A Simple Agent-Based Model of the Development of Trust in Hierarchical Organizations

Donovan Platt[1, *], Syed Arefinul Haque[2], Nai Seng Wong[3], William Leibzon[4,+], Xiongrui Xu[5,+], Rudi Minxha[6,+], and Lu Liu[7,+]

[1]University of the Witwatersrand, School of Computer Science and Applied Mathematics
[2]Northeastern University, Network Science Institute
[3]Monetary Authority of Singapore
[4]University of California Irvine, Institute for Mathematical Behavioral Sciences
[5]University of Electronic Science and Technology of China, CompleX Lab, Web Sciences Center
[6]Trumid Financial
[7]Pennsylvania State University, College of Information Science and Technology
[*]donovan.platt@students.wits.ac.za
[+]these authors contributed equally to this work

ABSTRACT

Hierarchical organizations typically rely on a chain of command in which supervisors instruct subordinates to perform certain actions in order to advance the interests and objectives of the organization. The trust between supervisors and subordinates is essential for the functioning of such hierarchical organizations, as it ensures the eventual execution of instructions. Despite this, blind trust and obedience may lead to the execution of instructions with negative consequences, which may have been avoided had a subordinate been more critical of a supervisor's instructions. We thus aim to model the development of trust between subordinates, their supervisors and the structures of the organization itself, subject to the understanding of the consequences of instructions by subordinates and the eventual outcomes of executed instructions. This is achieved through the construction of an agent-based model that is capable of replicating a number of intuitive behaviors and the use of this model to provide an indication of how trust develops over time.

Introduction

In this paper, we investigate the development of trust in hierarchical organizations, with emphasis on a leadership hierarchy in a military organization.

In general, trust is defined as a person's intention to accept vulnerability based on positive expectations of the intentions and behaviors of another person[5]. According to Allen and Braun[1], this allows individuals to manage the uncertainty associated with their interactions and enables them to jointly optimize the gains that will result from cooperative behavior.

While it is tempting to equate trust and cooperation, a key distinction must be made: trust is not cooperation, but rather facilitates cooperation.

The study of trust is particularly relevant to military organizations, as it is critical to unit cohesion and effectiveness. In turn, the high risks and interdependencies in military organizations create opportunities for us to observe the development and impact of trust.

Rousseau et al.[4] describe different forms of trust, namely:

- *Deterrence-based trust*, which involves imposing costly sanctions for any breaches of trust. It is questionable whether this is trust or just enforced cooperation.
- *Calculative trust*, which is based on rational choice, taking into account perceptions of the other party's intentions.
- *Relational trust*, which develops with repeated interactions over time, allowing the parties to obtain information from within the relationship that affects their willingness to trust.
- *Institutional trust*, which refers to the institutional factors that provide broad support for trust to develop.

We focus our investigations on relational, calculative and institutional trust.

Allen and Braun[1] identify the main drivers of trust to be:

- Credibility of competence
- Integrity, fairness and honesty
- Benevolence of motives
- Predictability of behavior

We focus primarily on credibility of competence, with good and bad outcomes tending to influence the evolution of trust in the model over time.

Based on the above theoretical insights from literature in the social sciences, we construct an agent-based model involving a network structure in an attempt to determine the development of trust and organizational success over time, given different initial organizational states.

