The Trust Dilemma: When Trust Impedes New Information Flows in an Organization

Mario Molina[1] and Jiin Jung[2]

[1] Department of Sociology, Cornell University
[2] Department of Psychology, Claremont Graduate University

1 Overview

In the present paper, we investigate how much volume of *new* information flows among members in an organization when they build trusting relationships over time. Although trust is usually thought to bring benefits to social relations – in terms, for instance, of the reduced risk of sharing valuable information among team members in organizations –, we argue that excessive levels trust may prevent individuals from interacting with others outside of their tight-knit clique and thus impede sharing and acquiring new information from other members. In fact, high levels of trust help increase group cohesion among their members, but this same fact makes information redundant among them over time, so that no new information will be flowing among group members. Since group members are usually more likely to interact with each other, their group cohesion prevents them from being exposed to new information and hence the total volume of new information that circulates in an organization will be lower than it should for the organization to benefit from it. We call this social phenomenon the *Trust Dilemma*.

From an individual perspective, trust is "a psychological state comprising the willingness to be vulnerable under conditions of risk and interdependence based upon positive expectations of the intentions or behavior of another" (Rousseau, Sitkin, Burt, & Camerer, 1998). Trust reduces the risk of sharing information (Penley & Hawkins, 1985). People share more useful, wide range of, and divergent knowledge when they trust one another (Kijkuit & van den Ende, 2011, Levin & Cross, 2004, Sosa, 2011). Trust-based social ties in work teams spur creative ideas and workplace innovation (Dokko, Kane, Tortoriello, 2014; Sosa, 2011). From a group perspective, however, trusting relations shape the structure of a group network as trusted members develop ties that become stronger via repeated social interaction and cooperation (Baldassari 2015; McAllister, 1995; Coleman, 1990). This suggests that close-knit groups will have over time only redundant information available to share with minimal access to other sources of new information. An extensive literature in sociology has shown that, while dense networks allow for social capital to emerge in the form of social norms that can be enforced (Coleman 1988), strong ties provide less opportunities to access new information (Granovetter 1973; Burt 1992; Aral and Alstyne 2011). Weak ties have the advantage of connecting socially distant members that may have other, new sources of information. However, as trust becomes necessary to share particularly special information, the flow of all new information available in a given organization will probably be restricted.

When strongly-tied, cohesive subgroups form within an organization, they may become information cocoons which can obstruct the flow of information across the boundaries of those subgroups. Maybe in political contexts, political minorities may sustain their existence by forming tight-knit subgroups can increase ideological diversity. But in organizational contexts, it can damage knowledge transfer across an entire organization, thus deteriorate organization performance.

We build a model in which (a) agents interact and share information; (b) the more agents interact and share information that is attractive for them, the more they can trust each other and share new, private information that one would only share with selective individuals. As agents interact more and more, they increase their probability of future interactions with the same agents, thus allowing a network structure evolve by forming cliques-type groups in which individuals trust to each other. This probability of future interactions is crucial in our model because it shows how attractive to each other individuals are within an organization. (c) We assume that agents have a trust threshold that controls whether agents are willing to exchange new information with other agents or not. Importantly, we assume that these thresholds are exogenously imposed and homogeneous in the population. Future versions of this paper will relax these assumptions. (d) When the probability of future interactions meets the trust threshold, agents expose their own unique information and interacting agents learn this new unique information and store them in their memory. This new information then becomes available for agents to share in future interactions with other trusted members. This stage is critical as it allows the flow of new information within an organization.

To inform our theory we measure two key system-level properties: the amount of the flow of new unique information in the organizational system and the network structure that agents form and that ultimately allows information flows. We explore how different trust thresholds affect these measures.

2 Model

We construct an algorithm with the rules of agents' interaction and informational sharing in an organization and implement the algorithm with minimal assumptions, and simulate the model to explore how this lead to network structures particularly the formation of trust-cliques within an organization.

