# Computationally Adjustable Cognitive Inertia

**Brian McLaughlan and Sebastian Bossarte**
Department of Computer and Information Science
University of Arkansas - Fort Smith, Fort Smith, AR, USA

**Abstract** - *Cognitive inertia is the tendency for beliefs to endure once formed. This paper proposes that cognitive inertia can be utilized in multi-agent systems as a first derivative of trust. In this role, cognitive inertia provides a mechanism for allowing agents to determine how quickly or how drastically they should re-evaluate their trust in other agents. Appropriate levels of cognitive inertia are experimentally determined for various combinations of high and low risk and reward scenarios. Methods are examined that can allow agents to alter cognitive inertia based on feedback from the environment and other agents. From these experiments, we have identified several variables that appear to be useful indicators of appropriate cognitive inertia as well as inertia determination methods appropriate for various generalized scenarios.*

**Keywords:** *Multi-agent systems, trust, cognitive inertia*

## 1 Introduction

The evolution of e-commerce has led to an increased interest in concepts relating to trust. In fact, the ability to computationally define trust has been a major factor in successfully developing electronic and online commerce [1]. These trust models are increasing used by automated agents for making decisions or suggesting courses of action. When looking for inspiration for developing these models it is useful to look beyond traditional artificial intelligence concepts and examine how research in other disciplines may apply.

One potentially useful concept, cognitive inertia, is commonly utilized in managerial science. Essentially, cognitive inertia is the human tendency to maintain previously valid beliefs even when new evidence no longer supports those beliefs. Typically, this concept has a negative connotation, and significant research in this topic is devoted to analyzing its effects [2] or circumventing its symptoms [3]. However, cognitive inertia does have a positive role in the maintenance of trust.

Cognitive inertia is the component that makes long-term relationships of trust possible. For instance, if an individual is considered a trustworthy friend, one mistake would not be enough to invalidate that friendship. Additional confirmation of untrustworthiness is often needed. However, that one offense may prompt a slight raise of one's guard. Additional offenses would then indicate a significant change in attitude toward the individual in question. In this sense, cognitive inertia can be considered the first derivative of trust. That is, it is the rate at which trust is modified when circumstances dictate a change.

Although trust can be defined in many ways, this paper defines trust as follows: Trust is a belief that another agent is reliable in the services it provides, and is honest when given the opportunity to defect [4].

While cognitive inertia is not a one-size-fits-all concept in which a single value is most appropriate in all situations, it may be possible to generate broad guidelines for determining how quickly an agent should alter its trust in a given situation. This ability to modify its reasoning methods in response to feedback from the environment and scenario could allow an agent to avoid costly mistakes from trusting a faulty source or rashly terminating an otherwise sound relationship.

There are many techniques for modeling trust. The techniques in this paper could be best categorized as a learning model, a famous version of which is based on game theory [5] where agents calculate the benefit of cooperating with another agent. However, this work is less like traditional game theory models involving payoffs and defections and is more like the emergent trust models taken from a fusion of complexity theory, marketing, and psychological theory [6].

The research presented in this paper addresses two questions regarding cognitive inertia. First, is it possible to determine broad rules for calculating an appropriate inertia value in a given situation? Second, is it possible to utilize environmental variables, agent performance, or self-examination to adjust inertia to more appropriate levels?

## 2 Approach

To explore the viability of the concept of cognitive inertia in trust relationships, a generic multi-agent system is proposed. In this system, agents broadcast information to each other. Information presented by some agents tends to be more reliable than information from other agents. Believing false information incurs a penalty while believing true information provides a reward. Conversely, ignoring false information is rewarded while ignoring true information is penalized.

Agents maintain trust levels in each other agent. These trust levels represent the percentage chance that the agent will believe the information provided (Figure 1). After receiving the reward or penalty, the agent has the opportunity to adjust

its trust in the sending agent based on its cognitive inertia. While an agent maintains trust levels for each other agent, it only has a single cognitive inertia value. Future experimentation can determine if this is sufficient or if a more discrete inertia would be beneficial.