Proposed Model

Network Structure

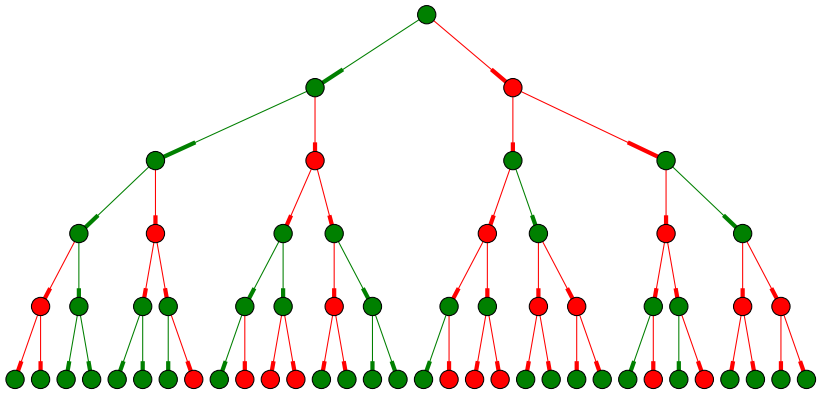

Figure 1. An example of the hierarchical network structure employed in our model, with green nodes representing agents who sent a good $(+1)$ order to their subordinates and red nodes representing agents who sent a bad (-1) order to their subordinates, noting that subordinates sometimes alter the order sent by their supervisor before sending it to their subordinates.

The model consists of a hierarchical network in which the root node in the hierarchy sends an order signal (-1 for a bad order and +1 for a good order) to its subordinates. Each subordinate then either transmits the received order or flips the order (i.e. transmits the opposite order) based on a decision rule that takes into account the:

- Trust the subordinate has in the supervisor who issued the order
- Extent to which the subordinate understands the consequences of the order
- Subordinate's trust in the institution

If a particular agent lacks confidence in the received order based on the aforementioned criteria, the agent sends the opposite order to its subordinate nodes in the network, where this process is again repeated.

Once the order reaches the leaf nodes of the network, we determine the outcome of the iteration (or mission in a military context), where a mission is successful if the number of good orders propagated through the network is greater than the number of bad orders.

Once the success or failure of the mission has been determined, each agent updates their trust in their supervisor and broader institutional trust based on a number of cases discussed in more detail in subsequent sections. The root node then sends another order, starting another mission, with this process repeating for a desired number of iterations.

In the above simulation, the root node sends both good and bad orders according to a binomial distribution, with the probability of a good order being given by the parameter p.

Finally in the simple simulations we present throughout the paper, we assume that each node has two children and that there are 5 levels in the hierarchy. This could easily be relaxed in future work.

Decision Rule

Agents decide whether to propagate or change an order received from a supervisor according to a decision rule that takes a variety of factors into consideration, including the agent's relational trust in their supervisor, the agent's calculative trust based on the agent's understanding of the consequences of the order, and the agent's institutional trust in the organization.

Notation

We define:

- β_t^i, the order received by agent i at time t, -1 being a bad order, $+1$ being a good order.
- τ_t^i, the confidence of the i-th agent in order β_t^i.
- α_t^i, the i-th agent's trust in their supervisor based on past experience (relational trust). We set α between 0 and 1.
- H^i, the information available to agent i, which we define as the extent to which the i-th agent understands the consequences of any given order. We set H between 0 and 0.5, with 0 implying no knowledge of the consequences of an arbitrary order and 0.5 implying moderate knowledge. We assume full knowledge $(H = 1)$ is not possible.
- δ_t, the agents' inherent skepticism of orders received from members of the institution, which we define to be inversely proportional to the overall level of institutional trust within the organization. Higher δ values imply that agents are less likely to accept orders and vice versa. We set δ between 0 and 1.
- n_t^1, the number of good orders propagated through the organization in mission t.
- n_t^{-1}, the number of bad orders propagated through the organization in mission t.

The parameter α quantifies the component of trust based on one's past experiences with a particular supervisor while δ quantifies the component of trust based on one's belief in an institution, though through an inverse relationship.

The initial value of α is set to be 0.5 in our simulations since there is no reason why a subordinate will be particularly trusting or mistrusting of a supervisor without having any prior experience with the supervisor.

