Mathematical Model of Dual Opinion Formation: Opinion Dynamics for Promoting Informational Health - The Interplay Between True and Socially Expressed Opinions and the Path to Informational Symbiosis

Hatadani Shunsuke[1] and Yasuko Kawahata[2]*

[1]*Rikkyo University, 24-40 Asahi-cho, Minami-ku, Sagamihara-shi, Kanagawa 252-0304, Japan*

[2]*Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo, 171-8501, Japan*

*Corresponding author: Yasuko Kawahata, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo, 171-8501, JAPAN

Introduction

Creating Informational Health through the Interaction of True and Socially Expressed Opinions

In contemporary society, understanding the processes of opinion formation and their interactions is extremely important for social consensus building, policy decisions, and the health of everyday communication. This paper focuses on Ishii et al.'s opinion dynamics model, which has made significant contributions to the mathematical understanding of opinion formation. By explicitly incorporating the dual structure of "true opinions" (honne) and "socially expressed opinions" (tatemae), which is particularly evident in Japanese society, we can construct a communication model that more closely reflects reality. In Japanese, "honne" refers to one's true opinions and feelings that exist internally, while "tatemae" refers to opinions and attitudes expressed externally in con- sideration of social context and interpersonal relationships. This duality is not simply falsehood or deception but functions as cultural wisdom for balancing social harmony and individual autonomy. However, social consensus formed without adequate understanding of this dual structure may accumulate potential conflicts and dissatisfactions beneath a superficial harmony, risking social division and collapse of trust in the long term.

In this paper, we propose an extension of Ishii et al.'s trust-distrust model in which each agent maintains two internal states: "true opinion" and "socially expressed opinion." This extension enables more detailed analysis of opinion formation processes under the influence of social pressure, power structures, and group norms. Particularly noteworthy are the conversion mechanisms between "true" and "socially expressed" opinions and how they influence group-level consensus formation and conflict resolution.

Informational Health and Communication Integrity

The increasing complexity and diversification of information environments in modern society significantly influence individual opinion formation. In environments where vast amounts of information are shared instantaneously, the quality and reliability of information, as well as the recipient's information processing ability, are being questioned, highlighting the importance of the concept of "Informational Health." Informational health refers to the state in which individuals and groups can make rational judgments and decisions based on appropriate information, and it can be considered a basic element of modern well-being alongside physical and mental health.

An opinion dynamics model that explicitly considers the dual structure of true and socially expressed opinions provides important insights for achieving informational health. Appropriate balance between maintaining social harmony (tatemae) and expressing genuine individual opinions (honne) can promote informational health in the following ways:

- Substantiation of social consensus: Enabling consensus formation that properly considers diverse true opinions rather than merely superficial agreement.

- Visualization and resolution of potential conflicts: Recognizing the differences in true opinions hid- den behind socially expressed ones, creating opportunities for mutual understanding through dialogue.

- Improving communication transparency: Better understanding of the intentions and contexts of statements through explicit recognition of the relationship between true and socially expressed opinions.

- Diversification of information selection: Promoting learning from diverse information sources through recognition of people with different true opinions.

Grassroots Co-creation of Informational Health

Achieving informational health requires not only top-down information regulation and media literacy education but also grassroots efforts through interactions in daily communication between individuals. The opinion dynamics model proposed in this paper, which considers true and socially expressed opinions, provides clues to understand the mechanisms of such grassroots co-creation of informational health.

Specifically, the following mechanisms for grass- roots informational health creation can be considered:

- **Creating Safe Spaces for Expressing True Opinions**: It is important to form environments where people can express their true opinions and deepen mutual understanding with adequate psychological safety rather than complete anonymity. In simulations, such environments can be modeled by appropriate setting of the trust parameter d_{ij}.

- **Promoting Contextual Understanding of Socially Expressed Opinions**: Enhancing the ability to more accurately infer the true opinions and intentions behind socially expressed ones by deepening the understanding that expression through socially acceptable means is not simply falsehood but contains specific social contexts and considerations. This is expressed by adjusting the parameters α and β in the sigmoid function.

- **Utilizing Meta-Communication**: Promoting dialogue about the nature of communication itself (meta-communication) to pro- vide opportunities for parties involved to adjust the appropriate balance between true and socially expressed opinions. This can be expressed as an extension of the function F (t, i, c, A, m) in the model.

- **Securing Diverse Opinion Expression Channels**: Allowing different expression forms (direct and indirect) in the same venue, enabling flexible opinion expression according to situations and contexts. This is expressed as a diversification of the conversion function $f(I_i^{real}, S_i, P)$ between true and socially expressed opinions. These efforts can potentially lead to a cyclical improvement of informational health by interacting with each other. For example, an increase in safe spaces for expressing true opinions can promote contextual understanding of socially expressed opinions, which in turn enhances the quality of meta-communication, leading to more effective use of diverse opinion expression channels.

Introduction

This paper provides a detailed analysis of the consensus formation model developed by Ishii et al., examining its fundamental characteristics. Ishii et al.'s model is a kind of multi-state Ising model that can evaluate agreement and disagreement with infinite granularity [1,2]. However, comparisons of some parameters within the model are insufficient, and their functional meanings are not adequately explained. This paper aims to re-examine this model and provide a detailed analysis of its basic properties.

Research Background

The consensus formation model developed by Ishii et al. is a type of multi-state Ising model that can evaluate agreement and disagreement with infinite granularity. Ishii et al.'s research demonstrates the basic characteristics of opinion consensus formation in dyads, tri- ads, and medium-sized groups1) [3]. However, com- parisons of some constant parameters are lacking, and their functional meanings are not addressed. This paper aims to review Ishii et al.'s consensus formation model and examine its basic characteristics in detail. Research on opinion dynamics has been approached previously using methods based on sociophysics and statistical physics [4,5]. A new approach to understanding sociological collective behavior based on the framework of critical phenomena in physics began when it was first proposed that constructing a simple mean-field approximation model to apply to the strike process in social factories would be the first step [6].

Theory

Ishii's Opinion Dynamics Theory

From the literature [1,3], the change in opinion is expressed by the following equation:

$$I_i(t+\Delta t) = I_i(t) - \propto I_i(t)\Delta t + c_i A(t)\Delta t + \sum_{j}^{n} D_{ij}\Phi(I_i(t),I_j(t))\{I_j(t)-I_i(t)\}\Delta t \quad (1)$$

Here, $\Phi(I_i(t),I_j(t))$ is a sigmoid function defined as follows:

$$\Phi(I_i(t),I_j(t)) = \frac{1}{1+\exp\{\beta(|I_j(t)-I_i(t)|-b)\}} \quad (2)$$

Equation (1) represents how opinions change due to the influence of others. First, the range of I_i is $-\infty \le I_i \le \infty$, with I_i being positive for affirmative opinions, negative for negative opinions, and $I_i = 0$ de- fined as a neutral opinion. The second term on the right side of equation (1) is a forgetting term, which discards part of one's current opinion to update it. The third term on the right side represents influences received from sources other than dialogue, such as media [7,8]. A(t) represents influence, and the coefficient c_i is set individually as susceptibility to influence (Figure 1).
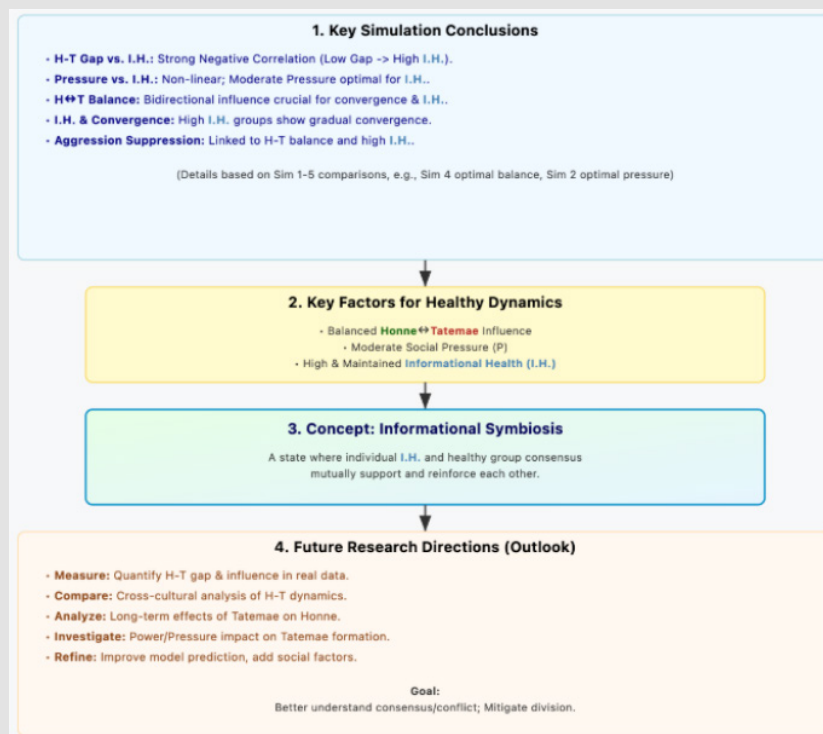

Figure 1: Research Brief: Findings, Synthesis, and Perspectives.

