	<i>I checked with our Travel Agency...</i>	discharge(C <sub>1</sub> )
Rob	Kim	<i>By Wednesday Aug 16 2001, please send all copies of your documentation...</i>	create(C <sub>2</sub> )
Kim	Rob	<i>Rob, please forgive me for not sending this in by Aug 15</i>	cancel(C <sub>2</sub> )
Sender	Receiver	Chat content	Operation
P <sub>1</sub>	P <sub>4</sub>	<i>Can you please give 1 green tile?</i>	
P <sub>4</sub>	P <sub>1</sub>	<i>I can</i>	create(C <sub>3</sub> )
P <sub>4</sub>	P <sub>1</sub>	<i>P<sub>4</sub> transferred P<sub>1</sub> 1 green tile</i>	discharge(C <sub>3</sub> )

study in two phases. In the first phase, subjects read 33 emails selected from the Enron email corpus [28, 29] and provided a trust value ranging from 0 to 1 between the senders and receivers of email. The emails were selected on the basis of their containing sentences that indicate commitment creation, satisfaction, or violation—such sentences having been identified using Kalia et al.’s [11] method. We augmented the dataset with 28 synthetic sentences indicating commitment satisfaction or violation, which do not occur frequently in the corpus. Subjects provided trust values based on their intuitions by reading these emails. We did not disclose the commitments identified. We did not provide any additional guidelines that might restrict a subject’s individual perception of trust. Once the subjects provided their estimated trust values, we mapped commitment operations to positive, negative, and neutral experiences. Table 3 shows an example of two rows created from the first two interactions between Kim and Dorothy given in Table 2. S1, ..., S6 in Table 3 represent the subjects who provided trust values based on the interactions between Kim and Dorothy. Based on the experiences collected from emails and trust values collected from subjects, we created 28 rows of data for each subject. Hence, for 30 subjects we obtain 28×30 or 840 rows. We provide additional details of our data, including a link to download it, in the Appendix (in Section ‘Data’).

Additionally, we asked subjects several questions based on the features discussed in Section ‘Considering strength’. The questions were about ranking these features in order of their perceived importance. We provide these questions in the Appendix in Section ‘Questionnaires’. Hypotheses H<sub>1</sub>, H<sub>2</sub>, and H<sub>3</sub> are about evaluating the proposed approaches with respect to their prediction accuracy.

For the second phase, we augmented Gal et al.’s [12] Colored Trails game with a chat interface. Each subject played three games of five rounds each. Figure 4 shows an instance of the game in progress. The game has a 4×4 board with a chat interface to communicate with one’s opponent. In each round, subjects were allocated a fixed number of colored tiles, a starting position, and a common goal position on the board. To reach the goal, a subject must provide the requisite tiles. During the game, subjects can communicate and

**Table 3** Different features and trust values from different subjects

Trust pairs	Experiences	S1	S2	S3	S4	S5	S6
Kim→Dorothy	2 Neutral, 0 Positive, 0 Negative	0.3	0.4	0.35	0.5	0.6	0.7
Dorothy→Kim	1 Neutral, 1 Positive, 0 Negative	0.7	0.9	0.8	0.7	0.6	0.95