Formally, we propose a World $\boldsymbol{W}$ that consists of Agents $\boldsymbol{A}$, a Reward value $R$, a Penalty value $P$, and a discrete Time variable $T$:

$$\boldsymbol{W} = \{ \boldsymbol{A}, R, P, T \}. \quad (1)$$

There exists a set of agents, $\boldsymbol{A} = \{ \boldsymbol{A_1}, \boldsymbol{A_2}, \ldots, \boldsymbol{A_n} \}$, such that

$$\boldsymbol{A_i} = \{ \boldsymbol{M_i}, \boldsymbol{TR_i}, C_i, H_i, S_i, \boldsymbol{F_i} \}. \quad (2)$$

$\boldsymbol{M}_i$ is the set of message observations for $\boldsymbol{A_i}$:

$$\boldsymbol{M}_i = \{ M_i(1), M_i(2), \ldots, M_i(k) \}. \quad (3)$$

$\boldsymbol{TR_i}$ is the set of trust values in other agents:

$$\boldsymbol{TR}_i = \{TR_i^{\,1}, TR_i^{\,2}, \ldots, TR_i^{\,n-1}\} \quad (4)$$

where $n$ is the number of agents in the system. Trust for each other agent is a value from 1 to 100.

$C_i$ is agent $\boldsymbol{A_i}$'s cognitive inertia level on a scale of 1 to 100, representing the percentage chance it will adjust its trust in another agent. $S_i$ represents its fitness score, and $H_i$ is the agent's honesty on a scale of 1 to 100, representing the percentage chance it will present accurate information. Finally, $\boldsymbol{F_i}$ represents the set of functionality available to the agent.

$$\boldsymbol{F_i} = \{ F_i^S, F_i^R \} \quad (5)$$

Each turn, agent $\boldsymbol{A_i}$ will send a true message, $M^{true}$, or a false message, $M^{false}$, to all other agents using the Send function, $F_i^S$:

$$\boldsymbol{M_x}(T+1), \forall\, x \in \boldsymbol{A}, x \neq i \leftarrow F_i^S(M_i^{true}(T)) \Leftrightarrow \quad (6)$$
$$Random(1{:}100) \leq H_i(\mathrm{T}).$$

$$\boldsymbol{M_x}(T+1), \forall\, x \in \boldsymbol{A}, x \neq i \leftarrow F_i^S(M_i^{false}(T)) \Leftrightarrow \quad (7)$$
$$Random(1{:}100) > H_i(\mathrm{T}).$$

Similarly, a Receiving function, $F_i^R$, is created for receiving messages from agent $\boldsymbol{A_j}$:

$$\boldsymbol{M_i}(T+1) \leftarrow F_i^R(F_j^S(x_j(T), \exists!\, x \in \{M^{true}, M^{false}\})). \quad (8)$$

Next, we define the method in which agent $\boldsymbol{A_i}$ believes or disbelieves a message from $\boldsymbol{A_j}$:

$$\mathsf{BEL}_i\,(M^x{}_j(\mathrm{T}), \exists!\, x \in \{true, false\}\,) \Leftrightarrow \quad (9)$$
$$Random(1{:}100) \leq TR_i^{\,j}(\mathrm{T}).$$

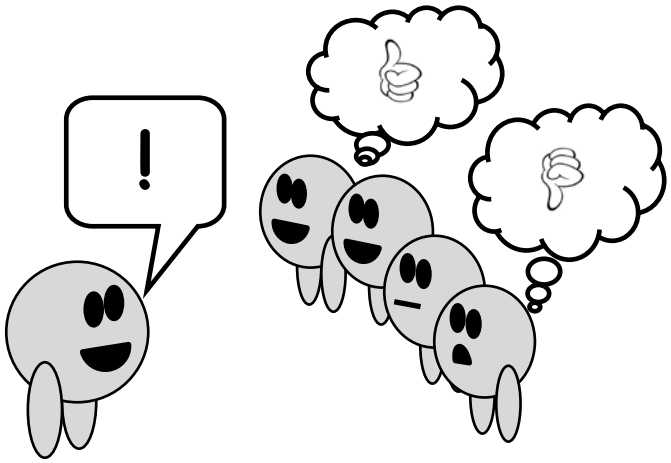

**Figure 1:** Agents believe or disbelieve messages based on their trust in the sending agent

That is, agent $\boldsymbol{A_i}$ will believe a message from agent $\boldsymbol{A_j}$ if and only if a randomly generated number from 1 to 100 is less than or equal to $\boldsymbol{A_i}$'s Trust in $\boldsymbol{A_j}$. While real-world scenarios would require a significant number of variables to adequately compute the trustworthiness of a message, this experiment abstracts the uncertainty as a stochastic variable.

