Trust Dynamics and Transfer across Human-Robot Interaction Tasks: Bayesian and Neural Computational Models

Harold Soh, **Shu Pan**, **Min Chen** and **David Hsu**
Dept. of Computer Science, National University of Singapore
{harold, panshu, chenmin, dyhsu}@comp.nus.edu.sg

Abstract

This work contributes both experimental findings and novel computational human-robot trust models for multi-task settings. We describe Bayesian non-parametric and neural models, and compare their performance on data collected from real-world human-subjects study. Our study spans two distinct task domains: household tasks performed by a Fetch robot, and a virtual reality driving simulation of an autonomous vehicle performing a variety of maneuvers. We find that human trust changes and transfers across tasks in a structured manner based on perceived task characteristics. Our results suggest that task-dependent functional trust models capture human trust in robot capabilities more accurately, and trust transfer across tasks can be inferred to a good degree. We believe these models are key for enabling trust-based robot decision-making for natural human-robot interaction.

1 Introduction

Trust forms the fabric of human-human relationships and, by extension, the relationships between humans and autonomous agents. For example, our decision to delegate tasks to another human or autonomous agent depends substantially on our trust in the agent [Xie *et al.*, 2019]. As AI enters our daily life, human trust in autonomous agents will impact how these systems are used or *misused*. Trust calibration is crucial for preventing over-trust, which results in unwarranted reliance in robots [Robinette *et al.*, 2016], or under-trust, which can cause poor utilization [Lee and See, 2004].

This extended abstract summarizes our recent work [Soh *et al.*, 2018][1] brings together empirical findings and new modeling tools to produce novel computational models of trust dynamics in human-robot interaction tasks. In contrast to prevailing approaches, e.g., [Lee and Moray, 1994; Xu and Dudek, 2015], our models leverage inter-task structure and are applicable across different tasks that a single agent must handle. As *predictive* models, they can be easily situated within a decision-theoretic framework, such as the partially observable Markov decision process [Kaelbling *et al.*, 1998], to calibrate trust in human-robot collaborative tasks [Chen *et al.*, 2018; Wang *et al.*, 2016; Nikolaidis *et al.*, 2017; Huang *et al.*, 2018].

We first describe results from a human-subjects study ($n = 32$) in two domains—household object handling and autonomous driving—and show that inter-task trust transfer depends on perceived task similarity, task difficulty, and observed robot performance. Our findings are consistent over both domains, even though the robots and the contexts are different. Specifically, the household domain involves a Fetch robot navigating and handling various household objects. The driving domain involves a simulated autonomous vehicle performing driving and parking maneuvers.

Based on the findings, we formalize trust as a context-dependent latent dynamic function that changes with observations of robot performance across tasks. We focus on the *representation* and *dynamics* of this trust function and develop two specific differentiable models: (i) a Bayesian Gaussian process (GP) [Rasmussen and Williams, 2006] model and (ii) a connectionist model based on recent advances in deep learning. The GP model explicitly encodes a prior assumption that human trust evolves via Bayes rule. In contrast, the neural model is largely data-driven. Both models leverage latent task space representations learned using word vector descriptions of tasks, e.g., “Pick and place an apple”. Our Bayesian and neural models better predict self-reported human trust compared to existing approaches. From one perspective, the GP model extends the single global trust variable used in prior work [Chen *et al.*, 2018; Xu and Dudek, 2015] to a *collection* of latent trust variables. To be clear, both models are computational analogues of trust, and neither model attempts to represent exact trust processes in the human brain. They offer conceptual frameworks for capturing the principles of trust formation and transfer.

2 Human Subjects Study

In this section, we summarize the key findings from our human subjects study. For more details, please refer to [Soh *et al.*, 2018]. In brief, we find that human trust changes and transfers across tasks in a structured manner: observations of robot performance have a greater affect on the trust dynamics over similar tasks compared to dissimilar tasks.

[1] “The transfer of human trust in robot capabilities across tasks”, presented at Robotics: Science and Systems 2019.