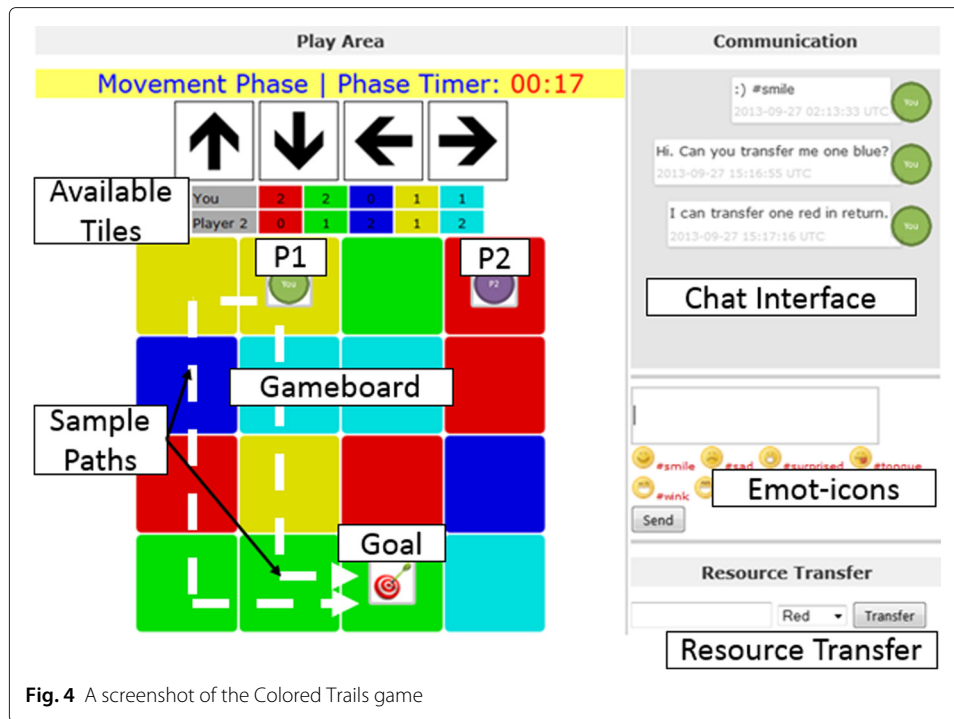


Fig. 4 A screenshot of the Colored Trails game

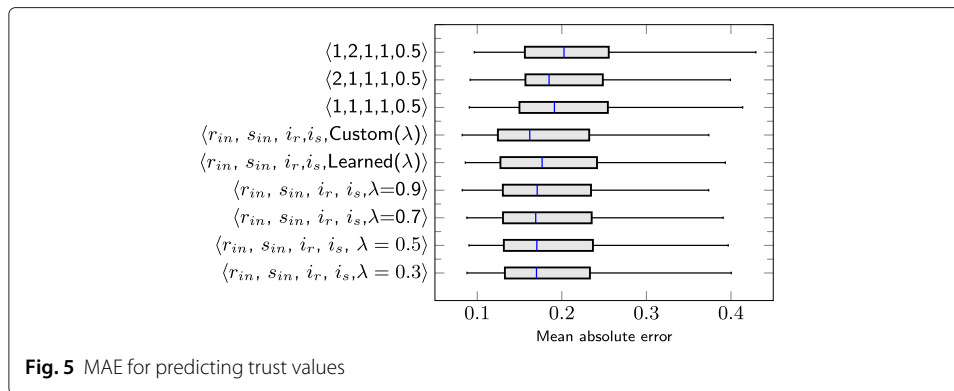
trade tiles via the chat interface. A subject could commit to an opponent to transfer specified tiles and could discharge or violate each commitment. After each round, subjects recorded their trust for their opponents on a five-point scale.

We randomly split our 30 subjects into five groups of six each, and each group into two subgroups of three subjects each. The players in any game sat in separate rooms and communicated through the chat tool. The subjects did not know the identities of the other subjects but knew they had the same opponent for all three rounds of each game. Hypotheses  $H_4$ ,  $H_{41}$ , and  $H_{42}$  are about evaluating whether a subject’s estimation profile differs across decision-making contexts.

### Results

[Verifying  $H_1$ ] We collected the trust values from the subjects from the emails assigned to them. We divided the data collected from subjects into three-fold training and test data and learned trust parameters for each subject ( $r_{in}$ ,  $s_{in}$ ,  $i_r$ ,  $i_s$ ,  $\lambda$ ) that minimize the mean absolute error (MAE) between predicted and actual trust values.