Equations

Our decision function is centered on the calculation of the quantity τ_t^i, representing the confidence of the i-th agent in the order received from their supervisor. We define the agent's confidence as the combination of the agent's trust in their supervisor as a result of past experience (relational trust), the agent's own information regarding the consequences of the order (calculative trust) and the agent's inherent skepticism, defined to be the inverse of the agent's trust in the organization (institutional trust). This gives

$$\tau_t^i = \alpha_t^i + H^i \beta_t^i - \delta_t \tag{1}$$

Based on their confidence in the order received from their supervisor, each agent decides to either keep or flip the received order before transmitting it to their own subordinates. Therefore, the i-th agent will set the order transmitted to its subordinates to be

$$\begin{cases} \beta_t^i & \tau_t^i \geq 0 \\ -\beta_t^i & \tau_t^i < 0 \end{cases} \tag{2}$$

After each iteration (or mission), each agent updates their relational trust in their supervisor based on a number of possible situations. We define the mission outcome to be good when the number of good orders propagated exceeds the number of bad

orders propagated. This leads to

$$\alpha_t^i = \begin{cases} \alpha_{t-1}^i + \varepsilon_1^i & n_{t-1}^1 > n_{t-1}^{-1}, \tau_{t-1}^i \geq 0 \\ \alpha_{t-1}^i - \varepsilon_1^i & n_{t-1}^1 > n_{t-1}^{-1}, \tau_{t-1}^i < 0 \\ \alpha_{t-1}^i - \varepsilon_1^i & n_{t-1}^1 \leq n_{t-1}^{-1}, \tau_{t-1}^i \geq 0 \\ \alpha_{t-1}^i + \varepsilon_1^i & n_{t-1}^1 \leq n_{t-1}^{-1}, \tau_{t-1}^i < 0 \end{cases} \tag{3}$$

where $\varepsilon_1^i \sim \mathcal{N}(\mu_1, \sigma_1)$, truncated between 0 and 1.

The above equation represents the following thought processes:

- "It was a good outcome and I listened to my supervisor. I therefore feel he is a person I can trust."
- "It was a good outcome and I did not listen to my supervisor. By not listening to him I feel that I contributed to the good outcome, so I feel he is a person I should trust less."
- "It was a bad outcome and I listened to my supervisor. By listening to him I feel I contributed to the bad outcome, so I feel he is a person I should trust less."
- "It was a bad outcome and I did not listen to my supervisor. By not listening to him, I feel I may have contributed to the bad outcome, so perhaps I should trust him more."

Similarly, we update the inherent skepticism of agents as a whole with decreasing δ in the case of a good outcome, leading to greater trust in the institution, and increasing δ in the case of a bad outcome, leading to lower institutional trust. This gives

$$\delta_t = \begin{cases} \delta_{t-1} - \varepsilon_2 & n_{t-1}^1 > n_{t-1}^{-1} \\ \delta_{t-1} + \varepsilon_2 & n_{t-1}^1 \leq n_{t-1}^{-1} \end{cases} \tag{4}$$

where $\varepsilon_2 \sim \mathcal{N}(\mu_2, \sigma_2)$, truncated between 0 and 1.

Model Validation

We perform basic model validation by attempting to showcase the model's ability to replicate intuitive behavior for different parameter combinations representing different initial levels of information and institutional trust within an organization.

This is much like the stylized-fact centric validation sometimes employed in financial agent-based models, where models are validated by demonstrating an ability to reproduce a number of empirically-observed behaviors[3].

More rigorous validation would require calibration to data, which is beyond the scope of this investigation, in part due to a lack of a compatible dataset.

Initial Parameter Values

We assume that relational trust, institutional trust and the available information are initially constant at all levels in the hierarchy, with $\alpha_0 = 0.5$ in all test cases and the available information and institutional trust varying for different test cases. Specifically, we have the following:

- High information, low institutional trust (high skepticism), $H = 0.5$ and $\delta_0 = 0.8$
- High information, high institutional trust (low skepticism), $H = 0.5$ and $\delta_0 = 0.2$
- Low information, low institutional trust (high skepticism), $H = 0.1$ and $\delta_0 = 0.8$
- Low information, high institutional trust (low skepticism), $H = 0.1$ and $\delta_0 = 0.2$

We further assume that $\mu_1 = 0.05$, $\sigma_1 = 0.001$, $\mu_2 = 0.017$ and $\sigma_2 = 0.0005$ for all test cases.