2.1 Agents Properties

We define two types of informational sets: shareable items and unique items. Shareable items are informational pieces that reveal cultural profiles of agents. Unique items, on the other hand, represent bits of information that are completely private and not shared unless the trust threshold is met (or exceeded). Shareable items and unique items are very different and constitute different types of information: while shareable items reveal how similar to one another agents are in terms of their cultural profiles, unique items are bits of information that can be transferred to other trusted agents. Shareable items cannot be learned and then transferred and they simply signal a particular cultural profile that can be attractive or not for other agents. We elaborate this idea in more detail below. In our model, the shareable set includes fours elements: the letters {A B C D} and the unique set including 22 elements {E F G H I J K L M N O P Q R S T U V W X Y Z}.

Each agent (a^j) has 2 sets of information with different elements: the shareable set has 2 elements randomly picked from the 4-item shareable set and unique set has 1 element is randomly picked from the 22-item unique set. In total, each agent has 3 information items, of which 2 signal an agent's cultural profile and 1 is a private bit of information that is transferable to other agent under certain circumstances.

Critically, we assume that when agents share shareable items, this information is not learned by other agents and therefore not stored in their memory. We consider shareable information as cultural attributes (e.g., demographics, nationality, natural inclination, propensity, personality) that cannot be (easily) adopted. This signaling process makes agents more or less attractive to each other depending on whether they are similar. We define similarity here as selecting the same item to share. In contrast, when agents share unique items, this information is readily added to target-agents' information set and can therefore be passed along in an organization's network. We think of unique items as work-related type of information that is directly relevant to organizational performance.

2.2 Multi-agent Systems

In our basic setup, 30 agents populate a fully-connected social network. We assume a context of team or organization where all members are acquainted with one another and possible to approach any one for potential information sharing. The initial weight of all ties will have the same value, $\frac{1}{(n-1)}$, which is the probability that two agents interact with each other. This basically assumes that agents initiate an interaction by selecting other agents with the equal probability and hence not giving any preference to some agents.

2.3 Interaction Rules

2.3.1 Intitial nomination and interaction

In an initial iteration, each agent nominates a target agent they want to interact with. Because there was no past interaction, the initial nomination will be uniformly random, such that all the other agents have $\frac{1}{(n-1)}$ probability to be nominated by a focal agent.

If two agents mutually nominate each other, they pair up and initiate interaction. If an agent does not get nominated back by the target agent, it stays alone and does not initiate interaction. The focal agent records past interactions with all the other agents, and after each iteration, it calculates relative past interaction with an interacting target agent and all the other agents. We assume that an increasing number of interactions with a target agent will always increase the probability of interaction with that agent. However, past interactions with a particular agent are assessed in their importance relative to all other past interactions. This introduces the following measure:

$$\delta_{ij\ (t)} = \frac{\delta_{ij\ (t)} + 1}{\sum_{i=1}^{n-1}\left(\delta_{ij\ (t)} + 1\right)} f \ \ \delta_{ij\ (t)} \geq 1$$

This parameter captures how important are past interactions with a target agent relative to all other past interactions that a focal agent has had before. Because the more agents interact, the smaller the relative past interaction, in order to make the value oscillate around 1, we add 1 to the relative past interaction as below:

$$\delta_{ij\ (t)} = \frac{\delta_{ij\ (t)} + 1}{\sum_{i=1}^{n-1}\left(\delta_{ij\ (t)} + 1\right)} + 1 f \ \ \delta_{ij\ (t)} < 1$$

2.3.2 Intitial information sharing

Between paired agents, focal agent (a^j) shares one information randomly picked from the two shareable information set. And a target agent (a^i) also shares one information randomly picked from the two shareable information set. If the two shareable information are the same, agent (a^j) and (a^I) become more attractive to each other for future interactions, which introduces a *sharedness score* (γ_{ij}) between self-agent a^j and target agent a^I:

$$\gamma_{ij(t)} = \gamma_{ij(t-1)} + \alpha$$

This parameter increases by α, which is externally determined. In our model, we use a fixed value of $\alpha = 0.05$. If the two shareable information are different, the *sharedness score* (γ_{ij}) between self-agent a^j and target agent a^i decreases by α.

$$\gamma_{ij(t)} = \gamma_{ij(t-1)} - \alpha$$

The self-agent a^j updates and remembers the sharedness score with the target agent a^i.