For verifying  $H_1$ , we calculated the MAE for  $\lambda$  ranging from 0.1 to 0.9. Then, we calculated the MAE by learning the  $\lambda$  (Learned( $\lambda$ )) itself. Based on the above MAEs, we obtained a customized  $\lambda$  (fixed or learned) for each subject. A customized  $\lambda$  for a subject refers to the value of  $\lambda$  for which the MAE is minimum. We represent the MAEs obtained using customized  $\lambda$ s for all subjects as Custom( $\lambda$ ) in Fig. 5. Finally, we arbitrarily assumed some fixed configurations of parameters (F1 =  $\langle 1, 1, 1, 1, 0.5 \rangle$ , F2 =  $\langle 2, 1, 1, 1, 0.5 \rangle$ , F3 =  $\langle 1, 2, 1, 1, 0.5 \rangle$ ). F1 indicates no bias in the initial trust perception where as F2 and F3 indicate positive and negative biases respectively.  $\lambda=0.5$  in fixed configurations indicates equal trust increments for the neutral experiences. The configurations can be changed by incrementing or decrementing different parameters. From the results, we observed



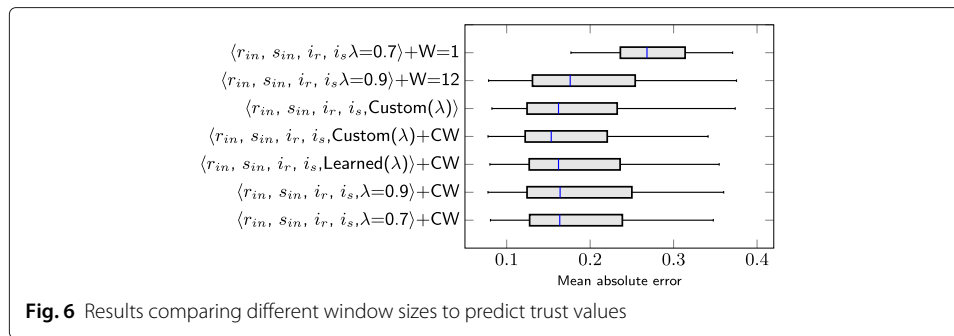
that the median of  $Custom(\lambda)$  (0.162) is less than the medians of all other approaches. To verify if the difference is significant, we evaluated hypothesis  $H_1$  via one-tailed t-test as shown in Table 4. From the results, we found that the difference is not statistically significant. Thus, we concluded that although the overall result (MAEs) seems to align with the hypothesis  $H_1$ , i.e., learning the trust parameters yields more accurate results than using fixed parameters for all subjects, the significance result doesn't confirm it. We leave the evaluation of  $H_1$  for further investigation.

[Verifying  $H_2$ ] For verifying  $H_2$ , first, we determined customized window sizes ( $CW = 1, 2, \dots, 12$ ) for each subject for all the values of  $\lambda$  (0.1, ..., 0.9, learned  $\lambda$ ). A customized window size ( $CW$ ) for a subject refers to the value of  $CW$  for which the MAE is minimum. We found that if we increase the window size further (i.e., greater than 12), the MAEs for the subjects do not decrease. We obtained MAEs for all the subjects based on various values of  $\lambda$  and  $CWs$ , and represent them as  $Custom(\lambda) + CW$  shown in Fig. 6. We compared  $Custom(\lambda) + CW$  with  $Custom(\lambda)$  obtained from  $H_1$  and MAEs obtained from other approaches. From the results, we found that the median of  $Custom(\lambda) + CW$  (0.153) is less than the median of  $Custom(\lambda)$  (0.162) and other approaches. From the one-tailed t-test results shown in Table 5, we found that the mean of  $Custom(\lambda) + CW$  is significantly lower than the means of approaches that consider fixed configurations (e.g.,  $\lambda=0.7 + CW$ ). However, for others, the t-test results show that the differences are not significant. Thus, we concluded that although the overall result (MAEs) seems to align with the hypothesis  $H_2$ , i.e., learning a specific discount window size yields more accurate results than

**Table 4** Statistical test results for  $H_1$

$\mu_{C(\lambda)} < \mu_{others}$	$\mu_{C(\lambda)}$	$\mu_{others}$	p-value
$\mu_{C(\lambda)} < \mu_{F_3}$	0.18	0.21	0.14
$\mu_{C(\lambda)} < \mu_{F_2}$	0.18	0.20	0.18
$\mu_{C(\lambda)} < \mu_{F_1}$	0.18	0.21	0.07
$\mu_{C(\lambda)} < \mu_{L(\lambda)}$	0.18	0.19	0.34
$\mu_{C(\lambda)} < \mu_{\lambda=0.9}$	0.18	0.19	0.36
$\mu_{C(\lambda)} < \mu_{\lambda=0.7}$	0.18	0.19	0.34
$\mu_{C(\lambda)} < \mu_{\lambda=0.5}$	0.18	0.19	0.33
$\mu_{C(\lambda)} < \mu_{\lambda=0.3}$	0.18	0.19	0.34

( $\mu$ -mean of the MAEs)



the baseline, the significance result doesn't confirm it. We leave the evaluation of  $H_2$  for further investigation.