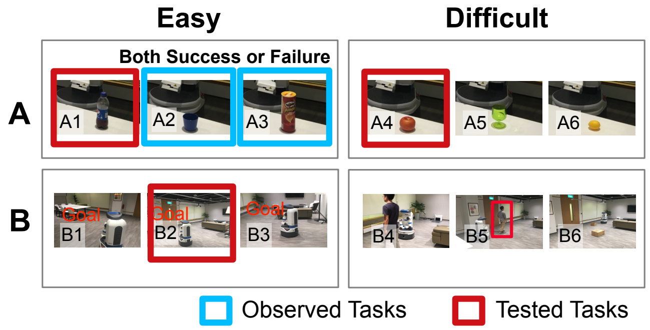


Figure 1: Trust Transfer Experiment Design (Household domain). Each row represents a task category: (A) picking and placing objects, and (B) navigation in a room. Tasks were further categorized into Easy and Difficult tasks. See text for more details.

Experimental design. Our experimental design is summarized in Fig. 1. We explored three factors as independent variables: task category, task difficulty, and robot performance. Each independent variable consisted of two levels: two task categories, easy/difficult tasks, and robot success/failure. We used tasks in two domains:

- **Household** comprising two common categories of household tasks, i.e., picking and placing objects, and navigation in an indoor environment. We used a real-world Fetch robot to perform live in-lab task demonstrations.
- **Driving** involving a simulated autonomous vehicle (AV) performing lane merging and parking, with dynamic and static obstacles. Participants interacted with the simulation system via a Virtual Reality (VR) headset that presented a first-person viewpoint from the driver seat.

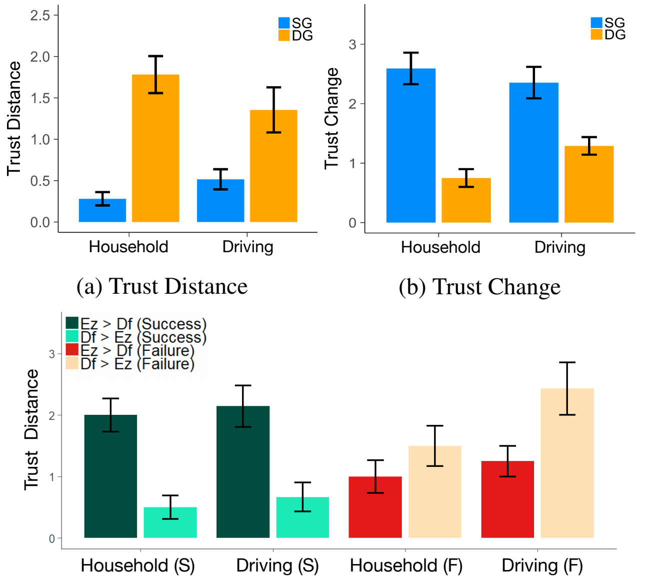
Note the same experiment protocol was conducted *independently* in each domain; having two separate domains enabled us to evaluate the robustness of our findings to different contexts.

Measures and procedure. The main dependent variable was subjective trust in the robot/agent a's capability to perform specific tasks. Participants were first asked to indicate their subjective trust on three "tested tasks". Participants were then randomly assigned to observe two tasks from a specific category and difficulty, and asked to indicate their trust in the robot to accomplish the observed tasks. Finally, participants were asked to re-indicate their trust on the three tested tasks. Participants indicated their degree of trust given task x at time t, denoted as $\tau^a_{x,t}$ and we computed two derivative scores:

- **Trust Distance** $d_{\tau,t}(x, x') = |\tau^a_{x,t} - \tau^a_{x',t}|$, i.e., the 1-norm distance between scores for x and x' at time t.
- **Trust Change** $\Delta\tau^a_x(t_1, t_2) = |\tau^a_{x,t_1} - \tau^a_{x,t_2}|$, i.e., the 1-norm distance between the scores for x at t_1 and t_2.

2.1 Key Results

We recruited 32 individuals (Mean age: 24.09 years, $SD = 2.37$, 46% female). For the driving domain, one participant's results were removed due to a failure to pass attention/consistency check questions.



(a) Trust Distance

(b) Trust Change

(c) Task Difficulty and Trust

Figure 2: (a) Trust distance between a given task and tasks in the same category group (SG) compared to tasks in a different category (DG). Trust in robot capabilities was more similar for tasks in the same group. (b) Trust changes due to observations were greater for tasks in SG versus DG. (c) Trust distance between the observed task and a more difficult task (Ez → Df) versus a simpler task (Df → Ez). Participants who observed successful demonstrations of a difficult task trusted the robot to perform simpler tasks, but not vice-versa.