2.3.3 Following nomination and interaction

After the initial iteration, each agent nominates a next interacting agent based on a probability of interaction that is calculated using δ_{ij} and sharedness score (γ_{ij}). First, we calculate $P_{ij(t)}$

$$P_{ij(t)} = P_{ij(t-1)} \cdot \delta_{ij(t)} \cdot \gamma_{ij(t)}$$

Since $P_{ij(t)}$ only accounts for the target agents a focal agent has interacted with, $P_{ik(t)}$ is evaluated as the complement of $P_{ij(t)}$, where k represents any other target agent not interacted with, at one particular time point.
Then, for each focal agent we normalize probability values for all the other target agents as follows:

$$P_{ij\ (t)} = \frac{P_{ij\ (t)}}{p\Re Pj\ (t) + \sum_{i=1}^{n-2} P_{kj(t-1)}}$$

$$P_{kj(t)} = \frac{P_{kj(t-1)}}{p\Re Pj\ (t) + \sum_{i=1}^{n-2} P_{kj(t-1)}}$$

Then, agents use these updated probabilities to select other agents for interaction.

Importantly, these probabilities are dynamically changing and therefore shaping the network structure agents are embedded in as they represent edge weights in a social network.

2.3.4 Following information sharing

At each iteration, after a focal agent picked target agent, focal agent compares P_{ij} to a *trust threshold* (λ). Because initial P values vary depending on a group size such that the larger a group size, the smaller $P_{initial}$, we parameterize (λ) as a function of $P_{initial}$ as below:

$$\lambda = \theta \cdot P_{initial} \qquad \theta \in \{0.5\ 1.5\ 2.0\ 2.5\ \ldots\ n\text{-}1\}.$$

If $P_{ij} < \lambda$, a focal agent shares only from shareable information set. If $P_{ij} \geq \lambda$, a focal agent opens up their unique information set and share one information item from its entire information set, which includes both the shareable and the unique sets. If the focal agent shares a unique information, a target agent incorporates the unique information into its unique information set. However, if the self-agent shares a shareable information, a target agent does not add it to its set. Importantly, we assume that sharing a unique item increases members attraction regardless of whether unique items match.

For example, let us say a^j = {B C H} and a^i = {B C R}. After some interactions and information sharing, suppose that focal agent has $P_{ij} = 0.6$ which is smaller than a trust threshold, 0.7. Focal agent will then share from their shareable information set. But if $P_{ij} = 0.8$, which is larger than the trust threshold 0.7, then a focal agent shares from the unique information. If a^j selects {H}, then the target agent now has a^i = {B C R H}.

2.3.5 Measurement of information flow in a system

We keep track of the contents of information that has been shared at a system level.

$$New\ information\ flow = \frac{\sum_{t=0}^{m} N(information\ newly\ shared\ at\ t)}{total\ N\ (information \in a\ system)}$$

where m = # iterations. This measures basically captures the amount of new information shared in a system at every iteration relative to all the new information available in the system.

2.4 An Example

For simplification, and to make this process transparent, let's consider the case that there are three agents in a system thus $P_{initial}$ = 0.5, each agent has two shareable and one unique information, Sharedness score is capitalized or penalized by α = 0.2. A trust threshold is 1.5 times larger than the initial P ($\lambda = \theta \cdot P_{initial} = 0.75$; θ=1.5).