[Verifying  $H_3$ ] From the first phase of our experiment, we obtained subjects' assessments of the weights of commitments. From their assessment we obtained different orderings among the feature values shown in Table 6. The ordering among the values for the creditor type shows that our initial assumption about it was incorrect. For the rest of the feature values, our assumptions correctly aligned with the subjects' assessments.

From the orderings shown in Table 6 we calculated a weight for each commitment (CWT). Based on the weights, we recalculated trust parameters using different  $\lambda$ s, namely, (0.1, ..., 0.9, learned  $\lambda$ ). For  $H_3$ , we calculated a customized  $\lambda$  (Custom( $\lambda$ ) + CWT) and compared the results with  $C(\lambda)$  from  $H_1$ . From the results shown in Fig. 7, we found that the median of Custom( $\lambda$ ) + CWT (0.161) is slightly less than the median of Custom( $\lambda$ ) (0.162). CWT in the figure means considering commitment weight. From the one-tailed t-test results shown in Table 7, the mean of Custom( $\lambda$ ) + CWT is not significantly lower than the means of other approaches. Thus, we concluded that although the overall result (MAEs) seems to align with the hypothesis  $H_3$ , i.e., inferring strengths of positive and negative experiences yields more accurate results than the baseline, the significance result doesn't confirm it. We leave the evaluation of  $H_3$  for further investigation.

[Verifying  $H_4$ ] We verified the subhypotheses  $H_{41}$  and  $H_{42}$  via two-tailed tests at the alpha level of 0.05. For both of them, there are no significant differences between the means of the correlation coefficient (R) obtained from the emails and the game, respectively ( $H_{41}$ :  $p$ -value = 0.32,  $H_{42}$ :  $p$ -value = 0.19). Therefore, subjects' trust assessment behaviors in emails and in games are not different, thereby rejecting Hypothesis  $H_4$ .

### Discussion

Our main contribution is a computational approach for trust that overlays a domain-independent concept describing the social relationships and outcomes of interpersonal

**Table 5** Statistical test results for  $H_2$

$\mu_{C(\lambda) + CW} < \mu_{others}$	$\mu_{C(\lambda) + CW}$	$\mu_{others}$	$p$ -value
$\mu_{C(\lambda) + CW} < \mu_{\lambda=0.7 + CW=1}$	0.17	0.27	0.00
$\mu_{C(\lambda) + CW} < \mu_{\lambda=0.9 + CW=12}$	0.17	0.20	0.08
$\mu_{C(\lambda) + CW} < \mu_{C(\lambda)}$	0.17	0.18	0.27
$\mu_{C(\lambda) + CW} < \mu_{L(\lambda) + CW}$	0.17	0.18	0.32
$\mu_{C(\lambda) + CW} < \mu_{\lambda=0.9 + CW}$	0.17	0.19	0.17
$\mu_{C(\lambda) + CW} < \mu_{\lambda=0.7 + CW}$	0.17	0.18	0.30

**Table 6** Ordering among the feature values obtained from subjects' assessments

Features	Ordering
1. Commissive, Directive	Commissive (2) > Directive (1)
2. Debtor's type	"I" (2) > "We" (1)
3. Creditor's type	multiple creditor (2) > single creditor (1)
4. Modal Verbs	"must, will" (4) > "would, should, shall" (3) > "can, could" (2) > "may" (1)
5. Deadlines	messages with "deadline" (2) > messages without "deadline" (1)

interactions. Previous theoretical approaches, both cognitive [7] and architectural [8, 9], have considered rich concepts but they are not easy to be used as bases for computing trust in the field. By contrast, previous computational approaches have largely worked in an ad hoc manner that bind the trust reasoning to a particular domain.