The $\Phi(I_i(t),I_j(t))$ in the fourth term on the right side is a sigmoid function expressed by equation (2), which becomes smaller as the difference in opinions between two agents becomes larger. The value of $\Phi(I_i(t),I_j(t))\{I_j(t)-I_i(t)\}$ becomes small when opinions are too close (small difference in opinions) and also when opinions are too far apart (large difference in opinions). The coefficient D_{ij} is defined as "trust toward the other party," with a larger value indicating more trust in the target and a smaller value indicating distrust [9,10]. Overall, equation (1) expresses the idea that "one's opinion changes under the influence of people whose opinions are moderately different from one's own, as well as media, etc."

Questions on Ishii's Opinion Dynamics Theory

Having explained the equation for opinion change, there are several questions about these equations. First, regarding the validity of the second and third terms on the right side of equation (1). Considering the second term, in the case where there is no influence from sources other than conversation, that is, when $c_i A(t)\Delta t = 0$, and when all participants in the dialogue agree and each person's opinion value is approximately the same, equation (1) becomes monotonically decreasing and converges to 0. While this seems natural just looking at the equation, it is unlikely that opinions naturally become neutral simply due to not being exposed to a topic/agenda for a sufficient amount of time (although there is room to consider changes in

opinion due to sudden mutations or changes in values, this is beyond the scope of this paper) (Figure 2). Also, while there is no explanation in the literature [1], it can be inferred from the graph that $\alpha = 0$, implicitly indicating that "opinions do not naturally become neutral." Thus, it is considered that the second term $-\alpha I_i(t)\Delta t$ in equation (1) should not exist.

Figure 2: Informational Health and Symbiosis Dashboard.
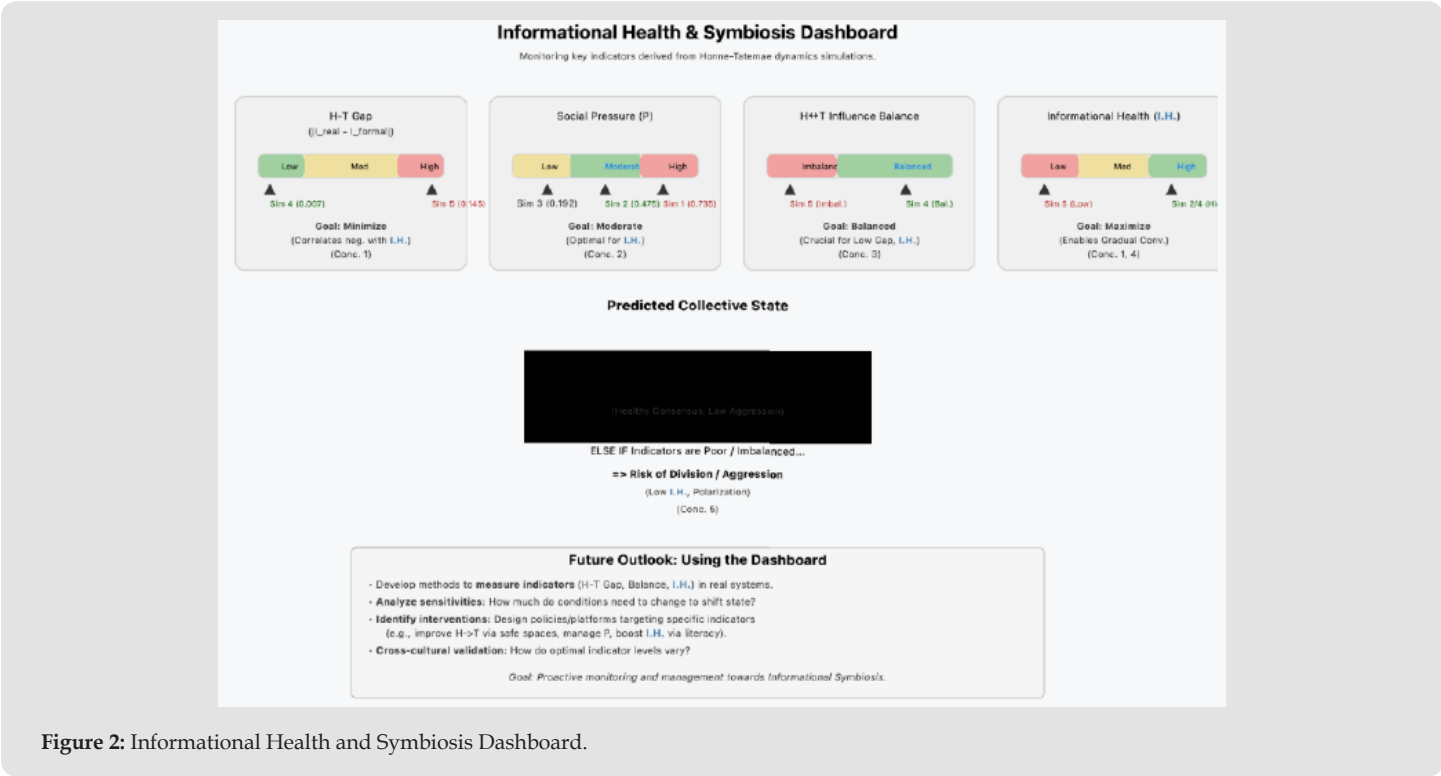
Next, consider the third term of equation (1) [11]. Similarly to the previous discussion, when all participants in the dialogue agree and each person's opinion value is approximately the same, taking the previous argument into account, equation (1) becomes a function dependent on $c_i A(t)$. This means that opinions change only by the amount of external influence, and if $c_i A(t) > 0, I_i(t)$ diverges to infinity, and if $c_i A(t) < 0, I_i(t)$ diverges to negative infinity. Such patterns might occur when considering unique patterns like extremists [12]. However, generally speaking, it is natural to think that even external forces like media have certain opinions [13], and that in cases where one is only influenced by and follows them, opinion values converge to a certain value rather than diverging infinitely, aligning with that external force (Table 1).

Table 1: Comparison of Communication Styles.

Communication Style	Advantages	Disadvantages
True opinion priority (direct expression)	Transparency, Reduced misunderstandings	Interpersonal friction, Disruption of social harmony
Socially expressed opinion priority (indirect expression)	Maintenance of social harmony, Smoother relationships	Avoidance of essential discussions, Accumulation of potential dissatisfaction
Appropriate balance between true and socially expressed opinions	Compatibility of essential discussions and social harmony, Promotion of informational health	Difficulty in maintaining balance, Cultural differences

Second, the issue is that constants set in equations (1) and (2) have not been given meanings. Specifically, these are α in equation (1) and β, b in equation (2) [3]. For α, based on the previous discussion, decay coefficient or forgetting coefficient would be appropriate, but this remains speculative without explicit docu- mentation. As for β, b in equation (2), the method for setting each value is not specified, making them determined by the researcher's discretion. While this is not necessarily bad in itself, I question the appropriateness of settling for researcher discretion when investigating basic properties (Figure 3).
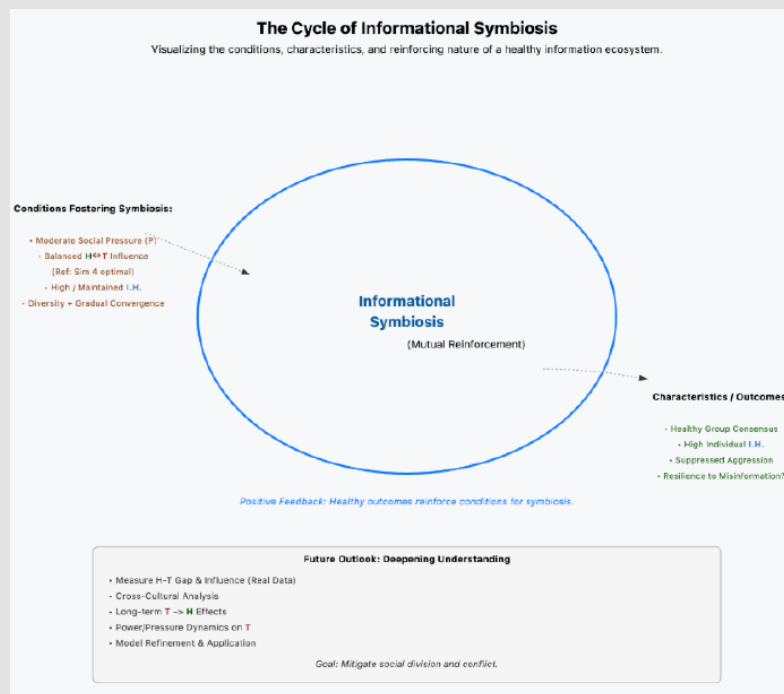
Figure 3: Cycle of informational symbiosis.

