

Proceeding Paper

Behavioral Influence of Social Self Perception in a Sociophysical Simulation [†]

Fabian Sigler ^{1,2,*}, Viktoria Kainz ^{1,2}, Torsten Enßlin ^{1,2,3}, Céline Boehm ⁴ and Sonja Utz ^{5,6}¹ Max Planck Institute for Astrophysics, Karl-Schwarzschildstraße 1, 85748 Garching, Germany² Faculty of Physics, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 Munich, Germany³ Excellence Cluster ORIGINS, Boltzmannstr. 2, 85748 Garching, Germany⁴ School of Physics, The University of Sydney, Physics Road, Camperdown, NSW 2006, Australia⁵ Department of Psychology, University of Tübingen, 72076 Tübingen, Germany⁶ Leibniz-Institut für Wissensmedien Tübingen, 72076 Tübingen, Germany

* Correspondence: f.sigler@campus.lmu.de

[†] Presented at the 42nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 3–7 July 2023.

Abstract: Humans make decisions about their actions based on a combination of their objectives and their knowledge about the state of the world surrounding them. In social interactions, one prevalent goal is the ambition to be perceived to be an honest, trustworthy person in terms of having a reputation of frequently making true statements. Aiming for this goal requires the decision whether to communicate information truthfully or if deceptive lies might improve the reputation even more. The basis of this decision involves not only an individual's belief about others, but also their understanding of others' beliefs, described by the concept of Theory of Mind, and the mental processes from which these beliefs emerge. In the present work, we used the *Reputation Game Simulation* as an approach for modeling the evolution of reputation in agent-based social communication networks, in which agents treat information approximately according to Bayesian logic. We implemented a second-order Theory of Mind based message decision strategy that allows the agents to mentally simulate the impact of different communication options on the knowledge of their counterparts' minds in order to identify the message that is expected to maximize their reputation. Analysis of the communication patterns obtained showed that deception was chosen more frequently than the truthful communication option. However, the efficacy of such deceptive behavior turned out to have a strong correlation with the accuracy of the agents' Theory of Mind representation.



Citation: Sigler, F.; Kainz, V.; Enßlin, T.; Boehm, C.; Utz, S. Behavioral Influence of Social Self Perception in a Sociophysical Simulation. *Phys. Sci. Forum* **2023**, *9*, 3. <https://doi.org/10.3390/psf2023009003>

Academic Editors: Udo von Toussaint and Roland Preuss

Published: 24 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: sociophysics; information theory; Theory of Mind; agent-based modeling; computational psychology

1. Introduction

Purposeful lying, particularly in a deceptive manner, involves a multifaceted process including understanding the nature of lies itself, producing the lie, and maintaining built-up lies without being spotted as a dishonest person [1]. Therefore, a well-put lie demands significant cognitive abilities, including the capacity to develop a robust *Theory of Mind* (ToM). ToM refers to the ability to infer and deduce the mental states of others by trying to understand what they know, think, belief, plan, etc. only considering informative aspects apart from the others' feelings [2]. The *depth* or *order* of the ToM denotes the number of recursive ToM beliefs that are included in a ToM representation, e.g., a belief of the form "A thinks that B thinks, that C thinks X" corresponds to a second-order ToM belief. Recent sociological studies find that the performance of young children in Theory of Mind-related tasks is not only closely connected to their mere ability to put targeted lies, but also to their tendency to do so in the first place [3]. Although lying may be considered antisocial behavior, this does not necessarily hold when telling white lies or

flattering to create a more pleasant social atmosphere [4,5]. Nonetheless, we frequently resort to deceitful lies to gain an advantage for ourselves [6,7], for example in the form of social advancement by attempting to move the own **reputation** in a desired direction. An approach to simulate the dynamics of reputation within social networks was taken by Enßlin, Kainz, and Böhm (2022) [8,9]. In the “Reputation Game Simulation” (RGS), virtual agents communicate their beliefs to others with the objective of maximizing the own reputation among their fellows. The agents are modeled not to retain every step of communication, but to let them store their information at hand in a compressed way with a fixed number of parameters that are updated according to Bayesian probabilistic reasoning and information theoretical principles.