Next, we define the methods in which the agent is rewarded or punished for the messages it has received:

$$\mathsf{BEL}_i\,(M^{true}(T)) \vee \neg\, \mathsf{BEL}_i\,(M^{false}(T)) \Rightarrow \quad (10)$$
$$S_i(\mathrm{T}+1) \leftarrow S_i(\mathrm{T}) + R.$$

$$\mathsf{BEL}_i\,(M^{false}(T)) \vee \neg\, \mathsf{BEL}_i\,(M^{true}(T)) \Rightarrow \quad (11)$$
$$S_i(\mathrm{T}+1) \leftarrow S_i(\mathrm{T}) - P.$$

When agent $\boldsymbol{A_i}$ believes a true message or disbelieves a false message, its score increases by the reward amount. Conversely, when agent $\boldsymbol{A_i}$ believes a false message or disbelieves a true message, its score decreases by the penalty amount.

Finally, we define agent $\boldsymbol{A_i}$'s ability to alter its trust in agent $\boldsymbol{A_j}$:

$$S_i(\mathrm{T}+1) > S_i(\mathrm{T}) \wedge Random(1{:}100) \leq C_i(\mathrm{T}) \Rightarrow \quad (12)$$
$$TR_i^{\,j}(\mathrm{T}+1) \leftarrow TR_i^{\,j}(\mathrm{T}) + (100\text{-}C_i)/10.$$

$$S_i(\mathrm{T}+1) < S_i(\mathrm{T}) \wedge Random(1{:}100) \leq C_i(\mathrm{T}) \Rightarrow \quad (13)$$
$$TR_i^{\,j}(\mathrm{T}+1) \leftarrow TR_i^{\,j}(\mathrm{T}) - (100\text{-}C_i)/10.$$

# 3 Experimentation

The proposed model was implemented in a Java simulation. The simulation consisted of 100 agents, each with a different cognitive inertia value:

$$C_i \leftarrow i, \forall\, i \in \boldsymbol{A}. \quad (14)$$

Each agent was given a random honesty value:

$$H_i \leftarrow Random(1{:}100), \forall\ i \in \boldsymbol{A}. \quad (14)$$

The honesty values for the group were weighted in one of three ways:

1. No weighting of honesty
2. Honesty weighted towards 75
3. Honesty weighted towards 25

Also examined was the model's ability to handle different reward and penalty scenarios. The experiment was executed with one of three reward/penalty combinations

1. Equal reward and risk: Reward $R = 1$, Penalty $P = 1$
2. High reward, low risk: Reward $R = 5$, Penalty $P = 1$
3. Low reward, high risk: Reward $R = 1$, Penalty $P = 5$

These parameters allow many different situations to be simulated. For instance, agents might be attempting to discover leads among data records for rooting out a wanted fugitive. In this scenario, there may be many dead-end leads and false positives shared with the group (honesty scores weighted towards 25), and those false positives don't cause significant problems (low penalty). However, a positive lead is a rare and significant event (high reward).

As another example, consider a group of agents that are interpreting and utilizing targeting data. Agents are expected to perform adequately (low reward and honesty weighted toward 75), and errors can cause catastrophic events such as targeting of friendly troops or non-combatants (high penalty).

The experiment was executed with one of five modifications to the method for adjusting cognitive inertia after each received message:

1. No change to inertia possible

2. Change inertia upwards for true received messages, downwards for false messages

$$F_i^R(M^{true}(T)) \Rightarrow C_i(T+1) \leftarrow C_i(T) + 1 \quad (15)$$
$$F_i^R(M^{false}(T)) \Rightarrow C_i(T+1) \leftarrow C_i(T) - 1 \quad (16)$$

3. Change inertia downwards for true received messages, upwards for false messages

$$F_i^R(M^{true}(T)) \Rightarrow C_i(T+1) \leftarrow C_i(T) - 1 \quad (17)$$
$$F_i^R(M^{false}(T)) \Rightarrow C_i(T+1) \leftarrow C_i(T) + 1 \quad (18)$$