Improved Opinion Dynamics Theory

Based on the contents up to the previous section, equations (1) and (2) are improved as follows [14]:

$$I_i(t+h) = I_i(t) + hF(t,i,c,A,m) + h\sum_{j}^{n} d_{ij}\{I_j(t) - I_i(t)\}\Phi(t,I,i,j,\alpha,\beta) \quad (3)$$

$$\Phi(t,I,i,j,\alpha,\beta) = \frac{1}{1+\exp\{\beta(|I_j(t) - I_i(t)| - \alpha)\}} \quad (4)$$

i, j each represent agents, which are elements of the analyzed set n [15]. Also, n is a subset of the entire society N, and i, j ∈ n ⊂ N. The forgetting term $-\alpha I_i(t)\Delta t$ in equation (1) has been deleted, and the influence from sources other than dialogue has been set as $F(t,i,c,A,m)$ [16]. Regarding the function F, m represents the source such as TV or social media, A represents the opinion value of the source, and c represents the influence agent I receive from source m. If we were to model this after equation (1), function F could be expressed as $F(t,i,c,A,m) = \sum_m c_{im} A_m(t) - I_i(t))$ [17]. Also, considering the condition mentioned earlier that "when only influenced by and following external forces, values con- verge to a certain value, and when not following, they converge after maintaining a certain distance," F could be expressed as $F(t,i,c,A,m) = \sum_m c_{im}(A_m(t) - I_i(t))\Phi^*$ [18]. Note that Φ∗ is a sigmoid function that operates between external force m (such as social media) and agent i. The sigmoid func-

tion defined in equation (2) has been changed to equation (4) for improved readability. Also, to prevent confusion with β, b in equation (2) has been changed to α [19].

Redefinition

In the previous section, α, β were given as arbitrary constants, but I am skeptical of this approach [20]. It is obvious that even with the same $|I_j - I_i|$, Φ takes different values if α, β are different. While this might not be problematic if treated simply as a function, since Φ is a function with the meaning that "one is less influenced by another's opinion the more distant it is from one's own," α, β should not be constants without specific meanings [21]. In equations (1) and (3), the influence from another's opinion is given as $\sum_{j\neq i} d_{ij}\{I_j(t) - I_i(t)\}\Phi$, and $\{I_j(t) - I_i(t)\}\Phi$ represents "the strength and direction of influence given according to the difference in opinion with the other" [22]. The "direction of influence" is given according to the relationship with the other's opinion value, and the direction of agent i's approach to or repulsion from agent j for $d_{ij} > 0$ is shown in Table 2.

Table 2: Direction of agent i's approach to or repulsion from agent j.

Direction of $I_i(t)$	
Relation of opinion values	Approach Repulsion
$I_i(t) > I_i(t)$	Increasing direction Decreasing direction
$I_i(t) < I_i(t)$	Decreasing direction Increasing direction

Now, $d_{ij}\{I_j(t)-I_i(t)\}\Phi$ represents "trust in the other (agent j)" from agent i's perspective and "the strength and direction of influence given according to the difference in opinion with the other (agent j)," but when considering dialogue with only socially ex- pressed opinions, it is more natural to think of d_{ij} not as "trust that agent i has for agent j itself" but as "the degree of consensus between agent j's opinion and social norms as seen by agent i" [23]. Therefore, I define "the degree of consensus between agent j's opinion and social norms as seen by agent i" as c_{ij}, and the following equation reflects this in equation (3) [24]:

$$I_i(t+h)=I_i(t)+hF(t,i,c,A,m)+h\sum_j c_{ij}\{I_j(t)-I_i(t)\}\Phi(t,I,i,j,\alpha,\beta)$$

(5)

This is simply replacing d in equation (3) with c, but its meaning differs from equation (3) [25]. Specific meanings should also be given to the parameters α, β in the sigmoid function?). Let's first consider parameter α. Since $|I_j(t)-I_i(t)|-\alpha$ in equation (4) is "(difference in opinion values) - α," α needs to be a parameter related to opinion values. When $|I_j(t)-I_i(t)|-\alpha$, the sigmoid function takes 0.5, which is an inflection point, so it would be reasonable to consider α as a kind of threshold[7)] (Figure 4).

Figure 4: Honne-Tatemae OD model: quadrant of findings and implications.

Proposal of the Trust-Distrust Model

To reflect more realistic social situations, Ishii et al. extended the conventional bounded confidence model and proposed the "Trust-Distrust Model" [1,2]. The characteristics of this model are as follows:

- Expansion of opinion value range: Extended from the conventional $0 \le I_i(t) \le 1$ to $-\infty \le I_i(t) \le +\infty$, allowing positive values to represent affirmative opinions and negative values to represent negative opinions.

- Introduction of trust coefficient signs: Redefined coefficient D_{ij} as a trust coefficient, with $D_{ij}>0$ representing a trust relationship and $D_{ij}<0$ representing a distrust relationship.

- Introduction of sigmoid function: Introduced sigmoid function $\phi(I_i,I_j)$ to limit the influence when the difference in opinions is very large or small.

- Consideration of mass media influence: Introduced the term $c_iA(t)$ to represent influence from external information (such as mass media).

- Consideration of interest decay: Introduced the term $-\alpha I_i(t)$ to represent the decline in interest over time.

The basic equation of the Trust-Distrust model incorporating these features is expressed as follows:

$$\Delta I_i(t)=-\alpha I_i(t)+c_iA(t)\Delta t+\sum_{j=1}^N D_{ij}\phi(I_i(t),I_j(t))(I_j(t)-I_i(t))\Delta t \quad (6)$$

Here, the sigmoid function $\phi(I_i, I_j)$ is defined as follows:

$$\phi(I_i, I_j) = \frac{1}{1 + \exp(\beta(|I_i - I_j| - b))} \quad (7)$$

This sigmoid function changes smoothly around $|I_i - I_j| = b$ and rapidly approaches 0 when the difference in opinions between two people exceeds the threshold b, enabling the expression of the realistic situation where "one is hardly influenced by people whose opinions are too different" [14]. Also, the term $(I_j(t) - I_i(t))$ expresses the property that one is not influenced by people with the same opinion as oneself (Figure 5).
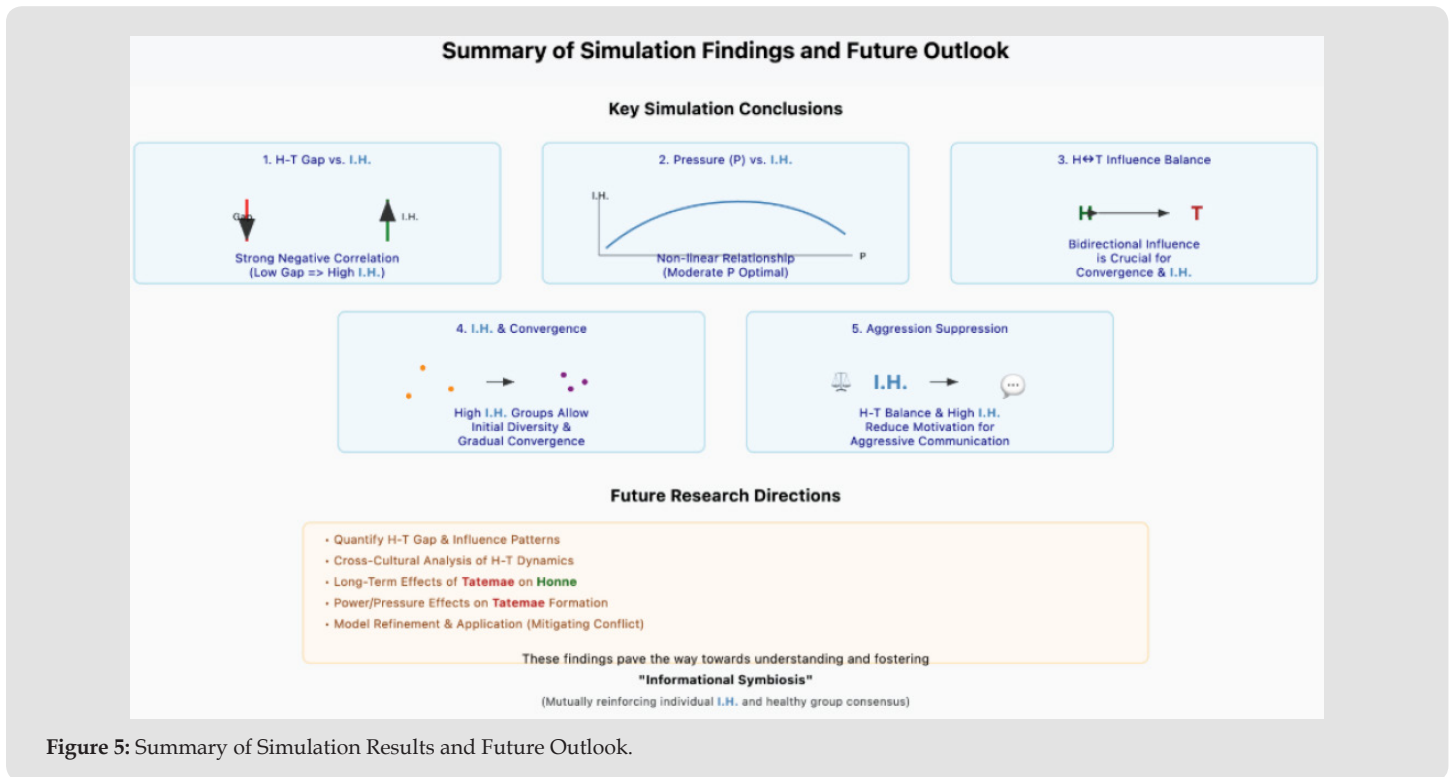

Figure 5: Summary of Simulation Results and Future Outlook.

Questions and Improvements on Ishii's Opinion Dynamics Theory

While Ishii et al.'s model has excellent features, there are also several questions [14]. This paper focuses particularly on the following points:

- Validity of the forgetting term: The term $-\alpha I_i(t)\Delta t$ in equation (6) means that opinions naturally converge to neutral (0) in the absence of external influence, which may not align with realistic human opinion formation processes.

- Validity of the external influence term: The term $c_i A(t)\Delta t$ alone in equation (6) allows opinions to potentially diverge to infinity if external influence continues, which is unrealistic.

- Ambiguity in parameter settings: The meanings and setting methods for parameters such as α, β, and b are not clearly defined.

To address these questions, this paper has improved the model as follows:

$$I_i(t + h) = I_i(t) + hF(t, i, c, A, m) + h\sum_j^n d_{ij}\{I_j(t) - I_i(t)\}\Phi(t, I, i, j, \alpha, \beta) \quad (8)$$

$$\Phi(t, I, i, j, \alpha, \beta) = \frac{1}{1 + \exp\{\beta(|I_j(t) - I_i(t)| - \alpha)\}} \quad (9)$$

The main improvements are as follows:

- Deletion of the forgetting term: Deleted $-\alpha I_i(t)\Delta t$ from equation (6) to create a more realistic model.

- Redefinition of external influence: Generalized external influence as function $F(t, i, c, A, m)$ to solve the divergence problem. In particular, forms such as $F(t, i, c, A, m) = \sum_m c_{im}(A_m(t) - I_i(t))\Phi^*$ can be considered, resulting in a model where opinions converge to a certain value.

- Clarification of parameters: Clearly defined parameter α in the sigmoid function as a threshold for opinion difference, and defined β as a function that varies according to trust in the other party.

- Redefinition of trust degree: Changed the interpretation of d_{ij} from "trust in the agent itself" to "the degree of consensus between the agent's opinion and social norms," modeling more realistic human relationship dynamics.

One's "True Opinion," Others' "Socially Ex- pressed Opinions," Superficial Trust Relation- ships, and Social Pressure: Discussion in the Model

$$I_i^{real}(t + \Delta t) = I_i^{real}(t) + \Delta t . F_{real}(I_i^{real}, I_j^{real}, d_{ij}^{real}) \quad (10)$$

$$I_i^{formal}(t + \Delta t) = I_i^{formal}(t) + \Delta t . F_{formal}(I_i^{real}, I_i^{formal}, I_j^{formal}, d_{ij}^{formal}, p)$$

$$(11)$$

Here, F_{real} is the opinion update function at the "true opinion" level, which is updated mainly based on others' "true opinions" and true trust relationships. On the other hand, F_{formal} is the opinion update function at the "socially expressed opinion" level, which is updated based on one's own "true opinion," others' "socially expressed opinions," superficial trust relationships, and social pressure P. Equations (10) and (11) are the core equations of the model proposed in this paper, which explicitly considers the dual structure of "true" and "socially expressed" opinions (Figure 6).
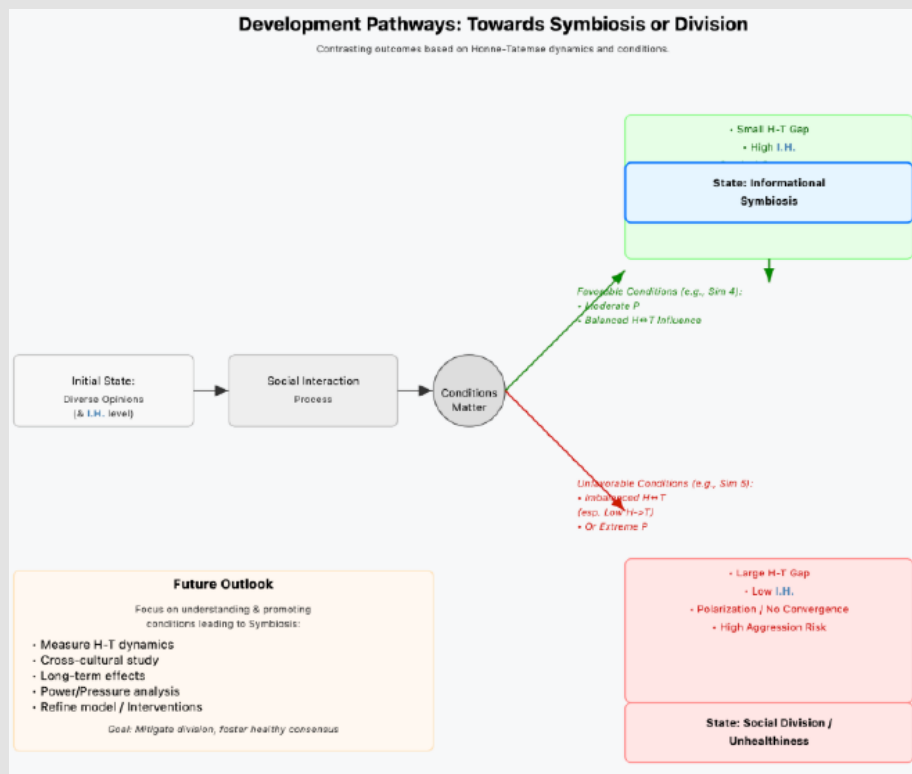
Figure 6: Development path: toward symbiosis or fragmentation.

Informational Health and Consensus Formation Model Considering the Dual Structure of True and Socially Expressed Opinions: Here, we analyze simulation results of a consensus formation model that explicitly incorporates the dual structure of "true and socially expressed opinions" that is particularly evident in Japanese society, extending Ishii et al.'s Trust-Distrust model. From the results of simulations using five different parameter sets, we consider the relationship between informational health and opinion formation processes, as well as the influence of social factors.