To send and receive messages, agents use one of many strategies that are implemented in the RGS based on conceivable real-life personality types and communication schemes. In this work, we explore a novel type of reputation-maximization-oriented strategy based on the implementation of a higher-order Theory of Mind.

2. The Reputation Game Simulation

2.1. Basic Principles

In the following, we summarize the most important aspects of the Reputation Game’s mechanisms. For the details, we refer to [8].

The game follows the interactions of n agents that exchange information of only one kind, namely, the honesty of other agents. In each round each agent once takes on the role of the speaker (denoted by agent a) and therefore chooses a communication partner b and a topic agent c to transfer a message about to b and then receives a reply from b about c . This means that a and b tell each other how honest they believe c ’s statements to be in general, which can either be the agents’ actual belief in form of an honest statement or a lie. The model is built around this type of information called the **reputation**. The agents use this image of others to decide about the trustworthiness of received messages to properly implement those into their beliefs while striving for maximizing their own reputation in the eyes of their fellows. A higher reputation of an agent also means a stronger influence of his messages on the group. So far, various strategies for achieving this goal are implemented in the RGS, varying in the way lies are used to deceptively influence other agents’ beliefs.

The agents’ beliefs are represented as the set of two parameters $I_{ab} = (\mu_{ab}, \lambda_{ab})$ that denote the number of honest and dishonest statements, respectively, that a believes b to have told during the game. Thus, following Bayesian logic, they describe a probability distribution over the possible lie frequencies of b in the eyes of a by $P(x_b|I_{ab}) = \text{Beta}(\mu_{ab} + 1, \lambda_{ab} + 1|I_{ab})$. The mean of a belief $\bar{x}_{ab} \in [0, 1]$, calculated from this probability distribution, refers to the reputation of b for agent a and is a measure of the frequency for a ’s belief for b to tell the truth.

Besides their belief on the behavior of others, the agent’s mental states also include the first-order Theory of Mind values of the form I_{abc} , denoting what a thinks b thinks about c . Proper information on the beliefs and plans of others, on the one hand, protects the agents against being manipulated, and, on the other hand, makes it easier the other way round to successfully deceive their counterparts.

After receiving a message, the new belief of an agent ideally would have the form of a superposition of probability distributions for the old belief (that is to be kept in case of being lied to) and one for the new belief, in case the message was honest. The weighting of these two beliefs is according to the assessment of the receiver that the message was a lie or truth, respectively. For the sake of keeping the form of two parameters μ and λ , the superposition is then compressed back into the form of Beta distribution by minimizing the appropriate information distance between those, namely, the Kullback–Leibler divergence.

Whether an agent a lies or tells the truth in a specific situation is not decided by himself but in a random process characterized by the intrinsic, agent-specific lie frequency $1 - x_a$.

If an agent is chosen to lie in the first place, his lies are constructed around first-order ToM values (what a thinks that b thinks about c , i.e., lying positive means transmitting

a message that is of the form $(\mu_{abc} + \alpha, \lambda_{abc})$, and for a negative $(\mu_{abc}, \lambda_{abc} + \alpha)$, with an α for that a expects to make the statement credible for b). This means the agents present the topic slightly better or worse than they believe the receiver to currently think about it. Transmitting the exact values of the first-order ToM is what is referred to as a **white lie**. After sending a message, the agents also update their belief on their own honesty according to whether they have just lied or not. This value is also influenced by the information the agents receive by others about themselves.