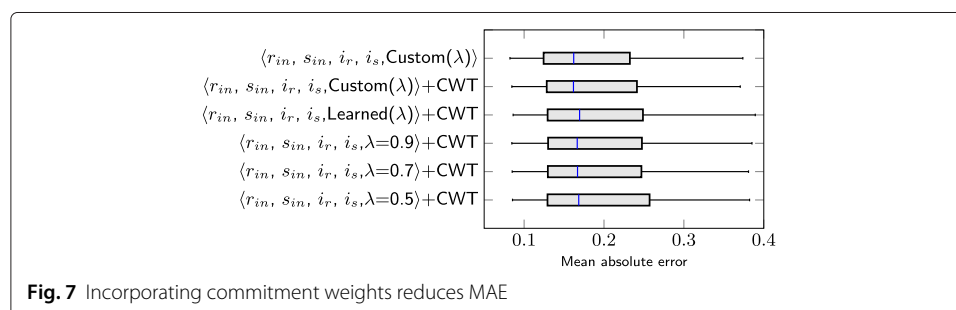
For the email dataset, comparing the means of the MAEs, our approach yields a correlation between subjects' intuitions regarding trust values and those computationally predicted values. Discounting windows customized for each subject yield improved predictions. Considering commitment weights improves predictions further, though not significantly. We additionally evaluate if subjects trust assessment behavior varies across two decision contexts, namely, as bystanders (reading emails exchanged between agents) and as game players (playing the Colored Trails game with opponents that involves competitive as well as cooperative elements). From the t-test results, we find that both hypotheses ( $H_{41}$  and  $H_{42}$ ) are rejected thereby suggesting that subjects' trust assessment behaviors in emails and in games do not vary.

The limitation of our results may be due to the following reasons: (1) lack of adequate data and (2) a greater fraction of experiences being judged neutral than positive or negative. Also, we lack an existing approach to compare our results. However, we submit that our contribution is valuable for having launched a new research direction on computational techniques unifying trust and commitments. Publishing imperfect results, as in this submission, might serve as an antidote to a systematic bias in academic research to favor "success stories" over accurate reporting of empirical results, a bias that is increasingly decried in the scientific disciplines, e.g., [30, 31].

**Future directions**

First, our dataset is not large. A challenge we faced was motivating subjects to provide trust values truthfully for a larger dataset.

Second, our work is limited to predicting trust updates and ignores certainty. According to Wang et al. [4] certainty is the measure of confidence that a trustor places in a trustee



**Fig. 7** Incorporating commitment weights reduces MAE

**Table 7** Statistical test results for H<sub>3</sub>

$\mu_{C(\lambda)+CWT} < \mu_{others}$	$\mu_{C(\lambda)+CWT}$	$\mu_{others}$	p-value
$\mu_{C(\lambda)+CW} < \mu_{C(\lambda)}$	0.18	0.18	0.50
$\mu_{C(\lambda)+CW} < \mu_{L(\lambda)+CWT}$	0.18	0.19	0.37
$\mu_{C(\lambda)+CW} < \mu_{\lambda=0.9+CWT}$	0.18	0.19	0.40
$\mu_{C(\lambda)+CW} < \mu_{\lambda=0.7+CWT}$	0.18	0.19	0.34
$\mu_{C(\lambda)+CW} < \mu_{\lambda=0.5+CWT}$	0.18	0.19	0.32

based on its experiences with the trustee. A truster's certainty increases with increasing number of consistent experiences of the truster with the trustee. Thus, certainty is crucial to trust. However, it is difficult to elicit certainty from subjects since certainty may be more subjective than even trust. A more careful social science style qualitative investigation, as suggested by a reviewer, may be appropriate. In the future, we plan to address these limitations by adopting an incentive scheme that motivates subjects to provide trust values truthfully.

Third, we plan to extend our model to Hidden Markov Models (HMM) [32, 33] that incorporate the temporal aspect of trust, i.e., an agent's current trust is computed based on its past trust. For the same, we will motivate subjects to provide intermittent trust labels by reading emails.