Creation of Informational Health through the Interaction of True and Socially Expressed Opinions: The dual structure of true and socially expressed opinions has functioned as cultural wisdom for balancing social harmony and individual autonomy, rather than mere falsehood or deception. The simulation results mathematically express how the dynamics of this dual structure affect informational health.

From the simulation results, particularly important findings include:

• In Simulation 4, the divergence between true and socially expressed opinions is minimal (0.007), while informational health maintains a high level (0.732). This simulation is characterized by a relatively high influence from socially expressed to true opinions (0.178) and a low true opinion update rate (0.107). This result shows that not only do socially expressed opinions approach true ones, but true opinions are also pulled toward socially expressed ones, resulting in increased consistency between the two (Figure 7).

• In Simulation 5, the divergence between true and socially expressed opinions is maximal (0.145), while the average informational health is also lowest (0.311). This result mathematically supports that divergence between true and socially expressed opinions negatively affects informational health.

• In Simulation 2, the average informational health is the highest (0.813), with moderate social pressure (0.475). This suggests that neither excessive social pressure nor excessive individualism promotes informational health (Figure 8).
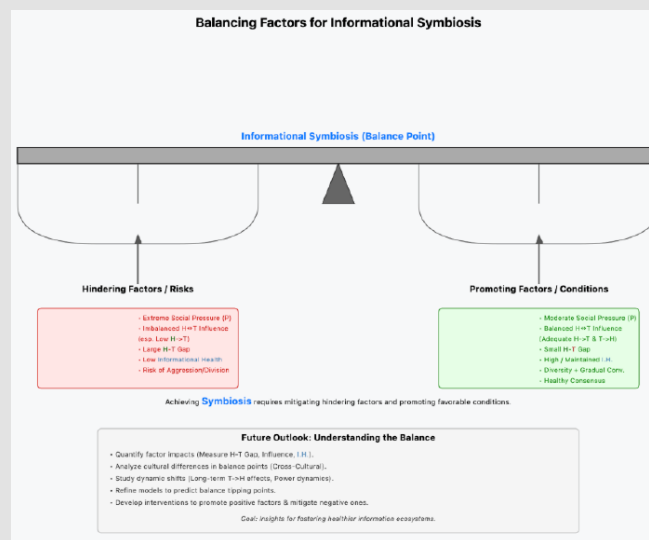

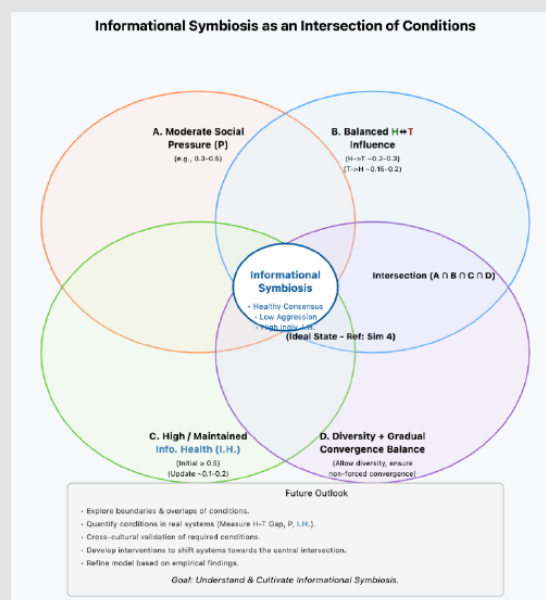**Figure 7:** Balancing Factors for Informational Symbiosis.

Figure 8: Informational Symbiosis as an Intersection of Conditions.

Relationship Between Social Pressure and Informational Health

Way to Informational Symbiosis: The simulation results provide important insights into the relationship between social pressure (P) and informational health. As shown in Table 3, in Simulation 1 with the highest social pressure (0.735), informational health is moderate (0.753), whereas in Simulation 2 with moderate social pressure (0.475), informational health is highest (0.813). In contrast, in Simulation 3 with relatively low social pressure (0.192), informational health maintains a good level (0.739).

Table 3: Summary of Parameters and Results for Each Simulation.

Parameter	Sim1	Sim2	Sim3	Sim4	Sim5
Social Pressure	0.735	0.475	0.192	0.316	0.429
True Opinion Update Rate	0.24	0.166	0.293	0.107	0.248
Socially Expressed Opinion Update Rate	0.48	0.468	0.328	0.253	0.437
True → Socially Expressed Influence	0.277	0.133	0.134	0.234	0.107
Socially Ex- pressed → True Influence	0.073	0.086	0.14	0.178	0.163
Informational Health Update Rate	0.191	0.106	0.199	0.12	0.182
Average Divergence between True and Socially Expressed	0.051	0.034	0.044	0.007	0.145
True Opinion Convergence (Variance)	0	0	0	0	0
Socially Expressed Opinion Convergence (Variance)	0	0	0	0	0
Final Average Informational Health	0.753	0.813	0.739	0.732	0.311

These results suggest the following non-linear relationship between social pressure and informational health:

- Excessively high social pressure may decrease informational health by promoting excessive conformity to socially expressed opinions, widening the gap from true opinions.

- Moderate social pressure may promote balanced opinion formation by encouraging a certain consideration for socially expressed opinions while also allowing expression of true opinions, thereby enhancing informational health.

- Even when social pressure is relatively low, a certain convergence of opinions may occur through interactions between agents, potentially maintaining informational health.

- Particularly noteworthy are the results of Simulation

- In this simulation, despite moderate social pressure (0.429), the influence from true to socially expressed opinions is lowest (0.107), creating a large divergence between true and socially expressed opinions and significantly reducing informational health. This indicates that not only the absolute magnitude of social pressure but also the pattern of interaction between true and socially expressed opinions has a significant impact on informational health (Table 4).

Table 4: Conditions for Achieving Informational Symbiosis.

Conditions for Informational Symbiosis	Specific Parameters Derived from Simulations
Moderate social pressure	Around 0.3-0.5 (from comparison of Simulations 2, 4, and 5)
Balance of mutual influence between true and socially expressed opinions	True→ Socially expressed in- fluence: Around 0.2-0.3, Socially expressed→ True influence: Around 0.15-0.2 (optimal pattern in Simulation 4)
Maintenance and improvement of informational health	Initial value: 0.5 or higher, Update rate: Around 0.1-0.2 (common points in Simulations 1, 2, 3, and 4)
Balance be- tween diversity and convergence of opinions	Allowing initial diversity and gentle convergence process (from all simulations)

Path to Informational Symbiosis: When integrating the simulation results, the possibility of a new social practice that could be called "Informational Symbiosis" emerges, based on informational health. This refers to a state where individual informational health and healthy group consensus formation mutually support and strengthen each other (Figure 9).
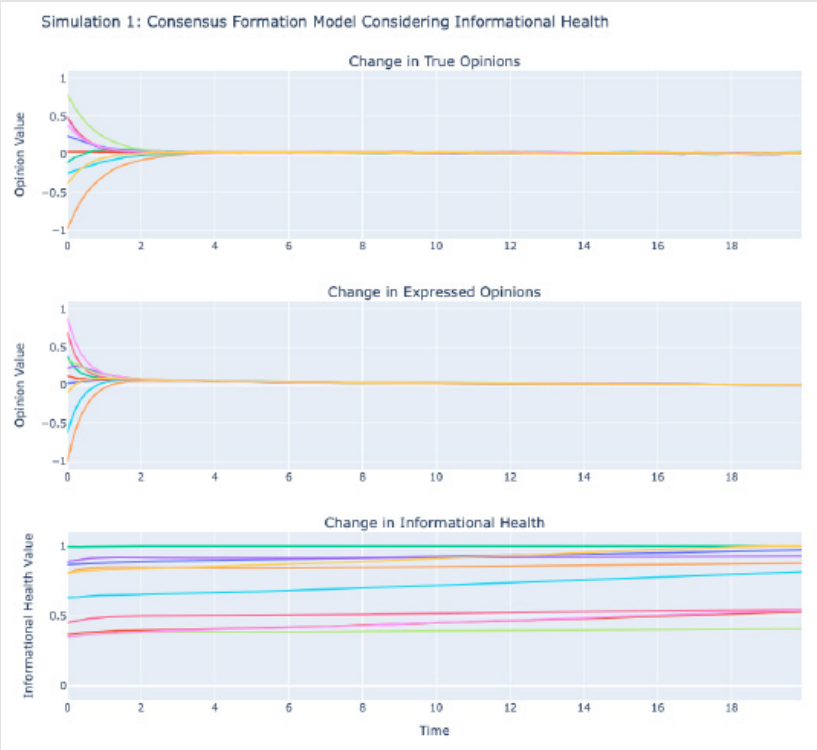
Figure 9: Simulation 1: Consensus Formation Model Considering Informational Health.