The agents use several indicators to detect lies in order to protect themselves from deception and adopting false information in their belief state. These include blushing (when agents tell a lie, they run the risk of exposing themselves by a subsequent blushing with frequency $f_b = 0.1$) or unusual surprise. The latter denotes the most crucial part of lie detection and means of evaluating the trustworthiness of a message according to the surprise it generates in the agent's mind when hearing it, i.e., the amount of information gained if the agent updated its initial belief to the information contained in the message. This surprise is calculated by the Kullback–Leibler divergence between the agents' current belief on the topic and the message. The value also is weighted by a factor κ , an agent specific value that denotes the mean of message surprises of the last ten conversations and therefore represents the current "social atmosphere".

Another indicator for honesty is the so-called *confession*. It means that, if a statement of an agent about himself is more negative than the receiver's current belief, the message is seen to be ultimately honest.

Besides the pure reputation-driven dynamics, the RGS contains a second principle that decides whether an agent should be lied positively or negatively about: whether they are regarded as a friend or an enemy. If an agent receives a message about himself wherein the mean reputation is above the perceived average, the sender is added to the receiver's list of friends, or to the list of enemies if the opposite was the case. If an agent decides to communicate false information, they will lie positively about friends, causing the receiver to put more trust in the expected positive subsequent messages of the friend about a . Following the same reasoning, a aims at minimizing the reputation of his enemies.

2.2. Implementation of a Second-Order Theory of Mind

We will now replace the lie direction rule for a new type of agent (called ToM-agent in the following) with the following considerations. The standard mental model of agents in the RGS includes a ToM up to the first order represented by $I_{abc} = (\mu_{abc}, \lambda_{abc})$ describing a 's belief about b 's zero order belief I_{bc} on c . The ToM-agent's knowledge not solely includes those first-order ToM values but also the very fact that this is the highest order that other agents conceive and an understanding of the mental processes leading to those beliefs. At least we demand the ToM-agents to be able to "imagine" how a hypothetical message affects a hypothetical mind. The rational consequence is to extend the ToM by an extra order, as this allows the ToM-agent to build up a complete picture of his fellows' mental states and, therefore, examine how they are influenced by possible messages and choose the option promising the best result in terms of its expected reputation. The ToM strategy therefore consists of using the extended picture of the surrounding to simulate the impact of different behavioral options on it. Following rational theory, the best response is again to undermine the strategy by increasing one's ToM by one more level and simulate the ToM-agents simulations, respectively, for the next response, resulting in an infinite number of iterations [10]. This is neither reasonable to model nor a process to be expected in real life. Sociology finds that, in practice, people do not apply deductive rational reasoning to the highest level but rather act according to a so-called **bounded rationality** that basically emerges from limitations in human brain capacities and from the resulting belief that others also react according to a bounded rationality [11]. Moreover, in the RGS, the uncertainty about the information of the ToM becomes more uncertain with each order of ToM, making the mental simulations more inaccurate and mostly useless at some point of depth. The proper strategic solution is, therefore, to identify the level of others' reasoning

and undermine their strategy by thinking exactly one order deeper than their counterparts, which is exactly what ToM-agents in the RGS do, where they know the others' depth of reasoning to be one.

We also see this approach legitimized by its practical relevance, as sociological studies show that humans in general are able to form and apply second-order ToM beliefs, while higher orders are observed less frequently [1–3,10].

2.3. Implementing a Lie Direction Decision Rule

ToM-agents use the following strategy to decide whether to be honest or lie, and if so, in what direction. The ToM-agents a use their second-order ToM to directly examine the effects of their talk on their counterpart's mental state and how this might influence further conversations and belief changes about himself. Therefore, a ToM-agent imagines a set of virtual agents (*dummies*) holding the mental state $\mathcal{I}_{b'}$ according to a 's first- and second-order ToM of his counterpart, agent b . Thus, $\mathcal{I}_{b'c} = I_{abc}$ and $\mathcal{I}_{b'ac} = I_{abac}$, $\mathcal{I}_{b'ca} = I_{abca}$. In the current model, the second-order parameters of the latter shape are hard to extract from conversations as it is only relevant and updated in conversations in which a is not involved. The ToM-agent is therefore not able to have an independent value for the second-order coefficients and his best guess is that c 's beliefs were grasped in the equal way by both agents. We therefore assume the ToM-agent to replace $\mathcal{I}_{b'ca} = I_{abca} = I_{aca}$, meaning a assumes b to hold the same first-order belief on c 's opinion on a as a himself.