Our results are summarized in Fig. 2. Briefly, our main findings support the intuition that human trust transfers across tasks and similar tasks are more likely to share a similar level of trust. The trust distances are significantly lower compared to tasks in other categories (DG) for the household ($t(31) = -5.82$, $p < 10^{-5}$) and driving domains ($t(30) = -2.755$, $p < 10^{-2}$).

Fig. 2b shows that trust changes across tasks due to performance observations of a specific task were also moderated by the perceived similarity of the tasks. The trust change for SG is significantly greater than DG in both domains. The trust change for DG was non-zero (one-sample t-test, $p < 10^{-2}$ across both domains), which suggests that trust transfers between task categories, but to a lesser extent.

Finally, we analyzed the relationship between perceived difficulty and trust transfer by splitting the data into two conditions: participants who observed successful demonstrations, and those that observed failures (Fig. 2c). For the success condition, the trust distance among tasks was significantly less for tasks perceived to be easier than the observed task in both the household domain ($t(14) = 4.58$, $p < 10^{-3}$) and driving domain ($p < 10^{-3}$). For the failure condition, the results were not statistically significant (at the $\alpha = 1\%$ level), but suggest that belief in robot inability would transfer more to difficult tasks compared to simpler tasks.

3 Multi-Task Trust Models

The results from our human subjects study indicate that trust is relatively rich mental construct. Here, we present a computational model where trust is a *latent dynamic function*:

$$\tau_t^a(\mathbf{x}) : \mathbb{R}^d \rightarrow [0, 1]$$

that maps task features, $\mathbf{x}$, to real-values indicating trustworthiness of the robot to perform the task. This functional view of trust captures trust differences across tasks, and can be extended to include other contexts, e.g., the current environment, robot properties, and observer characteristics.

To model trust dynamics, we propose a Markovian function g that updates trust,

$$\tau_t^a = g(\tau_{t-1}^a, o_{t-1}^a) \tag{1}$$

where $o_{t-1}^a = (\mathbf{x}_{t-1}, c_{t-1}^a)$ is the observation of robot a performing a task with features $\mathbf{x}_{t-1}$ at time $t-1$ with performance outcome c_{t-1}^a. The function g changes trust given observations of robot performance, and as such, is a function over the space of trust functions. In this work, we consider binary outcomes $c_{t-1}^a \in \{+1, -1\}$ indicating success and failure respectively, but differences in performance can be directly accommodated via "soft labels" $c_{t-1}^a \in [-1, +1]$ without significant modifications to the following methods.

In this work, we propose and evaluate two different forms for τ_t^a and g: (i) a Bayesian approach where we model a probability distribution over latent functions via a Gaussian process, and (ii) a connectionist approach utilizing a recurrent neural network (RNN).

3.1 Bayesian Gaussian Process Trust Model

Our first model formalizes trust formation as a cognitive process in a rational Bayesian framework [Griffiths *et al.*, 2009], i.e., the human is learning about the capabilities of the robot f^a by combining prior beliefs about the robot with evidence (observations of performance) via Bayes rule,

$$p_t(f^a | o_{t-1}^a) = \frac{P(c_{t-1}^a | f^a, \mathbf{x}_{t-1}) p_{t-1}(f^a)}{\int P(c_{t-1}^a | f^a, \mathbf{x}_{t-1}) p_{t-1}(f^a) df^a}, \tag{2}$$

where p_t is the posterior distribution, $P(c_{t-1}^a | f^a, \mathbf{x}_{t-1})$ is the likelihood of observing the robot performance c_{t-1}^a given the task $\mathbf{x}_{t-1}$ and latent function f^a. The human estimates robot trustworthiness by integrating over the posterior:

$$\tau_t^a(\mathbf{x}_i) = \int P(c_i^a | f^a, \mathbf{x}) p_t(f^a) df^a \tag{3}$$