Characteristics of informational symbiosis include:

- **Appropriate Balance Between True and Socially Expressed Opinions**: As in Simulations 2 and 4, a state where there is appropriate interaction between true and socially expressed opinions without extreme divergence (Table 5).

- **Dynamic Balance Between Diversity and Convergence**: A process where initial diversity of opinions is respected, yet a certain convergence occurs over time.

- **Mutual Creation of Informational Health**: A virtuous cycle where individual informational health improves through interaction with others, which in turn improves the information environment for the entire group.

- **Natural Suppression of Aggressive Communication**: A state where the motivation for aggressive expression is reduced, and healthy dialogue is promoted by maintaining an appropriate balance between true and socially expressed opinions.

Table 5: Key Indicators and Parameters for Simulation 1.

Indicator	Value
Average True-Expressed Divergence	0.0103
True Opinion Convergence	0
(Variance)	
Expressed Opinion Convergence	0
(Variance)	
Final Average Informational	0.7619
Health	
Parameter	**Value**
Social Pressure	0.5436
True Opinion Update Rate	0.2724
Expressed Opinion Update Rate	0.4109
True→ Expressed Influence	0.1526
Expressed→ True Influence	0.0613
Informational Health Update	0.1232
Rate	

Simulation 4 is considered closest to the ideal state of informational symbiosis. In this simulation, the di- vergence between true and socially expressed opinions is minimal, and informational health maintains a high level. Additionally, the process of gradual convergence from initial diversity of opinions is observed. The simulation results suggest the following elements as important conditions for achieving such a state: To apply these conditions to real society, not only improvement of individual media literacy and critical thinking skills but also reconsideration of social institutions, design of dialogue venues, and cultural norms are necessary. Particularly important is the perspective that an appropriate balance of the dual structure of true and socially expressed opinions leads to informational health and healthy consensus formation, rather than viewing it merely as a problem (Table 6).

Table 6: Key Indicators and Parameters for Simulation 2.

Indicator	Value
Average True-Expressed Divergence	0.0449
True Opinion Convergence	0
(Variance)	
Expressed Opinion Convergence	0
(Variance)	
Final Average Informational	0.7916
Health	
Parameter	**Value**
Social Pressure	0.2892
True Opinion Update Rate	0.1465
Expressed Opinion Update Rate	0.3373
True→ Expressed Influence	0.2197
Expressed→ True Influence	0.0665
Informational Health Update	0.1864
Rate	

Prospects for Opinion Dynamics Considering "True and Socially Expressed Opinions"

By extending the Trust-Distrust model from the perspective of "true and socially expressed opinions," we can gain a deeper understanding of the complexity of opinion formation processes in society. In particular, the asymmetry and external influence effects observed from the simulation results are useful for quantitatively analyzing the influence of social power structures and pressures on opinion formation as "socially expressed opinions." Various developments are expected, such as methods for measuring the divergence between "true" and "socially expressed" opinions, comparative analysis in societies with different cultural backgrounds, and examination of the long-term influence of "socially expressed opinions" on "true opinions" (Table 7). The advancement of such research may make the complex mechanisms of social consensus formation more understandable and provide insights for mitigating social division and conflict (Figure 10).

Table 7: Key Indicators and Parameters for Simulation 3.

| Indicator	Value
Average True-Expressed Divergence	0.0831
True Opinion Convergence	0
(Variance)	
Expressed Opinion Convergence	0
(Variance)	
Final Average Informational	0.6852
Health	
Parameter	**Value**
Social Pressure	0.5948
True Opinion Update Rate	0.2166
Expressed Opinion Update Rate	0.2464
True→ Expressed Influence	0.2903
Expressed→ True Influence	0.1771
Informational Health Update	0.1534
Rate	
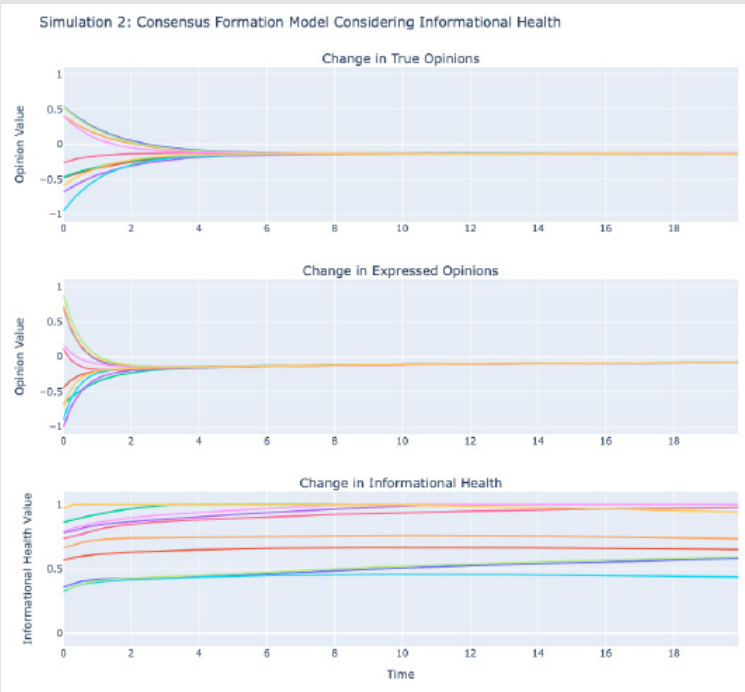
Figure 10: Simulation 2: Consensus Formation Model Considering Informational Health.

Simulation Results

Overview of the Informational Health Model

This research constructs a model that considers individual infor-mational health in the consensus formation process between multi-ple agents and conducts simulation analysis. In this model, each agent possesses both true opinions and expressed opinions, and the degree of informational health influences the divergence and changes be-tween these opinions (Table 8).

Table 8: Key Indicators and Parameters for Simulation 4.

Indicator	Value
Average True-Expressed Divergence	0.0303
True Opinion Convergence	0
(Variance)	
Expressed Opinion Convergence	0
(Variance)	
Final Average Informational	0.6538
Health	
Parameter	**Value**
Social Pressure	0.5602
True Opinion Update Rate	0.3283
Expressed Opinion Update Rate	0.3733
True→ Expressed Influence	0.2628
Expressed→ True Influence	0.1765
Informational Health Update	0.0619
Rate	

Simulation Settings

The following parameters are set in this model:

- I_{real} : True opinion (range from -1 to 1)

- I_{formal} : Expressed opinion (range from -1 to 1)

- d_{real} : Trust matrix at the true opinion level (range from 0 to 1)

- d_{formal} : Trust matrix at the expressed opinion level (range from 0 to 1)

- E: Informational health (range from 0 to 1)

- P: Social pressure (range from 0 to 1)

- S: Social status or role (range from 0 to 1)

- α_{real} : Update rate for true opinions

- α_{formal} : Update rate for expressed opinions

- β_{health} : Update rate for informational health

- γ_{real_formal} : Influence from true opinions to expressed opinions

- γ_{formal_real} : Influence from expressed opinions to true opinions

Five simulations were conducted, each consisting of 10 agents. The simulation period was set to 200-time steps, with a time increment dt of 0.1.