Fourth, there is no reason to be limited to commitments: indeed, we have begun work on bringing in cognitive aspects such as goals and emotions, suitably elicited from subjects, as a basis for creating commitments and judging commitment outcomes and overall trust.

Fifth, a subtle potential benefit of our approach is that it seeks to understand communications and can thus provide more natural explanations for trust estimates than an approach that is purely heuristic. Evaluating this potential benefit would require additional human study, which we defer to future work.

## Endnote

<sup>1</sup>From the Turkish word that brings together the concepts of trust and reliance.

## Appendix

### Data

We provide the data content in the following URL: "<http://tinyurl.com/q59joom>". In the data folder there are two sub folders: (1) *Email\_Documents* and (2) *Processed\_Data*. In the *Email\_Documents* folder, there are 14 files. Each file represents email exchanges between Kimberly and one of her colleague at Enron. In the *Processed\_Data* folder, there are 15 files. Each file represents data in the format given in Table 3. Each files contains experiences observed by the truster from the trustee and corresponding trust values of the truster toward the trustee as assigned by subjects. There are four kinds of processed data files.

- *Processed\_Data\_Emails\_Experiences\_Trust\_Values* file contains experiences obtained from emails between senders and receivers and corresponding trust value assigned by subjects.
- *Processed\_Data\_Game\_Experiences\_Trust\_Values* file contains experiences obtained from chats exchanged between subjects during their game play and corresponding trust value assigned by subjects



- *Processed\_Data\_Emails\_Windows\_Size\_1-12\_Trust\_Values* files contain experiences based on different window sizes (from 1 to 12) and corresponding trust values assigned by subjects. The maximum window size considered is 12 since we mention in Section 'Results' that increasing the window size beyond 12 does not improved the result (reduce the MAE).
- *Processed\_Data\_Emails\_Experiences\_Strength\_Trust\_Values* file contains experiences strength computed from emails based on ranks of features provided in Table 6 and corresponding trust values assigned by subjects.

### Questionnaires

We asked the following questions to subjects to collect their perceived importance about features discussed in Section 'Considering strength'.

Commissive over directive. To assess whether a commissive carry a greater weight than a directive, we asked subjects the following question (choice 1 indicates a directive whereas choice 2 indicates a commissive).

- According to you which is more important?
  - 1. Please review the attached agreement for the Big Sandy Interconnect.
  - 2. I will review the attached agreement for the Big Sandy Interconnect.
  - 3. 1 and 2 are equal

Debtor's type. To assess if a single debtor type carry a greater weight than multiple debtors, we asked subjects the following questions.

- According to you which is more important?
  - 1. We will follow up with Mike to make sure he understands how the numbers were derived.
  - 2. I will follow up with Mike to make sure he understands how the numbers were derived.
  - 3. 1 and 2 are equal
- According to you which is more important?
  - 1. SENDER: Kimberly; RECEIVER: Steven, Teb, Mark, Mansoor, Earl, Stephen, Robert, Jan, Mark, Mansoor, Earl, Stephen, Robert, Jan; Please review the attached work order for the tap and side valve for the new Agave interconnect.
  - 2. SENDER: Kimberly; RECEIVER: Steven; Please review the attached work order for the tap and side valve for the new Agave interconnect.
  - 3. 1 and 2 are equal

Creditor's type To assess whether multiple debtors carry a greater weight than a single creditor, we asked subjects the following question (Choice 1

- According to you which is more important?
  - 1. SENDER: Kimberly; RECEIVER: Lorraine, Lohman, Michelle, Mark, Paul; Would you please send me your bullets by the end of today (before we leave for the Cirque show).
  - 2. SENDER: Kimberly; RECEIVER: Lorraine; Would you please send me your bullets by the end of today (before we leave for the Cirque show).
  - 3. 1 and 2 are equal
- According to you which is more important?
  - 1. SENDER: Michelle; RECEIVER: Rich Cc: Earl; Kimberly; I will speak to Mark Kraus (EOG Commercial) to recap, in case he was not aware of the results.