We place a Gaussian process (GP) prior over f^a,

$$p_0(f^a) = \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{4}$$

where $m(\cdot)$ is the prior mean function, and $k(\cdot, \cdot)$ is the kernel or covariance function. In standard machine learning scenarios, GPs are often assumed to have a zero-mean prior, $m(\cdot) = 0$. However, as our human subject studies have shown, perceived task difficulty results in asymmetric trust transfer. As such, we use a linear prior mean function $m(\mathbf{x}) = \boldsymbol{\beta}^\top \mathbf{x}$. In addition, we assume tasks to live on a low-dimensional manifold, i.e., a "psychological task space" and use a projection kernel:

$$k(\mathbf{x}, \mathbf{x}') = \exp(-(\mathbf{x} - \mathbf{x}')^\top \mathbf{M} (\mathbf{x} - \mathbf{x}')) \tag{5}$$

with a low rank matrix $\mathbf{M} = \mathbf{\Lambda} \mathbf{L} \mathbf{\Lambda}^\top$ where $\mathbf{\Lambda} \in \mathbb{R}^{d \times k}$ and $\mathbf{L}$ is a diagonal matrix of length-scales capturing axis-aligned relevances in the projected task space. Given binary success outcomes, we use a probit likelihood [Neal, 1997] and iteratively update trust via approximate Bayesian inference [Csató and Opper, 2002; Soh and Demiris, 2014]. Note that this trust update forms the Markovian function g and is not adapted by learning.

3.2 Neural Trust Model

The Gaussian process trust model is based on the assumption that human trust is fundamentally Bayesian. Here, we consider an alternative "data-driven" model based on recent advances in deep learning where g is learned. We learn task representations or "embedding" space $Z \subseteq \mathbb{R}^k$ and model trust as a parameterized function over this space. Using recurrent neural networks, the model "compresses" the human's prior interaction history into a trust vector $\boldsymbol{\theta}_t$. Trust is then computed as a simple sigmoid function of the dot product between task embeddings and this trust vector,

$$\tau_t^a(\mathbf{x}; \boldsymbol{\theta}_t) = \mathrm{sigm}(\boldsymbol{\theta}_t^\top f_z(\mathbf{x})) = \mathrm{sigm}(\boldsymbol{\theta}_t^\top \mathbf{z}), \tag{6}$$

where $f_z(\mathbf{x})$ is a function that maps task features $\mathbf{x}$ to task representations $\mathbf{z}$. The trust function τ_t^a is fully parameterized by $\boldsymbol{\theta}_t$ and its linear form has benefits: it is efficient to compute given a task representation $\mathbf{z}$ and is interpretable in that the latent task space Z can be examined, similar to other dot-product spaces, e.g., word embeddings [Mikolov *et al.*, 2013].

We model the trust update function g using a RNN with parameters θ_g,

$$\boldsymbol{\theta}_t = \mathrm{RNN}(\boldsymbol{\theta}_{t-1}, \hat{\mathbf{z}}_{t-1}; \theta_g). \tag{7}$$

We use the Gated Recurrent Unit (GRU) [Cho *et al.*, 2014], which is a variant of long short-term memory [Hochreiter and Schmidhuber, 1997] with strong empirical performance [Jozefowicz *et al.*, 2015].

Similar to the GP, Z can be seen as a psychological task space in which the similarities between tasks can be easily determined. We project observed task features $\mathbf{x}$ into Z via a nonlinear function, specifically, a fully-connected neural network,

$$\mathbf{z} = f_z(\mathbf{x}) = \mathrm{NN}(\mathbf{x}; \theta_z) \tag{8}$$

where θ_z is the set of network parameters. Similarly, the robot's performance c^a is projected via a neural network, $\mathbf{c}^a = \mathrm{NN}(c^a; \theta_c^a)$. During trust updates, both the task and performance representations are concatenated, $\hat{\mathbf{z}}_i = [\mathbf{z}; \mathbf{c}^a]$ and input to the GRU.

4 Computational Experiments

Our computational experiments were designed to answer three specific questions: (**Q1**) Do our models outperform existing approaches on unseen participants? (**Q2**) Do the models generalize to held-out tasks? (**Q3**) How important is it to model differences in initial perceptions of task difficulty?

Our first experiment (**E1**) was a variant of 10-fold cross-validation where we held-out data from 10% of the participants (3 people). The second experiment (**E2**) held-out all

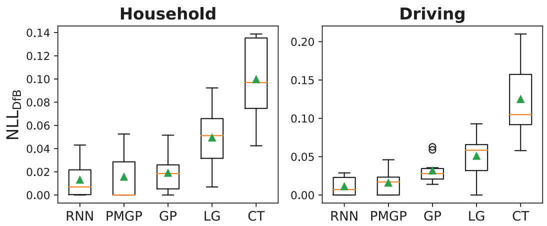


Figure 3: $\mathrm{NLL}_{\mathrm{DfB}}$ scores for experiment **E1** with medians (lines) and means (triangles) shown. The proposed neural and Bayesian Trust models (RNN and PMGP) achieve better scores on previously unseen participants, than the alternative models that do not consider trust transfer.