Analysis of Simulation Results

Figure 1 shows the results of the first simulation. The top graph displays the temporal change in true opinions, the middle graph shows the change in expressed opinions, and the bottom graph represents the change in informational health. For both true and expressed opinions, the initial values were randomly distributed between -1 and 1, but over time, all agents' opinions converged toward a neutral value (approximately 0). However, it is characteristic that informational health maintains different values for each agent and fluctuates over time (Table 9). In the second simulation, the social pressure is set lower at 0.2892 compared to the first simulation, but both true and expressed opinions again converge to neutral values. However, informational health improves overall, showing the highest final average value of 0.7916 among the five simulations. The true-expressed opinion divergence in this simulation is 0.0449, which is larger than in the first simulation (Figure 11). In the third simulation, social pressure is set at the highest value of 0.5948. This simulation shows the maximum true-expressed opinion divergence of 0.0831 among the five simulations, suggesting that high social pressure influences the divergence between true and expressed opinions. Additionally, informational health is relatively low, with some agents showing a declining trend over time.
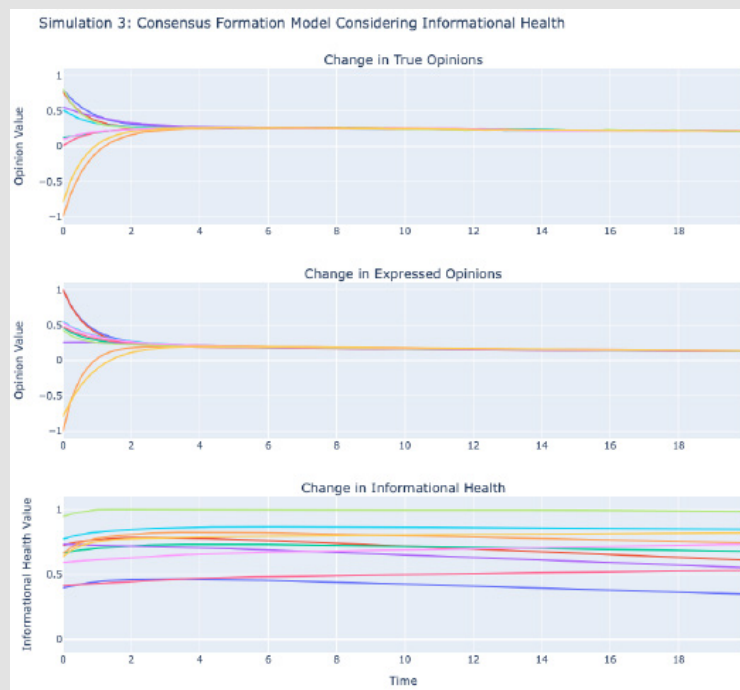

Figure 11: Simulation 3: Consensus Formation Model Considering Informational Health.
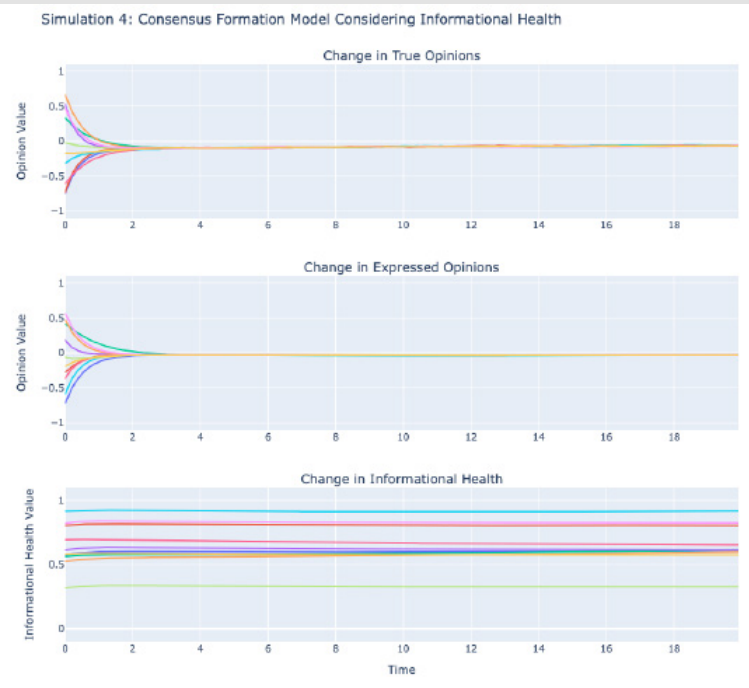
Table 9: Key Indicators and Parameters for Simulation 5.

Indicator	Value
Average True-Expressed Divergence	0.0205
True Opinion Convergence	0
(Variance)	
Expressed Opinion Convergence	0
(Variance)	
Final Average Informational	0.7328
Health	
Parameter	**Value**
Social Pressure	0.2729
True Opinion Update Rate	0.4183
Expressed Opinion Update Rate	0.104
True→ Expressed Influence	0.2834
Expressed→ True Influence	0.1859
Informational Health Update	0.0573
Rate	

In the fourth simulation, true opinion convergence is said to be the highest, though the true opinion con- vergence (variance) values are approximately 0 in all simulations. The final average informational health is relatively low at 0.6538, with variation in health values among agents (Table 10). In the fifth simulation, social pressure is set relatively low at 0.2729, while the true opinion update rate shows the highest value at 0.4183. As with other simulations, both true and expressed opinions converge to nearly neutral values. The final average informational health is 0.7328, maintaining a consistently high level (Figure 12). Figure 6 shows the correlation between simulation parameters and results. From this correlation heatmap, the following important relationships are observed:

- There is a positive correlation between Social Pressure and True-Expressed Divergence. This suggests that as social pressure increases, the gap between true opinions and expressed opinions widens.

- There is a negative correlation between Informational Health Average and True-Expressed Divergence. This means that agents with higher informational health have greater consistency between their true and expressed opinions.

- There is a negative correlation between Expressed True Influence and informational health. When expressed opinions strongly influence true opinions, informational health tends to decline (Figure 13).

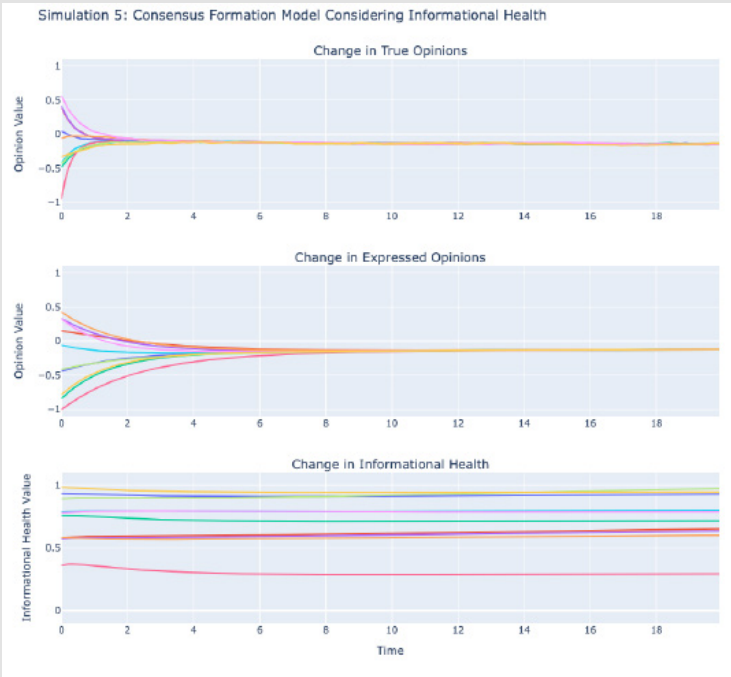**Figure 12:** Simulation 4: Consensus Formation Model Considering Informational Health.

Figure 13: Simulation 5: Consensus Formation Model Considering Informational Health.

Table 10: Mean and Standard Deviation of Simulation Parameters and Results.

Indicator	Mean Dev	± Std
True-Expressed Divergence	0.0378	±
	0.0283	
True Opinion Convergence	0	±
(Variance)	0	
Expressed Opinion Convergence	0	±
(Variance)	0	
Informational Health Average	0.7251	±
	0.0559	
Social Pressure	0.4521	±
	0.1574	
True Opinion Update Rate	0.2764	±
	0.104	
Expressed Opinion Update Rate	0.2944	±
	0.1227	
True→ Expressed Influence	0.2418	±
	0.0569	
Expressed→ True Influence	0.1335	±
	0.0636	

Comprehensive Analysis of Simulation Results

The following trends were confirmed from the results of the five simulations:

- **Opinion Convergence Tendency**: In all simulations, agents' true opinions and expressed opinions tend to converge toward a neutral value (approximately 0) over time. This is a result of the opinion update mechanism in the model design.

- **Impact of Social Pressure**: In cases with high social pressure (Simulation 3: 0.5948, Simulation 4: 0.5602), there is a tendency for greater divergence between true and expressed opinions. In particular, Simulation 3 shows the maximum true- expressed opinion divergence of 0.0831.

- **Diversity of Informational Health**: Informational health shows diverse values both between simulations and between agents. The highest informational health is observed in Simulation 2 (0.7916), while the lowest is in Simulation 4 (0.6538) (Figure 14).