2.4. Finding the Optimal Message

If an agent has chosen communication partner b and topic c , he finds himself with different options for the messages that he is about to be sent, which are a white, a positive, and a negative lie or a 's honest opinion. In the first step, the ToM-agent constructs all these possible messages on the basis of his first-order ToM. Then five virtual dummy minds for b are created, of which four are updated under one of the possible messages each, whereas the last is confronted with an arbitrary lie with a subsequent blushing. We denote these options in the following by w , $+$, $-$, t , and blush , respectively. a needs to consider the possibility of being exposed as a liar by blushing, as this diminishes the expected gain in reputation by telling a lie. Blushing happens with a fixed frequency $f_b := P(\text{blush}|\text{lie})$ for all agents, which is known to all agents. This procedure is illustrated in Figure 1. The quality of the message is then measured by the expected reputation of a after the conversation. For the honest statement, this is equal to the reputation of a in the eyes of the simulated dummy after the update under the honest information. For the lies, the expected value is calculated as the sum of the reputation after the update under the successful lie, multiplied with the frequency of not blushing, and the update after the lie followed by blushing multiplied with the frequency for that to happen.

The message promising the highest possible expected value is then actually sent to the receiver. As only one positive and one negative lie are simulated, the chosen message is in general close to but not necessarily the ultimate best option.

When talking about the receiver b , b 's mind update on a is not influenced by the content of the message but only by the amount of trust that he places in it to be true. Therefore, in this scenario, the white lie option or the truth happens to be the most promising (as currently the ToM-agent does not aim for friendship and its benefits). When talking about oneself, it is not possible to give a generalized answer about what lie, if not the truth, is the optimal message as it arises from a more complex interplay of parameters. Besides the intuitive result of lying positively being the most promising option in some cases, sometimes lying in the downward direction might boost the reputation even more. This is due to the RGS's concept of confession, meaning that if a receiver agent b receives a message from an agent a that presents himself worse than b currently thinks of a , b considers the statement to be ultimately true, as the only purpose of sending a potentially damaging message must lie in the intention of making a true statement. Therefore, if a manages to put a negative lie on his counterpart that lies only slightly below his opinion,

the negative impact of the messages content becomes overcompensated by the one certainly true statement that *b* adds to his belief on *a*.

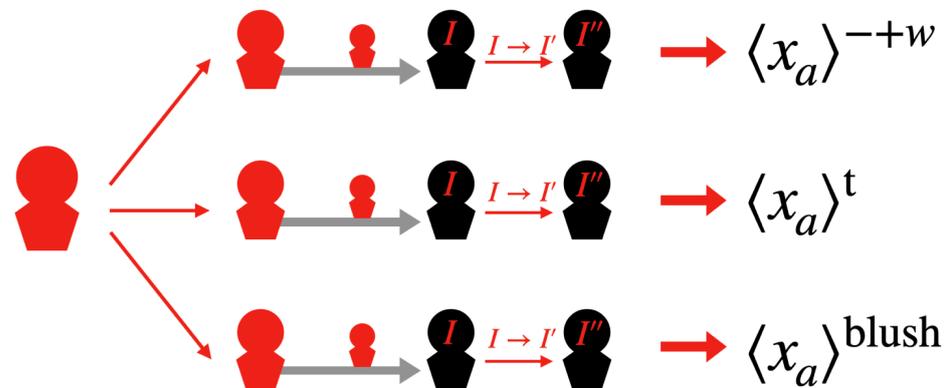


Figure 1. Red agent simulating black’s update of possible messages in a conversation about red, who imagines three conversations in which he tells the four options of +, −, w, and t, as well as the blush option to *b* and simulates the resulting reputation in the corresponding dummies.