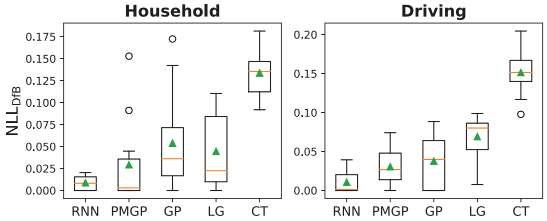


Figure 4: $\mathrm{NLL}_{\mathrm{DfB}}$ scores for experiment **E2** with medians (lines) and means (triangles) shown. Compared to existing models, the proposed methods—RNN and PMGP—are able to better predict trust on previously unseen tasks.

trust data associated with one task and trained on the remaining tasks. In both experiments, we evaluated 5 different models: **RNN**: The neural trust model; **PMGP**: The GP trust model with prior mean; **GP**: A zero-mean Gaussian process trust model; **LG**: A linear Gaussian trust model similar to the linear Gaussian updates used in OPTIMo [Xu and Dudek, 2015], and trained in the same manner as the RNN and GP models; and **CT**: A baseline model with constant trust.

In our evaluations, training involved learning the relevant model parameters, e.g., $\boldsymbol{\beta}$ and $\mathbf{M}$ for the GP. For *each* participant, each model was updated *only* with the tasks and outcomes that the participant observed. Prediction and evaluation was carried out on both pre-and-post-update trust scores. All models were trained and tested using the data collected in our human subjects study. Preliminary cross-validation runs were used to ascertain good parameters (details in [Soh *et al.*, 2018]). For task features, we used 50-dimensional GloVe word vectors [Pennington *et al.*, 2014] computed from the task descriptions. Models were optimized via ADAM algorithm [Kingma and Ba, 2014] using the Bernoulli likelihood of observing the trust scores (as soft labels); when trust is unobserved, the models can be trained using observed human actions.

To mitigate significant differences across the folds (each participant/task split), we compared the methods using relative Difference from Best (DfB) scores:$\mathrm{NLL}_{\mathrm{DfB}}(i,k) = \mathrm{NLL}(i,k) - \mathrm{NLL}^*(i)$, where $\mathrm{NLL}(i,k)$ is the NLL achieved by model k on fold i and $\mathrm{NLL}^*(i)$ is the best score among the tested models on fold i. $\mathrm{MAE}_{\mathrm{DfB}}$ is similarly defined. Our key results still hold when comparing NLL and MAE scores.

4.1 Results

Results for **E1** are summarized in the boxplots shown in Fig. 3; we only show the $\mathrm{NLL}_{\mathrm{DfB}}$ measure as the $\mathrm{MAE}_{\mathrm{DfB}}$ was very similar. The RNN and PMGP models appear comparable, and outperform the baselines on both datasets. This suggests our models — which account for trust differences and task transfer — achieved better trust predictions on unseen participants (**Q1**).

Turning our attention to **E2**, we see that the RNN and PMGP once again outperform the other models (Fig. 4). Interestingly, the gap between them is larger on **E2**, with the RNN achieving lower scores than the PMGP. This suggests that the RNN learnt a better mapping from the task feature space to the latent task space. Nevertheless, both models are able to make accurate trust predictions on *unseen* tasks (**Q2**).

Finally, to answer **Q3**, we examined the differences between the GP models. The PMGP model always achieved lower or similar scores to the GP model, suggesting that the difficulty modeling enabled better performance. While not conclusive evidence—the PMGP is more flexible due to the extra parameters—it does suggest the importance of accounting for the asymmetries resulting from prior beliefs.