- 2. SENDER: Michelle; RECEIVER: Rich, Earl, Kimberly; I will speak to Mark Kraus (EOG Commercial) to recap, in case he was not aware of the results.
- 3. SENDER: Michelle, RECEIVER: Rich; I will speak to Mark Kraus (EOG Commercial) to recap, in case he was not aware of the results
- 3. 1 and 2 are equal
- 4. 2 and 3 are equal
- 5. 1 and 3 are equal

Modal verbs. To assess which modal verb convey high confidence over others, we asked subjects the following question.

- 1 Rank the following sentences in a order that ranges over a scale of 1 to 8 where 1 indicates the lowest and 8
  - 1. I would call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 2. I must call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 3. I can call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 4. I will call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 5. I could call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 6. I should call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 7. I may call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].
  - 8. I shall call you on Monday to discuss so we can give it to Danny quickly [Assign Rank: 1–8].

Deadlines. To assess if noun phrases with deadlines may convey more importance than noun phrases without deadlines we asked subjects the following questions

- According to you which is more important?
  - 1. Would you please send me your bullets by the end of today?
  - 2. Would you please send me your bullets?
  - 3. 1 and 2 are equal

#### **Competing interests**

The authors declare that they have no competing interests.

#### **Authors' contributions**

AK developed the model, methods, and experiments. ZZ evaluated the data collected from experiments. MS provided the overall direction and helped realizing the model and methods. All authors read and approved the final manuscript.

#### **Acknowledgment**

We thank the anonymous reviewers for their helpful comments. This research is partially supported by a US Army Research Laboratory (ARL) Oak Ridge Institute for Science and Education (ORISE) Fellowship and the US Department of Defense through the Science of Security Lablet.

#### **Author details**

<sup>1</sup>Department of Computer Science, North Carolina State University, 27695 Raleigh, USA. <sup>2</sup>IBM Corporation, Research Triangle Park, 27709 Durham, USA.