In the case of talking about a third agent *c* that is not the receiver or the ToM-agent *a* himself, *a* not only influences *b*’s opinion on *a* but also on *c* and therefore also takes into account the possible subsequent messages from *c* to *b* about *a*. Depending on what *a* expects *c* to tell *b* about himself, it might be reasonable to accept a different effect on his reputation in the first communication with *b*. If *a*’s statement on *c* leads to a different reputation of *b* about *c*, then *c*’s positive or negative messages about *a* will be evaluated differently by *b* and the expected value for the message options changes. In order to investigate the effect of the conversation from *c* to *b* about the ToM-agent, he now imagines the four possible messages that *c* could tell *b* on the basis of its knowledge about *c*’s mental state, as shown in Figure 2. Those are, again, the honest statement, positive and negative lie, and a lie followed by unintended blushing. The honest statement *a* expects *c* to make is given by the first-order ToM I_{aca} , which represents what *a* thinks *c* to believe about himself. As *c* will construct his lies to *b* around his first-order values I_{cba} , *a* will construct his expectation for them around his second-order values I_{acba} .

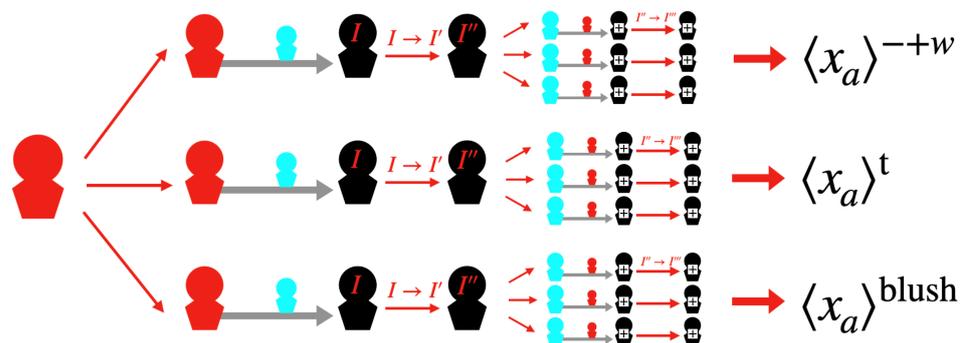


Figure 2. Red agent simulating black’s update of possible messages in a conversations about cyan, followed by simulations from cyan to black about red for each of the updated dummies. Each of the five dummies for the options of *a* is copied three times to simulate the effect of the subsequent message from cyan to black on red for each of the dummies from the first conversation.

For making predictions about *c*’s behavior, *a* assumes friendships to be symmetrical, meaning, if *a* sees *c* as a friend, he automatically assumes that *c* also considers *a* as a friend, applying the same logic if *c* is an enemy of *a*. This makes *a* expect *c* to tell *b* a positive lie about himself in the first case and a negative in the latter. If *c* is neither friend nor enemy in the eyes of *a*, then he expects positive and negative lies with probability of $\frac{1}{2}$ each.

The ToM-agent calculates the expected reputations for each combination of messages from himself about c , followed by a possible message from c to b . Then he evaluates the expected value for each of his options by considering the messages by c and the blushing to happen. The honest message that a expects c to make about a is equal to the first-order parameter I_{aca} . However, when b receives the message, he will extract the new information by comparing the content to his first-order I_{bca} . As we construct the ToM-agent so that he assumes that this value in the form I_{abca} is equal to I_{aca} , the content of the true messages in the subsequent messages in the simulation only influences the belief of b in the trustworthiness of c but not the belief in a , which is of actual interest. Therefore, the analysis of the strategy shows that it is advantageous for the ToM-agent to lie positively about agents that he considers to consider him a friend and negative about agents that he thinks see him as an enemy. Due to the assumption of symmetry, this is equivalent to the current rule of lying positively/negatively about agents the sender himself considers friend/enemy. Only if the first-order value for friendship and the second-order ToM values of I_{acba} were tracked and updated as independent parameters, the rule would become more differentiated. Anyhow, the current structure of the RGS does not provide the mechanism to do so. Moreover, the uncertainty of ToM values increases with the order, which was also a reason to introduce the bounded rationality in the first place.