2.4.1 An initial condition at t=0

To demonstrate these processes we consider a minimal case example where there are three agents in a team. a^1 = {B C H}, a^2 = {B C R}, a^3 = {A B P}. There are total six information bits available in the system is {A B C H R P}.These agents have two shareable information and one unique information in their information set. Initially at t = 0, they are all connected to one another and a weight of the three ties is all 1/3. They have no past interaction, such thatδ is an empty matrix. Initial shareness scores (γ) for each pair is 1.

$$N(past\,interaction)_{t=0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\delta_{t=0} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$\gamma_{t=0} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

2.4.2 Nomination/interaction and information sharing at t=1

$$P_{t=1} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

Let's say a^1 randomly nominated a^3, a^2 randomly nominated a^1 and a^3 randomly nominated a^1 with an equal probability of 1/2. Because a^1 and a^3 mutually nominated each other, they pair up. Now they update a memory of past interaction and δ.

$$N(past\,interaction)_{t=1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\delta_{t=1}=\begin{pmatrix} 0 & 1 & \frac{2}{3}+1 \\ \frac{1}{3}+1 & 0 & \frac{1}{3}+1 \\ \frac{2}{3}+1 & 1 & 0 \end{pmatrix}$$

Next, a^1 randomly picks one shareable information out of two. Let's say, a^1 chose B and a^3 chose A. Because the two shareable information don't match, they update their sharedness score to 1 – 0.2 = 0.8. The new information flow score is 2/6.

$$\gamma_{t=1}=\begin{pmatrix} 0 & 1 & 0.8 \\ 1 & 0 & 1 \\ 0.8 & 1 & 0 \end{pmatrix}$$

2.4.3 Nomination/interaction and information sharing at t=2

After the first nomination and information sharing processes, agents calculate probability for next interaction for each of the other agents.

$$pre\ P_{t=2}=\begin{pmatrix} 0 & 0.5 & 0.26 \\ 0.5 & 0 & 0.5 \\ 0.26 & 0.5 & 0 \end{pmatrix}$$

$$P_{t=2}=\begin{pmatrix} 0 & 0.5 & 0.34 \\ 0.66 & 0 & 0.66 \\ 0.34 & 0.5 & 0 \end{pmatrix}$$

Now, agents nominate interacting partner based on $P_{t=2}$. Let's say a^1 nominated ^{2}, a^2 randomly nominated a^1 and a^3 nominated a^2. Because a^1 and a^2 mutually nominated each other, they pair up. Now they update a memory of past interaction and δ.

$$N(past\ interaction)_{t=2}=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\delta_{t=2}=\begin{pmatrix} 0 & 1.66 & 1.66 \\ 1.58 & 0 & 1.33 \\ 1.42 & 1.33 & 0 \end{pmatrix}$$

Next, a^1 randomly picks one shareable information out of two. Let's say, a^1 chose B and a^2 chose B. Because the two shareable information match, they update their sharedness score to 1 + 0.2 = 1.2. There is no new information addressed in the system, and the new information flow score remains the same, 2/6.

$$\gamma_{t=2}=\begin{pmatrix} 0 & 1.2 & 0.8 \\ 1.2 & 0 & 1 \\ 0.8 & 1 & 0 \end{pmatrix}$$

2.4.3 Nomination/interaction and information sharing at t=3

Agents update probability values for next interaction for all other agents.

$$preP_{\ t=3}=\begin{pmatrix} 0 & 0.996 & 0.34 \\ 1.25 & 0 & 0.66 \\ 0.34 & 0.5 & 0 \end{pmatrix}$$

$$P_{t=3}=\begin{pmatrix} 0 & 0.67 & 0.38 \\ 0.79 & 0 & 0.62 \\ 0.21 & 0.33 & 0 \end{pmatrix}$$