Each test case is associated with two simulations, the first simulation dealing with the case of effective ($p = 0.8$) leadership and the second dealing with the case of ineffective ($p = 0.2$) leadership.

Validation Procedure

For each of the cases previously described, we simulate 500 missions and determine the mean number of good decisions across all of the simulated missions. We then determine if the mean number of good decisions across the various cases is consistent with the following set of intuitive expectations:

- We expect that a greater amount of information should usually lead to better decisions. The more agents know about the consequences of orders the better the decisions should be on average.
- When a large quantity of information is available to agents, the nature of trust within the organization should not significantly affect the number of good decisions being made, since trust is required in cases of uncertainty, which is largely mitigated by significant information availability[5].
- A degree of healthy skepticism allows for better decision making on average, especially when one does not have access to an abundance of information. In this case, one should rather be less trusting in nature than gullible.

It is worth noting that we do not claim that the above cases are definitive or exhaustive, but rather present a starting point to determine the model's sensibility.

Validation Results

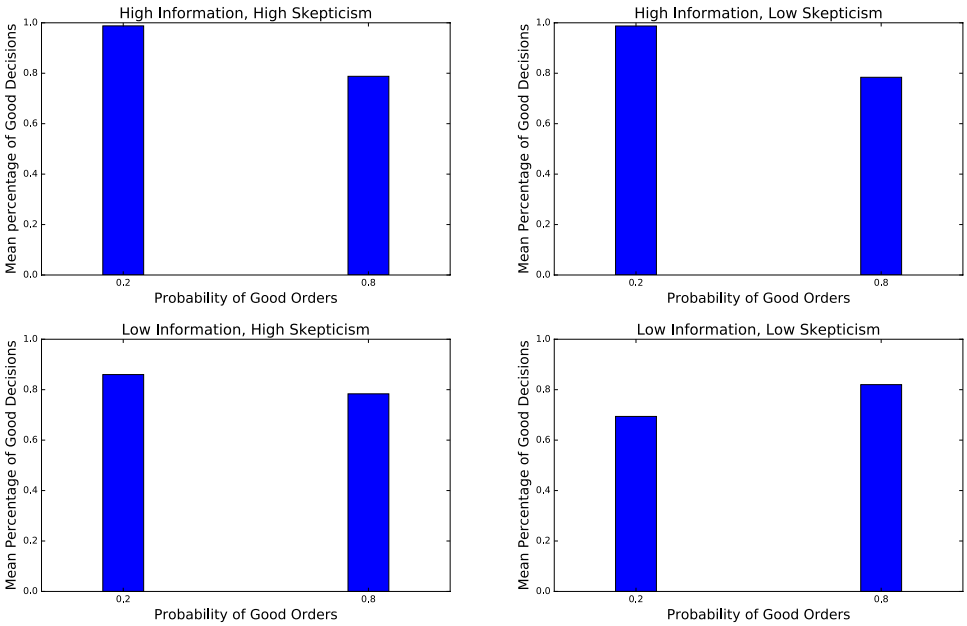

Figure 2. Mean percentages of good decisions made for each of the considered test cases, averaged over 500 missions.

Referring to Figure 2 and comparing the upper two panels to the bottom two panels, we see that for both effective and ineffective leadership and irrespective of the initial levels of institutional trust, higher levels of information available to agents generally result in superior decision-making performance in most cases. This confirms the model's consistency with the first of our validation cases.

Referring to the upper two panels of Figure 2, we see that when a large amount of information is available to the agents, differing levels of institutional trust do not have a large effect on the mean percentage of good decisions being made, confirming the model's consistency with the second of our validation cases.

Finally, referring to the bottom two panels of Figure 2, we see that when the availability of information is low, higher levels of skepticism protect against poor leadership, illustrated by a significantly greater mean percentage of good decisions than those

demonstrated in the case of lower levels of skepticism. While higher skepticism does result in slightly poorer performance than low skepticism in the case of good leadership, this discrepancy is significantly less than that of low and high skepticism in the case of poor leadership. In other words, high skepticism results in significantly better performance in the case of poor leadership, while low skepticism only results in slightly better performance in the case of good leadership. Therefore, higher skepticism is desirable and we have shown our model's consistency with the third of our validation cases.