2.5. ToM-Agent Update Rules

We also introduced additional information update rules that are unique to ToM-agents compared to the agents of the previous version. First, after sending a simulated message to b , the ToM-agent replaces his representation for b 's mind with the mind of the dummy that was updated under the chosen message.

We also assume the ToM-agents to be deaf towards messages about themselves. As their lie decision does not depend on random processes but on prudent considerations, we find it reasonable to assume the ToM-agents to be aware of their exact lie and truth count and therefore should not be influenceable in their self reputation. Anyhow, those messages are still used to update the ToM-agents' view on the honesty of the sender.

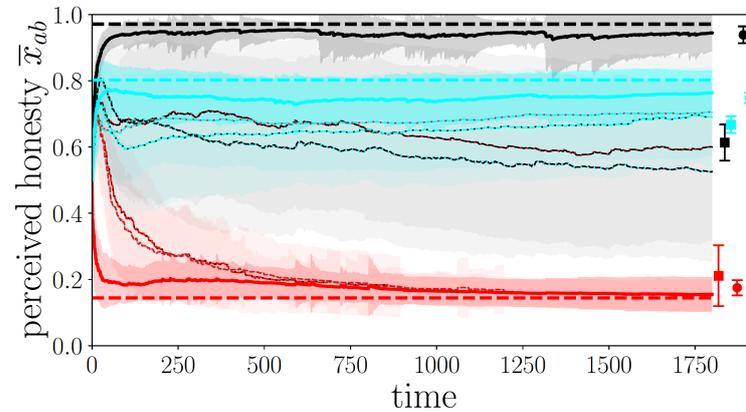
An agent that understands the general structure of a mind in the RGS so well that he is able to perform the described simulations should also be aware that other agents like himself perform an update on their self reputation after telling a message. Therefore, after receiving a message from a , the ToM-agent b will update his first-order belief I_{baa} . In a two-agent game, the ToM-agent is therefore capable of keeping track of its counterpart's exact mental state at all times, under a known start parameter. Moreover, a lie can be spotted easily if the content of the message does not precisely match the first ToM-agent's first-order ToM. In this constellation, ToM-agents therefore always manage to reach and keep the top reputation after a small number of rounds by effectively outplaying their counterpart. In larger groups in general, the first-order ToM representation is not equal to the real values, which is why we stick to smart lie detection (as described in Enßlin et al. [8]) for more than two agents.

3. Results

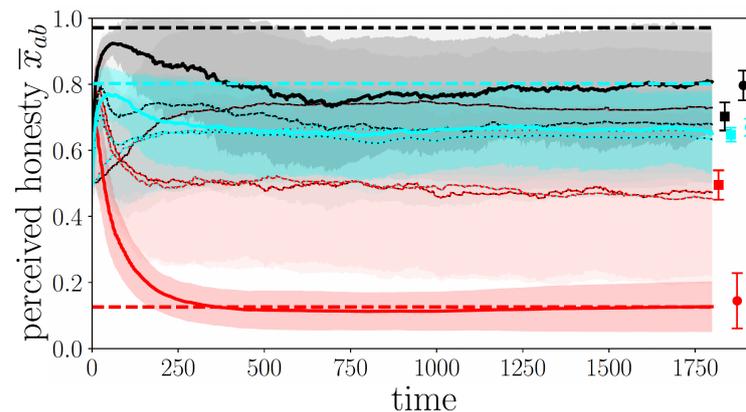
Figure 3 shows a reputation statistic for a ToM-agent (red) among two ordinary agents.