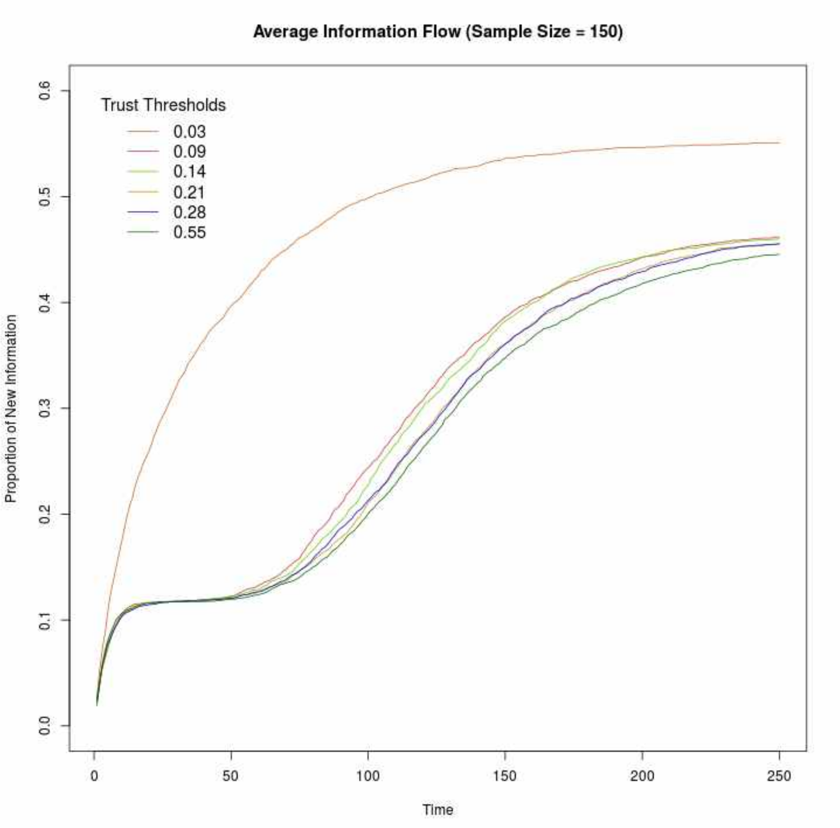
Now, agents nominate interacting partner based on $P_{t=3}$. Let's say a^1 nominated ^{2}, a^2 nominated a^1 and a^3 nominated a^2. Because a^1 and a^2 mutually nominated each other, they pair up. Now they update a memory of past interaction and δ.

Next, at the information sharing stage, because a^1*'s* trust toward a^2 (P_{21} =0.79) is larger than the trust threshold (0.75). a^1 opens up its unique information set and share one information randomly picked from the three. Let's say it share its unique information, H. However, a^2*'s* trust toward a^1 (P_{12} =0.67) does not reach the trust threshold (0.75). a^2 shares one of shareable information, C. Now a^2 adds the new unique information to its set, a^2 = {B C R H}. Two new information H and C were shared in the system, and the new information flow score is 4/6.

3 Running Simulations

We simulate our model to explore how a trust threshold affect the degree to which agents share *new unique* information and what is the shape of the network structure that allows or prevents this flow (i.e., formation of trust-cliques). For each iteration (1, … , *t*), we sample 150 and calculate the amount of new information flowing in the system. We average across these values. We run this process 250 times for different trust thresholds.

The plot below summarizes our results. In one particular case, we test a trust threshold that is right below $\frac{1}{(n-1)}$ - where $n = 30$ - that reflects a limit case in which all agents have a very low threshold to start sharing new information. In all other cases, the trust threshold is higher than this probability. As observed, in the case of a trust threshold that is lower than the value to uniformly select a partner, new information flows easily in an organization over time with a decay at iteration = 100. But more interestingly, the other cases reveal the scenario where agents need to invest some time in social interactions that build up trusting partners. However, these higher levels of trust thresholds produce lower levels of new information flows in the system, with some interesting differences between trust thresholds.



In general, the higher the trust threshold, the less new information becomes available in the system for the benefit of the organization. In all these cases, shareable information flows easily in the system, as expected, after which the flow of information gets stuck for some time. This period of stagnation is necessary to build trusting relations. At the end of iteration 50, we observe that the flow of new information starts increasing again but a different rates. Higher levels of trust thresholds are generally different to each other, even when a minor change in probabilitty occurs. Compare, for example, the difference between the red line (λ=0.09) and the dark green line (λ=0.55) suggests that high levels of trust thresholds damage the flow of new information in the system and therefore affects organizational performance by avoiding the organization to to benefit from the new information items that agents possess.