The Development of Relational and Institutional Trust Over Time

Considering that we have now validated the model to some extent, we use it to demonstrate how trust develops over time within our hypothetical, hierarchical organization. We consider the same simulation and parameter cases employed in the validation procedures, but now focus on the average level of relational trust and overall level of skepticism present in the simulation over the 500 conducted missions for each test case.

High Information Cases

We begin with an analysis of test cases involving high levels of information among agents ($H = 0.5$).

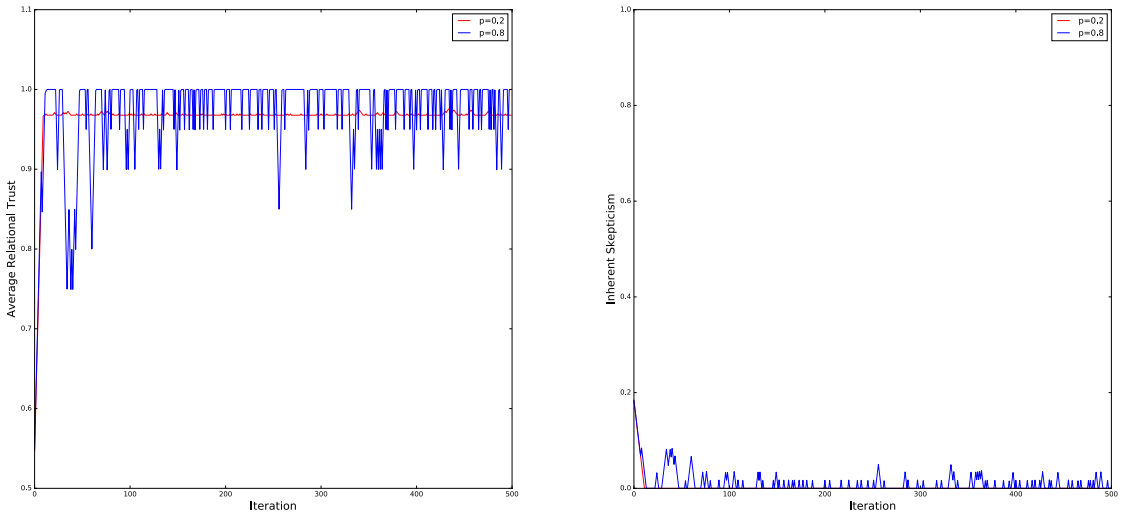

Figure 3. Evolution of relational trust and skepticism over time, $\delta_0 = 0.2$ and $H = 0.5$.

As was apparent in our validation cases, the behaviors presented in Figures 3 and 4 are very similar, since institutional trust does not play a significant role when agents are highly aware of the consequences of orders. Therefore, we discuss Figure 3 exclusively.

We observe that in the case of bad leadership, inherent skepticism rapidly decreases to 0 and the average relational trust within the organization rapidly rises to approximately 0.9, before reaching a stable equilibrium. This behavior occurs because the agents at the second level of the hierarchy quickly realize that the orders emanating from the root node are generally bad, which they can identify as a result of the significant amounts of information available to them, and thus begin to distrust the root node completely, making decisions based entirely on their own available information rather than their trust in the root node.

As a consequence of this behavior, the relational trust between the root node and its subordinates quickly converges to 0.

All other orders received by agents at the remaining levels of the hierarchy are always good, since any bad orders are corrected by the agents at the second level. This causes relational trust to converge to 1 between the remaining subordinates and their supervisors. When combined with the fact that the relational trust between the root node and its subordinates converges to 0, we obtain an average of approximately 0.9, which stabilizes.