- **Mutual Influence Between True and Expressed Opinions**: The influence from true opinions to expressed opinions (γ_{real_formal}) consistently shows higher values than the influence from expressed opinions to true opinions (γ_{formal_real}). This suggests that internal true thoughts have a stronger influence on expressed opinions (Figure 15).

Figure 14: Correlation between Simulation Parameters and Results.
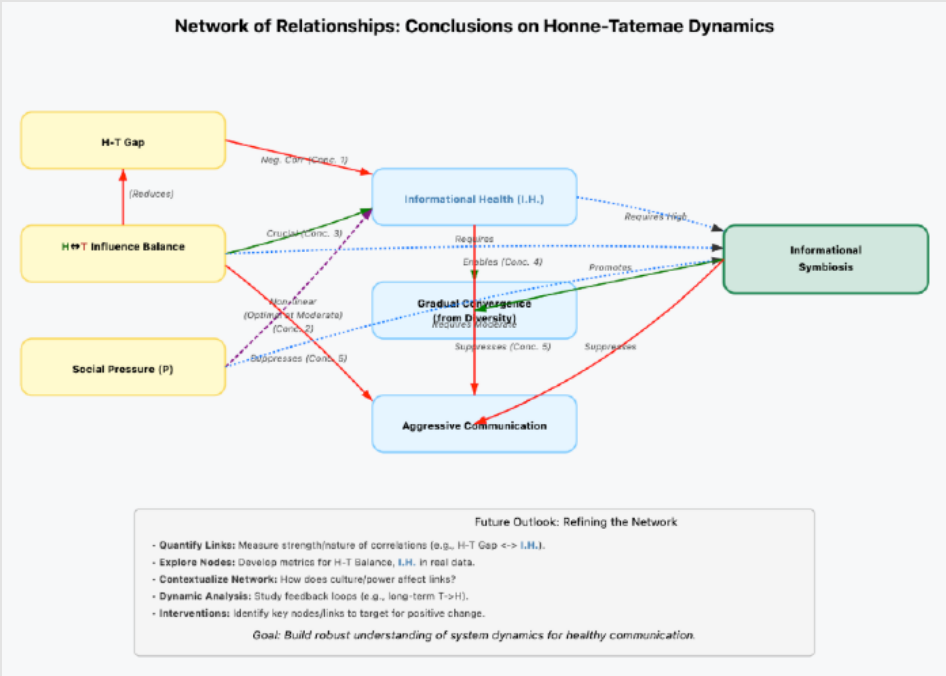

Figure 15: Relationship Networks: Conclusions on the Honne-Tate Dynamics.

Conclusion

In this paper, we extended Ishii et al.'s Trust-Distrust model to perform simulations of a consensus formation model incorporating the dual structure of true and socially expressed opinions and informational health. From the results of simulations using five different parameter sets, the following main conclusions are drawn:

- The smaller the divergence between true and socially expressed opinions, the higher the tendency for informational health. In particular, in Simulation 4, the divergence between true and socially expressed opinions is minimal (0.007) and informational health maintains a high level (0.732).

- The relationship between social pressure and informational health is non-linear, with a tendency for informational health to be highest under moderate social pressure (Simulation 2: 0.475).

- The balance of interaction between true and socially expressed opinions plays an important role in the convergence process of opinions and the maintenance of informational health. In particular, the balance between the influence from true to socially expressed opinions and from socially expressed to true opinions is important.

- In groups with high informational health, even starting from diverse initial opinions, there is a tendency for gentle convergence to occur over time.

- For the suppression of aggressive communication, maintaining an appropriate balance between true and socially expressed opinions and informational health is important.

These findings can serve as a starting point for a new social practice that could be called "Informational Symbiosis." Informational symbiosis refers to a state where individual informational health and healthy group consensus formation mutually support and strengthen each other. To achieve this state, appropriate social pressure, proper balance of mutual influence between true and socially expressed opinions, environmental preparation for the maintenance and improvement of informational health, and balance between diversity and convergence of opinions will be discussed.

Acknowledgments

References

1. Ishii A, Kawahata Y (2018) Opinion Dynamics Theory for Analysis of Consensus Formation and Division of Opinion on the Internet. In Proceedings of the 22nd Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES2018) 71-76.

2. Ishii A (2019) Opinion Dynamics Theory Considering Trust and Suspicion in Human Relations. In Group Decision and Negotiation: Behavior, Models, and Support. Lecture Notes in Business Information Processing In: Morais D, Carreras A, deAlmeida A, Vetschera R (Eds.)., 351: 193-204.

3. Ishii A, Kawahata Y (2019) New Opinion dynamics theory considering interpersonal relationship of both trust and distrust. In Proceedings of ABCSS2019 in Web Intelligence, p. 43-50.

4. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. Rev Mod Phys 81: 591-646.

5. Sˆırbu A, Loreto V, Servedio V D P, Tria F (2017) Opinion Dynamics: Models, Extensions and External Effects. In Participatory Sensing, Opinions and Collective Awareness In: Loreto V, et al. (Edt.)., p. 42.

6. Galam S, Gefen (Feigenblat) Y, Shapir Y (1982) Sociophysics: A new approach of sociological collective behaviour: I. Mean-behaviour description of a strike. Journal of Mathematical Sociology 9(1): 1-13.

7. Vaz Martins T, Pineda M, Toral R (2010) Mass media and repulsive interactions in continuous-opinion dynamics. Europhys Lett 91(4): 48003.

8. Baum M B, Potter P B K (2008) The relationships between mass media, public opinion, and foreign policy: toward a theoretical synthesis. Annu Rev Polit Sci 11: 39-65.

9. Urena R, Kou G, Dong Y, Chiclana F, Herrera Viedma E (2019) A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. Information Sciences 478: 461-475.

10. Ishii A, Kawahata Y (2019) Opinion dynamics theory considering interpersonal relationship of trust and distrust and media effects. In The 33rd Annual Conference of the Japanese Society for Artificial Intelligence JSAI.

11. Ishii A, Okano N (2021) Sociophysics Approach of Simulation of Mass Media Effects in Society Using New Opinion Dynamics. In Intelligent Systems and Applications (Proceedings of the 2020 Intelligent Systems Conference (IntelliSys)) 3: 13-28.

12. Galam S, Javarone M A (2016) Modeling Radicalization Phenomena in Heterogeneous Populations. PLOS ONE.

13. McCombs M (2018) Setting the Agenda: Mass Media and Public Opinion. John Wiley Sons.

14. Nozimi O (2022) Opinion dynamics analysis of phenomena that lead to social fragmentation and isolation. Doctoral Thesis.

15. Ishii A, Okano N (2021) Social Simulation of a Divided Society Using Opinion Dynamics. Proceedings of the 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (in press).

16. Okano N, Ishii A (2021) Opinion dynamics on a dual network of neighbor relations and society as a whole using the Trust-Distrust model. In Springer Nature - Book Series: Transactions on Computational Science & Computational Intelligence (The 23rd International Conference on Artificial Intelligence (ICAI'21)).

17. Chen, Zhongtian, Hanlin Lan (2021) Dynamics of public opinion: Diverse media and audiences' choices. Journal of Artificial Societies and Social Simulation 24(2).

18. Perra N, Rocha L E (2019) Modelling opinion dynamics in the age of algorithmic personalization. Sci Rep 9: 7261.

19. Douven I (2010) Simulating Peer Disagreements. Studies in History and Philosophy of Science Part A 41(2): 148-157.

20. Abelson R P (1964) Mathematical models of the distribution of attitudes under controversy. In: Frederiksen N, Gulliksen H (Eds.)., Contributions to Mathematical Psychology.

21. Salzarulo L A (2006) continuous opinion dynamics model based on the principle of meta-contrast. Journal of Artificial Societies and Social Simulation 9(1): 13.

22. Flache A (2018) About renegades and outgroup haters: Modeling the link between social influence and intergroup attitudes. Advances in Complex Systems 21(06n07): 1850017.

23. Asch S E (1951) Effects of group pressure upon the modification and distortion of judgments. In: Groups, Leadership and Men Research in Human Relations, pp. 177-190.

24. Wang S W, Huang C Y, Sun C T (2014) Modeling self-perception agents in an opinion dynamics propagation society. Simulation 90(3): 238-248.

25. Huang C Y, Wen T H (2014) A novel private attitude and public opinion dynamics model for simulating pluralistic ignorance and minority influence. J Artif Soc Soc Simul 17(3).

Submission Link: https://biomedres.us/submit-manuscript.php

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

https://biomedres.us/