Ordinary agents had rather low but fixed lie frequencies, whereas the ToM-agent was free to lie whenever he saw this to be the most promising option according to the rules described above. On average, he showed to tell the truth in only about 14% of the conversations. For an ordinary agent with a comparable, fixed lie frequency, the reputation, as well as his self-reputation, converges against this low value. In the ToM case, the agent's self-reputation (by construction of the ToM update rules) always represents the real value. The way others see the ToM-agent's honesty improved significantly to a value of around 50%, meaning that others held a very distorted image about the ToM-agent by expecting him to tell honest statements in about half of the conversations. We also observe simulations where a very dishonest ToM-agent can even achieve nearly top reputations, but also

sometimes expose themselves as deceptive liars and therefore fail completely. As we are dealing with a highly chaotic system, it is often difficult to identify the exact reason for the shape of the reputation dynamic of a game. However, the success of the strategy was found to have a strong correlation with the quality of the ToM-agents' information on the others' beliefs, as shown in Figure 4.



(a) Ordinary agent (red) among ordinary agents



(b) ToM-agent (red) among ordinary agents

Figure 3. Comparison of the average achieved reputation and standard deviation for the ToM-agent with the ordinary agent in 100 simulations over the time of 300 rounds. The unicolored lines represent the agents' self-reputations, whereas the red line with black dots denotes the reputation of red in the eyes of black. The dashed lines stand for the honesty frequencies of the agents.

The range around the white lie, in that a lie in general turns out to be better than the truth in the first place, commonly happens to be a lot more narrow than the deviation of the first-order ToM towards the real belief described by it. If this occurs, the deception turns out to be ineffective regardless of the chosen lie option. Dropping a large number of lies close to their counterparts' first-order information and therefore confirming their counterparts' opinion can turn out to be disadvantageous for ToM-agents as future lies then need to be very close to this value. As the ToM-agent's amount of lies is not restricted by an intrinsic lie frequency, he can become very persistent in achieving his goal by permanently dropping lies on a specific agent. Especially in cases where those lies are based on a wrong set of ToM values and agents are confronted with a large variety of different opinions on a specific topic, this can create a high level of social chaos and a high volatility in reputations, especially the self-reputation of the most honest agents.

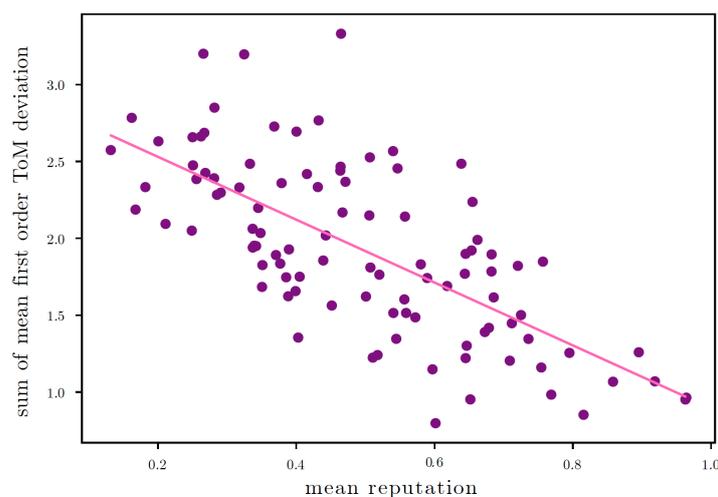


Figure 4. Correlation between the sum of the absolute deviations of first- and second-order Theory of Mind information from the real values and the success of the simulating ToM strategy in terms of the achieved mean reputation in the game.

Ideally, a ToM-agent would identify a statement to be a lie if its content deviates from his first-order ToM values for the sender's belief of the topic. One needs to take into account the possibility that the agent actually intended to send an honest message but the ToM value could be dramatically erroneous and therefore the message's content cannot be used as a measure for its truth. Instead, we take a product of the Kullback-Leibler divergence between the message and the first-order ToM as well as the ToM-agents' belief on the topic. Therefore, the ToM-agents still suffer from the so-called "echo chamber effect" already known from ordinary agents [8]. If a ToM-agent identifies a gain in reputation by manipulating an agent's mind in a certain direction and effectively influencing his counterpart to believe false information, the ToM-agent himself might adopt these false beliefs if the influenced agents themselves start to spread those false beliefs.