In the case of good leadership, however, we see that while skepticism rapidly decreases to values around 0 and relational trust rapidly increases to values around 0.9, we do not observe an equilibrium state, with relational trust and skepticism tending to oscillate. This is a result of the agents at the second level of the hierarchy building up trust in the root node, unlike in the case of bad leadership. Since the root node sometimes sends bad orders, however, these subordinates will sometimes allow a bad order to propagate due to this trust which may, at times, supersede their judgment based on available information. This sometimes results in bad outcomes and a sudden increase in inherent skepticism.

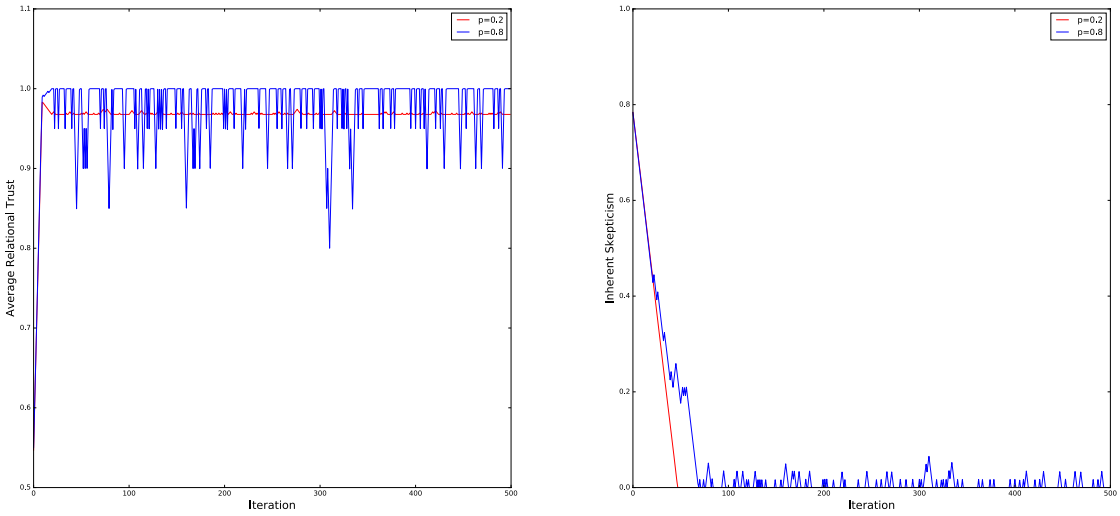

Figure 4. Evolution of relational trust and skepticism over time, $\delta_0 = 0.8$ and $H = 0.5$.

Therefore, we see that in an organization with poor leadership where the majority of members are well-informed, overall organizational stability can still emerge, with the organization's leader becoming more a figurehead than a practical component of the organization. In contrast to this, we have demonstrated that while a leader with a good track record would generally be considered desirable, any mistakes made by such leaders may eventually propagate through the organization and cause instability, since trust may begin to supersede the understanding of consequences within the organization.