Received: 8 April 2015 Accepted: 15 December 2015

Published online: 05 January 2016

## References

- Scissors LE, Gill AJ, Geraghty K, Gergle D (2009) In CMC we trust: The role of similarity. In: Proceedings of the 27th International Conference on Human Factors in Computing System. ACM, Boston. pp 527–536
- Adali S, Sisenda F, Magdon-Ismail M (2012) Actions speak as loud as words: Predicting relationships from social behavior data. In: Proceedings of the 21st International Conference on World Wide Web. ACM, Lyon. pp 689–698
- DuBois T, Golbeck J, Srinivasan A (2011) Predicting trust and distrust in social networks. In: Proceedings of 3rd International Conference on Social Computing. IEEE, Boston. pp 418–424
- Wang Y, Hang C-W, Singh MP (2011) A probabilistic approach for maintaining trust based on evidence. *J Artif Intell Res* 40:221–267
- Teacy WTL, Patel J, Jennings NR, Luck M (2006) TRAVOS: Trust and reputation in the context of inaccurate information sources. *J Autonomous Agents Multi-Agent Syst* 12(2):183–198
- Jøsang A (1998) A subjective metric of authentication. In: Proceedings of the 5th European Symposium on Research in Computer Security (ESORICS), volume 1485, of Lecture Notes in Computer Science. Springer, Louvain-la-Neuve, Belgium. pp 329–344
- Castelfranchi C, Falcone R (2010) Trust Theory: A Socio-Cognitive and Computational Model. Agent Technology. John Wiley & Sons, Chichester, UK
- Chopra AK, Paja E, Giorgini P (2011) Sociotechnical trust: An architectural approach. In: Proceedings of the 30th International Conference on Conceptual Modeling (ER), volume 6998 of Lecture Notes in Computer Science. Springer, Brussels. pp 104–117
- Singh MP Trust as dependence: A logical approach. In: Proceedings of the 10th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS). IFAAMAS, Taipei. pp 863–870
- Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. *Acad Manag Rev* 20(3):709–734
- Kalia AK, Nezhad HRM, Bartolini C, Singh MP (2013) Monitoring commitments in people-driven service engagements. In: Proceedings of the 10th IEEE International Conference on Services Computing. IEEE, Santa Clara. pp 160–167
- Gal Y, Grosz B, Kraus S, Pfeffer A, Shieber S (2010) Agent decision-making in open-mixed networks. *Artif Intell* 174(18):1460–1480
- Singh MP (1999) An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artif Intell Law* 7(1):97–113
- Singh MP (2008) Semantical considerations on dialectical and practical commitments. In: Proceedings of the 23rd Conference on Artificial Intelligence (AAAI). AAAI Press, Chicago. pp 176–181
- Kalia AK, Singh MP (2015) Muon: Designing multiagent communication protocols from interaction scenarios. *J Autonomous Agents Multi-Agent Syst* (JAAMAS) 29(4):621–657
- Telang PR, Singh MP (2012) Specifying and verifying cross-organizational business models: An agent-oriented approach. *IEEE Trans Serv Comput* 5(3):305–318
- Osman N, Sierra C, McNeill F, Pane J, Debenham J (2014) Trust and matching algorithms for selecting suitable agents. *ACM Trans Intell Syst Technol* 5(1):16:1–16:39
- Kastidou G, Larson K, Cohen R (2014) A trust model for truthful disclosure of anticipated contributions in multiagent systems. In: Proceedings of the 3rd Workshop on Incentives and Trust in E-Communities. AAAI, Québec City. pp 19–24
- Burnett C, Oren N (2012) Sub-delegation and trust. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, Valencia. pp 1359–1360
- Jøsang A, Hayward R, Pope S (2006) Trust network analysis with subjective logic. In: Proceedings of the 29th Australasian Computer Science Conference. ACSC, Australian Computer Society, Inc., Hobart, Australia. pp 85–94
- Singh MP, Chopra AK, Desai N (2009) Commitment-based service-oriented architecture. *IEEE Comput* 42(11):72–79
- Wang Y, Singh MP (2010) Evidence-based trust: A mathematical model geared for multiagent systems. *ACM Trans Autonomous Adaptive Syst* (TAAS) 5(4):14:1–14:28
- Colman TF, Li Y (1996) An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J Optimization* 6(2):418–445
- Zhang J, Cohen R (2013) A framework for trust modeling in multiagent electronic marketplaces with buying advisors to consider varying seller behavior and the limiting of seller bids. *ACM Trans Intell Syst Technol* 4(2):24:1–24:22
- Nartey M, Yankson FE (2014) A semantic investigation into the use of modal auxiliary verbs in the manifesto of a Ghanaian political party. *Int J Humanities Soc Sci* 4(3):21–304
- Burchardt A, Erk K, Frank A (2005) A WordNet detour to FrameNet. In: Proceedings of the GLDV 2005 Workshop GermaNet II, Bonn. pp 1–15. [http://www.cl.uniheidelberg.de/~frank/papers/gnws05\\_burchardt\\_erk\\_frank-final.pdf](http://www.cl.uniheidelberg.de/~frank/papers/gnws05_burchardt_erk_frank-final.pdf)
- Comerio M (2013) Value-based service contract selection. In: Proceedings of the 2013 IEEE International Conference on Services Computing. IEEE, Santa Clara. pp 611–618
- Fiore A, Heer J (2004) UC Berkeley Enron email analysis. [http://bailando.sims.berkeley.edu/enron\\_email.html](http://bailando.sims.berkeley.edu/enron_email.html)
- Klimt B, Yang Y (2004) The Enron corpus: A new dataset for email classification research. In: Proceedings of the 15th European Conference on Machine Learning, volume 3201 of LNCS. Springer, Pisa. pp 217–226
- Kicinski M (2013) Publication bias in recent meta-analyses. *PLoS ONE* 8(11):e81823
- Scargle JD (2000) Publication bias: The “file-drawer” problem in scientific inference. *J Sci Exploration* 14(1):91–106. [http://www.scientificexploration.org/journal/jse\\_14\\_1\\_scargle.pdf](http://www.scientificexploration.org/journal/jse_14_1_scargle.pdf)
- Liu X, Datta A (2012) Modeling context aware dynamic trust using Hidden Markov Model. In: Proceedings of the 26th National Conference on Artificial Intelligence. AAAI Press, Toronto. pp 1938–1944
- Vogiatzis G, MacGillivray I, Chli M (2010) A probabilistic model for trust and reputation. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems. IFAAMAS, Toronto. pp 225–232