4. Discussion and Summary

We have set up a working model of a second-order Theory of Mind agent that uses this mental ability to deliberately deceive others with the goal of maximizing his own reputation, by choosing the communication option that promises the highest expected reputation after simulating different possibilities in the other agents' minds. As it turned out, such ToM-agents showed a strong tendency to tell lies and significantly increased their mean reputation performance compared to the conventional ordinary strategy. It must be noted the success in the individual scenario mostly depends on the quality of the ToM information used for the decision making.

Outlook

We consider the following to be reasonable future extensions of the existing model:

- Include an independent update on I_{abca} : These would give the ToM-agent the opportunity to evaluate the influence of the expected honest statement a third agent might tell to one's counterpart in a subsequent conversation and would lead to more differentiated lie decision rules.
- First-order ToM for friendship: If the ToM-agent holds a picture of their counterparts' friend and enemy lists, he might also have a more accurate belief on the direction of lies from third agents in subsequent conversations, as we found the assumption of symmetrical friendship to be true in many but by far not all cases.
- Choice of topic and conversation partner: Additionally to letting the ToM-agent decide which message about a fixed topic to send to a fixed receiver, one could also let him calculate the expected value for all messages to any agent about any topic in order

to exploit even more options for increasing one's reputation. Moreover, the option to withdraw from conversation could be added, as we found that in some situations none of the possible messages promise an expected reputation that is above the one the ToM-agent expects the receiver to currently hold.

Author Contributions: Conceptualization, F.S., T.E., V.K., S.U. and C.B.; software, F.S., T.E. and V.K.; formal analysis, F.S.; visualization, F.S.; supervision, T.E. and V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, J.Y.S.; Imuta, K. Lying and Theory of Mind: A Meta-Analysis. *Child Dev.* **2021**, *92*, 536–553. [[CrossRef](#)] [[PubMed](#)]
2. Böckler-Raettig, A. *Theory of Mind*; utb.: Stuttgart, Germany, 2019.
3. Ding, X.P.; Wellman, H.; Wang, Y.; Fu, G.; Lee, K. Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychol. Sci.* **2015**, *26*, 1812–1821. [[CrossRef](#)] [[PubMed](#)]
4. Warneken, F.; Orlins, E. Children tell white lies to make others feel better. *Br. J. Dev. Psychol.* **2015**, *33*, 259–270. [[CrossRef](#)] [[PubMed](#)]
5. Biziou-van Pol, L.; Haenen, J.; Novaro, A.; Liberman, O.A.; Capraro, V. Does telling white lies signal pro-social preferences? *Judgm. Decis. Mak.* **2015**, *10*, 538–548. [[CrossRef](#)]
6. Kathleen, O.M.; Carnevale J., P. A nasty but effective negotiation strategy: Misrepresentation of a common-value issue. *Personal. Soc. Psychol. Bull.* **1997**, *23*, 504–515.
7. Steinel, W.; De Dreu, C. Social Motives and Strategic Misrepresentation in Social Decision Making. *J. Personal. Soc. Psychol.* **2004**, *86*, 419. [[CrossRef](#)] [[PubMed](#)]
8. Enßlin, T.; Kainz, V.; Böhm, C. A Reputation Game Simulation: Emergent Social Phenomena from Information Theory. *Ann. Phys.* **2022**, *534*, 2100277. [[CrossRef](#)]
9. Enßlin, T.; Kainz, V.; Böhm, C. Reputation Communication from an Information Perspective. *Phys. Sci. Forum* **2022**, *5*, 15. [[CrossRef](#)]
10. Nagel, R. Unraveling in Guessing Games: An Experimental Study. *Am. Econ. Rev.* **1995**, *85*, 1313–1326.
11. Arthur, W.B. Inductive Reasoning and Bounded Rationality. *Am. Econ. Rev.* **1994**, *84*, 406–411.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.