5 Discussion

Human trust in automation is a large interdisciplinary research endeavor spanning multiple fields, including human-factor engineering, psychology, and human-robot interaction [Lee and Moray, 1992; Muir, 1994; Hancock *et al.*, 2011; Chen *et al.*, 2018; Soh *et al.*, 2018]. Even so, crucial gaps remain in our understanding of when and how humans trust autonomous agents. In particular, how do we formalize trust, its representation and dynamics given variability among people, tasks, and robots? This work takes a key first step towards conceptualizing and formalizing computational models for predicting human trust *across tasks*. The following highlights the interesting new questions arising from our findings:

- **Combining Bayesian and neural trust models.** A natural question is whether we can formulate a "structured" trust update that combines the simplicity and interpretability of the Bayes update, while allowing for additional flexibility via a neural network. Preliminary experiments show such a hybrid model is able to achieve better prediction accuracy.

- **Trust in intent.** In this work, we limited our investigation to trust in the robot's capabilities. However, it is also essential to examine trust in the robot's "intention", e.g., its policy [Huang *et al.*, 2018] and decision-making process. In very recent work [Xie *et al.*, 2019], we examined how human mental models of both these factors influence decisions to trust robots.

- **Trust-based decision making.** Our prior work [Chen *et al.*, 2018] showed that decision-making (via a POMDP) using a human trust model results in policies that improved overall task performance. Integrating more complex models, such as the ones proposed in this work, into a decision-making framework for real-time trust inference and calibration remains a key open problem.

References

[Chen *et al.*, 2018] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. Planning with trust for human-robot collaboration. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 307–315, New York, NY, USA, 2018. ACM.

[Cho *et al.*, 2014] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Conference on Empirical Methods in Natural Language (EMNLP)*, pages 1724–1734, 2014.

[Csató and Opper, 2002] Lehel Csató and Manfred Opper. Sparse on-line gaussian processes. *Neural Computation*, 14(3):641–668, March 2002.

[Griffiths *et al.*, 2009] Thomas L Griffiths, Christopher G Lucas, Joseph J Williams, and Michael L Kalish. Modeling human function learning with Gaussian processes. *Advances in Neural Information Processing Systems*, pages 553–560, 2009.

[Hancock *et al.*, 2011] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5):517–527, 2011.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–80, 1997.

[Huang *et al.*, 2018] Sandy H. Huang, Kush Bhatia, Pieter Abbeel, and Anca D. Dragan. Establishing (appropriate) trust via critical states. In *Workshop on Explainable Robotic Systems, HRI*, 2018.

[Jozefowicz *et al.*, 2015] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. An empirical exploration of recurrent network architectures. In *International Conference on Machine Learning*, pages 2342–2350, 2015.

[Kaelbling *et al.*, 1998] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Lee and Moray, 1992] John Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992.

[Lee and Moray, 1994] John D Lee and Neville Moray. Trust, self-confidence, and operators' adaptation to automation. *Intl. J. of Human-Computer Studies*, 40(1):153–184, 1994.

[Lee and See, 2004] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004.

[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013.

[Muir, 1994] Bonnie M Muir. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994.

[Neal, 1997] Radford M Neal. Monte carlo implementation of Gaussian process models for Bayesian regression and classification. *arXiv preprint physics/9701026*, 1997.

[Nikolaidis *et al.*, 2017] Stefanos Nikolaidis, Swaprava Nath, Ariel D. Procaccia, and Siddhartha Srinivasa. Game-theoretic modeling of human adaptation in human-robot collaboration. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 323–331. ACM, 2017.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

[Rasmussen and Williams, 2006] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[Robinette *et al.*, 2016] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. Overtrust of robots in emergency evacuation scenarios. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 101–108, 2016.

[Soh and Demiris, 2014] Harold Soh and Yiannis Demiris. Spatio-temporal learning with the online finite and infinite echo-state gaussian processes. *IEEE Transactions on Neural Networks and Learning Systems*, 26:522–536, June 2014.

[Soh *et al.*, 2018] Harold Soh, Shu Pan, Chen Min, and David Hsu. The transfer of human trust in robot capabilities across tasks. In *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.

[Wang *et al.*, 2016] Ning Wang, David V. Pynadath, and Susan G. Hill. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 109–116, 2016.

[Xie *et al.*, 2019] Yaqi Xie, Indu P Bodala, Desmond C Ong, David Hsu, and Harold Soh. Robot capability and intention in trust-based decisions across tasks. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 39–47. IEEE, 2019.

[Xu and Dudek, 2015] Anqi Xu and Gregory Dudek. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 221–228. ACM, 2015.