4. Inertia moves toward 50 when a false message is received, away from 50 when true

$$F_i^R(M^{true}(T)) \wedge C_i(T) < 50 \Rightarrow \quad (19)$$
$$C_i(T+1) \leftarrow C_i(T) - 1$$
$$F_i^R(M^{true}(T)) \wedge C_i(T) > 50 \Rightarrow \quad (20)$$
$$C_i(T+1) \leftarrow C_i(T) + 1$$
$$F_i^R(M^{false}(T)) \wedge C_i(T) < 50 \Rightarrow \quad (21)$$
$$C_i(T+1) \leftarrow C_i(T) + 1$$
$$F_i^R(M^{false}(T)) \wedge C_i(T) > 50 \Rightarrow \quad (22)$$
$$C_i(T+1) \leftarrow C_i(T) - 1$$

5. Inertia moves toward 50 when a true message is received, away from 50 when false

$$F_i^R(M^{true}(T)) \wedge C_i(T) < 50 \Rightarrow \quad (23)$$
$$C_i(T+1) \leftarrow C_i(T) + 1$$
$$F_i^R(M^{true}(T)) \wedge C_i(T) > 50 \Rightarrow \quad (24)$$
$$C_i(T+1) \leftarrow C_i(T) - 1$$
$$F_i^R(M^{false}(T)) \wedge C_i(T) < 50 \Rightarrow \quad (25)$$
$$C_i(T+1) \leftarrow C_i(T) - 1$$
$$F_i^R(M^{false}(T)) \wedge C_i(T) > 50 \Rightarrow \quad (26)$$
$$C_i(T+1) \leftarrow C_i(T) + 1$$

To allow each agent to stabilize at what it felt was an appropriate configuration, the simulation was executed for 500 time cycles, with each agent sending one message to the group during each time cycle. This number of time cycles gave each agent adequate time to settle into particular trust and cognitive inertia values. At the end of the simulation, fitness scores of each agent were tabulated. Each of the 45 possible scenario combinations was executed 1000 times to get a good average statistic.

# 4 Results

Figure 2 shows the average scores of agents in a simple low risk/low reward scenario containing agents with a wide range of honesty values. The "No modifier" line shows the baseline desirability of each cognitive inertia value. Thus, a low inertia would be most appropriate here. Agents with an honesty value that clusters around 25 tend to perform similarly. However, agents with high average honesty (Figure 3) tend to perform better with a cognitive inertia in the 65 to 80 range. This trend holds fairly steady for all combinations of high/low risk and reward parameters.

The method of adjusting cognitive inertia is particularly interesting. As summarized in Figure 4, the formula that achieves the best performance varies by scenario. When the average honesty is high, pushing the inertia towards the