4 Discussion

Our goal was to present a minimal model capable of embodying the key mechanisms of information flows across trusted members of an organization. Our model is motivated by what we call the *Trust Dilemma*, which is basically the puzzle that higher levels of trust in an organization damages the flow of new information and therefore its overall performance. On the one hand, trust relations are wanted and usually needed in an organization because they generally increase group cohesion. But on the other hand, these same beneficial relations can affect the flow of information in an organization by forming network cliques that make individuals interact with the same others and therefore be less exposed to opportunities where new information can become available.

As shown in our model simulations, new unique information flows are hurt by increasing trust thresholds. The model, however, are purported to have much wider and deeper application. In order to fully engage the theoretical literature and applied issues that relate to our theory, model expansions must be made. We are interested in finding ways to mitigate the *Trust Dilemma*. One effective way addressed in the literature is to exchange people across different teams within an organization (Kane, 2005), so that they are forced to form new trusting relations that can ultimately promote the exchange of new information. One potential problem, however, is that a newcomer in a team often fails to be accepted by oldtimers and thus as members become less attractive to each other the probability to exchange new information decreases with these interactions. But if we consider the shareable information as cultural elements and unique information as skill sets, for instance, our model predicts that a newcomer should have enough shareable information in common with oldtimers - which was flowing in the organization after certain time - so that they can develop trust to the level where they may be able to expose unique information.

A critical problem of our model is that it manipulates trust thresholds to explore how systemic information flow is affected. If we were to design a policy to improve this organizational flow purely based on our results, we would need to suggest that trust thresholds of an organization's members should be lower, which certainly does not make any sense. In future extensions of this work, we want to explore how certain feasible policies – such as the one proposed above - can impact the flow of new information. Our model is also limited in the sense that we do not allow multiple interactions to occur at a particular time. Only dyadic interactions take place. If we extend this to multiple interactions, we would be able to observe how the network structure evolves to form clear groups with multiple agents within an organization. Another interesting, more real-life extension of this model would be to have a distribution of trust thresholds in our population of agents and allow for these thresholds to endogenously change with repeated interactions. For instance, a straightforward extension can be easily implemented so that successful interactions, where information exchange took place successfully, can affect not only the probability of future interaction but also decrease the trust thresholds in a given group. This would allow agents more or less embedded in certain groups to modify their trust thresholds with respect to other members of that group as well.

References

Aral, S., & Van Alstyne, M. (2011). The Diversity-Bandwidth Trade-off. *Journal of Sociology, 117,* pp. 90-171

Baldassarri, D. (2015). Cooperative Networks: Altruism, Group Solidarity, Reciprocity and Sanctioning in Ugandan Farmer Organizations**. *American Journal of Sociology, 121*, pp.355-395**

Burt, R. (1992). *Structural Holes.* Harvard University Press, Cambridge, MA.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94(Supplement)*, S95-S120.

Dokko, G., Kane, A. A. & Tortoriello, M. (2014). One of us or one of my friends: How social identity and tie strength shape the creative generativity of boundary-spanning ties. Organization Studies 35(5): 703-726.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78, pp.1360-1380.

Kijkuit, B. & Ende, J. van den (2010). With a Little Help from Our Colleagues: A Longitudinal Study of Social Networks for Innovation. *Organization Studies, 31*, pp. 451-479.

Levin, D. Z., & Cross, R. (2004). The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. *Management Science, 50*, pp. 1477-1490

Penley, L., & Hawkins, B. (1985). Studying interpersonal communication in organizations: A leadership application. *Academy of Management Journal*, 28, pp.309-26.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. F. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23,* pp. 393-404.

Sosa, M. E. (2011) Where Do Creative Interactions Come From? The Role of Tie Content and Social Networks. *Organization Science, 22,* pp.1-21.