Low Information Cases

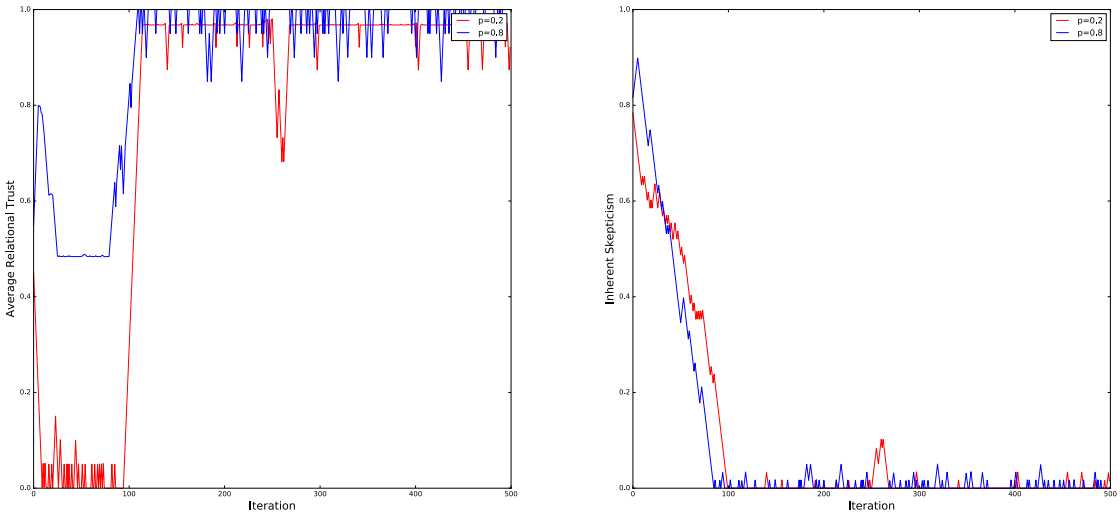

Figure 5. Evolution of relational trust and skepticism over time, $\delta_0 = 0.8$ and $H = 0.1$.

We now continue with an analysis of test cases involving low levels of information among agents ($H = 0.1$).

We observe in both Figure 5 and Figure 6 that despite individual agents having limited knowledge of the consequences of the orders received, by determining who to trust within the organization over time, the skepticism of agents eventually decreases to zero, regardless of its starting level, though at a slower rate than in the high information cases. This behavior occurs in cases

of both good and bad leadership.

This therefore suggests that regardless of the quality of leadership, by observing outcomes and developing relationships, the organization can eventually develop a robust resilience to poor orders, even if individual agents are not particularly well-informed about orders themselves. This demonstrates the pivotal role of trust and mistrust, especially in organizations where disobedience can occur if perceived to be for the overall good of the organization.

While good leadership is therefore not essential for overall success, we do observe that in both 5 and Figure 6 that good leadership leads to higher average levels of relational trust within the network, a fairly intuitive result.

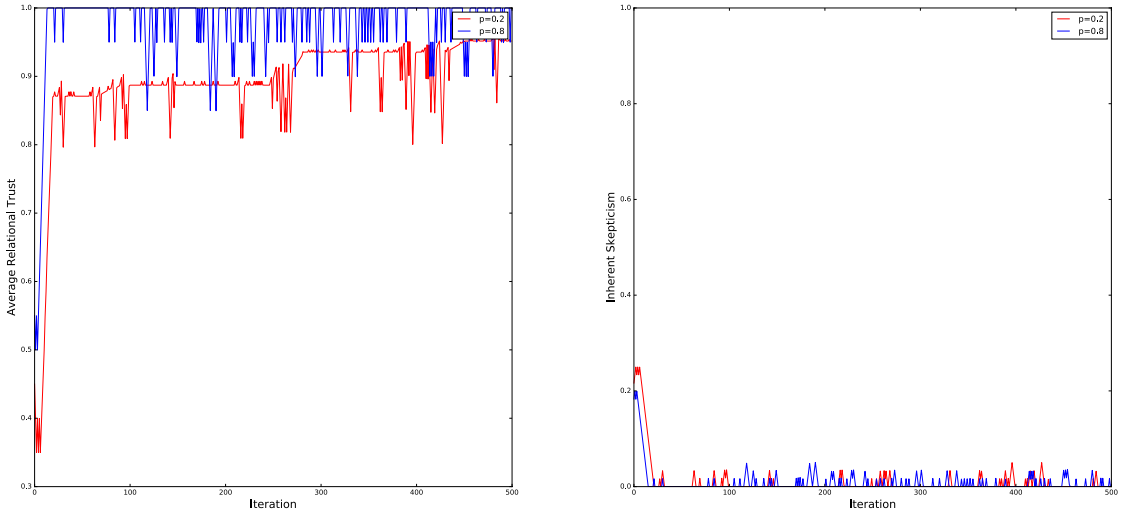

Figure 6. Evolution of relational trust and skepticism over time, $\delta_0 = 0.2$ and $H = 0.1$.

Possible Extensions

A possible extension is to explore whether the build-up and dissipation of trust occur at the same speed. A common notion is that trust develops incrementally over time, but can be lost relatively quickly when it is breached. A further area for study is whether trust decays with time if there is no fresh stimulus to sustain it. These extensions could have implications for how relational and institutional trust are modeled.