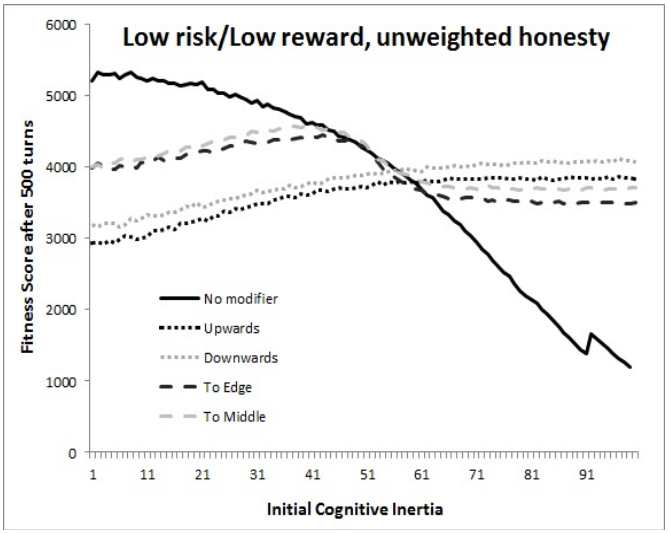


**Figure 2:** Comparison of inertia modification techniques in a community with random honesty

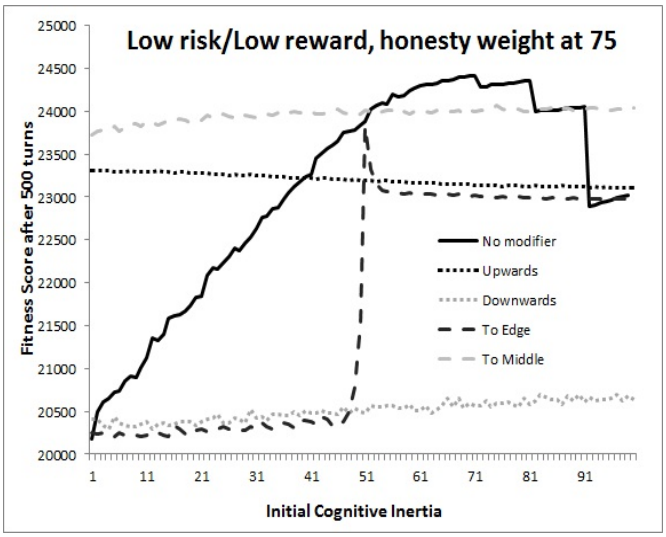


**Figure 3:** Comparison of inertia modification techniques in a community with honesty weighted towards 75

middle when a true statement is received and towards the edge when a false statement is received works best. This makes sense when considering that an ideal inertia value is slightly above 50, and most statements will be true.

However, other scenarios are less defined. Clustering scores to the edge or middle tends to provide adequate results, but linear pushes upwards or downwards tends to either be very good or very poor. Low risk/high reward scenarios tended to do poorly with linear movement, while high risk/low reward scenarios tended to do excellently with these formulas. As before, high honesty systems didn't follow this pattern.

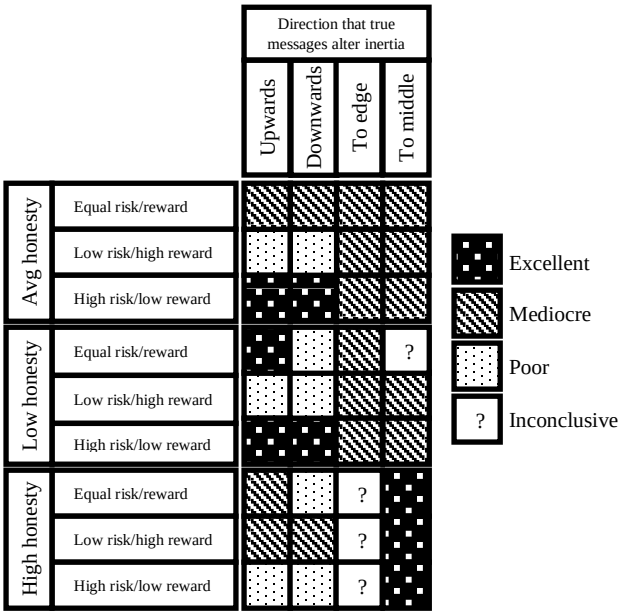


**Figure 4:** Summary of inertia modification techniques in various scenarios

# 5 Conclusions and future work

The simulations show that there are patterns in the data, suggesting that the notion of cognitive inertia is a valid method of adjusting trust in multi-agent systems. However, the anomalies suggest that a more complex representation is probably needed. For instance, agents may be better served by a trust vector, combining trust and cognitive inertia for each other agent in the system. Additionally, time may be a valid factor, with agents keeping track of how long they have held a particular trust in another agent. Another concept that may be of use is social distance [7] in which agents are not necessarily directly known to each other. Notions such as these could allow for new algorithms for adjusting trust inertia that might better serve the agents than ones presented here. Our future work in this topic will explore these possibilities.

# 6 References

[1] He, J. (2011) "Understanding the sources and impacts of trust in e-commerce: A meta-analysis." *AMCIS 2011 Proceedings – All Submissions.* Paper 142.

[2] Tripsas, M., and Gavetti, G. (2000) "Capabilities, cognition, and inertia: Evidence from digital imaging." In *Strategic Management Journal* 21:10-11.

[3] Messner, C., and Vosgerau, J. (2010) "Cognitive inertia and the implicit association test." In *Journal of Marketing Research*, 47:2.

[4] Dasgupta, P. (1998) "Trust as a commodity." In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*. Blackwell.

[5] von Neuman, J., and Morgenstern, O. (1944) *The Theory of Games and Economic Behavior*. Princeton University Press, Princeton NJ.

[6] Jarratt, D., Bossomaier, T., Thompson, J. (2007) "Trust as an emergent phenomenon in wealth management relationships." In *Global Business and Economics Review* 9:4, 335-352.

[7] Binzel, C., and Fehr, D. (2010) "Social Relationships and Trust," SFB 649 Discussion Papers, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.