More work could be done on the drivers of institutional trust. Garfinkel[2] posited that institutional arrangements function like a personal guarantor to establish a "world-in-common", i.e. shared explicit and tacit knowledge between the trustor and the trustee that reduces the risk of breaches of trust. Such institutional structures may include laws or regulations, professional standards or codes of conduct that are not legally binding, reputation, and other formal or informal norms of behavior. It would be interesting to study whether institutional trust varies over time and across individuals and the impact that could have on the organization. For example, how does punishment for flipping orders affect behavior and do observations of order flipping by others embolden a person to do the same?

Another area for further investigation is how a supervisor's trust in his/her subordinates and the institution affects his/her behavior and the implications for the organization. For example, how would a supervisor respond to a subordinate who frequently flips orders? Do supervisors tend to have more or less institutional trust than their subordinates? Does asymmetry in trust levels, i.e. agent i trusts agent j more than agent j trusts agent i, affect the ability of the organization to achieve its mission?

Interactions between sub-units within an organization could be a further area of study. This could yield useful insights on the flow of information (or misinformation) and the emergence of norms that could have implications for trust levels and organizational outcomes.

It may also be possible to apply the model, with suitable adaptations, to other contexts where trust could be important, with schools, emergency services and financial markets being prominent examples.

Conclusion

In this paper, we have constructed a relatively simple toy model capable of recovering a number of intuitive behaviors associated with the role of trust within a hierarchical organization and used it to gain a number of insights into the development and nature of trust within such organizations.

As one would expect, we found that in organizations where the information accessible to members is relatively sparse, which is likely the case in many hierarchical organizations where the upper echelons of leadership have sole access to most organizational secrets, that good leadership inspires an overall higher level of relational trust within the organization than in the case of poor leadership.

Despite this, however, we have demonstrated that in our hypothetical, hierarchical organization that good leadership is not a prerequisite for organizational success, provided that members of the organization are allowed to disobey orders if perceived that doing so is in the interests of the organization. We showed that in such organizations, agents are capable of learning to identify who and who not to trust based on iterative adjustments of trust based on the overall outcome of specific tasks or missions, eventually resulting in a robust organization in the cases of both good or bad leadership.

In the case of highly-informed agents, we demonstrated that bad leadership may lead to greater stability, with agents tending to make decisions based on available information rather than trust, eventually relegating the leader to a symbolic role. In contrast to this, good leadership may result in the development of trust which is desirable in most cases, but may also result in the propagation of bad orders if a historically good leader unexpectedly gives a bad order, which may lead to disaster for the organization.

References

1. Allen, C. D. & Braun, W. G. Trust: Implications for the Army Profession. Conference for the International Society for Military Ethics, San Diego, USA (2012).
2. Garfinkel, H. A. Conception of, and Experiments with "Trust" as a Condition of Stable, Concerted Actions in *Motivation and Social Interaction. Cognitive Determinants* (ed. Harvey, O. J.) 187-238 (Ronald Press Co, 1963).
3. Panayi, E., Harman, M. & Wetherilt, A. Agent-based modelling of stock markets using existing order book data in *Multi-Agent-Based Simulation XIII* (ed. Giardini, F. & Amblard, F.) 101-114 (Springer-Verlag, 2013).
4. Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. Not so different after all: A cross-discipline view of trust. Acad. Manage. Rev. **23**(3), 393-404 (1998).
5. Twyman, M. Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. Judgm. Decis. Mak. **3**, 111-120 (2008).

Author Contributions

D.P. designed the experiments, wrote the result, model, validation and conclusion sections of the paper and contributed to model design. S.H. defined the original line of investigation and implemented the model in Python. N.W. performed a survey of literature and wrote the introduction and possible extension sections of the paper. W.L., X.X., R.M. and L.L. contributed to the model design and discussion of results.

Additional Information

The views expressed in this paper are those of the authors and do not necessarily reflect those of their respective institutions. The model is still very much in need of development and should thus not be taken to provide immediately applicable solutions to real-world problems.