



[Michael Macy and Yoshimichi Sato \(2008\)](#)

Reply to Will and Hegselmann

Journal of Artificial Societies and Social Simulation vol. 11, no. 4 11

<<http://jasss.soc.surrey.ac.uk/11/4/11.html>>

For information about citing this article, click [here](#)

Received: 02-Sep-2008 Accepted: 11-Sep-2008 Published: 31-Oct-2008



Keywords:

Replication, Social Dilemmas, Simulation Methodology, Cooperation, Trust, Agent-Based Modelling

Introduction

1.1

Will and Hegselmann ([2008](#)) (hereafter W&H) make several important points that we very much support:

1. Most importantly, they call attention to the need for replication of computational models.
2. They point out the importance of publishing source code and/or making the code publicly available.
3. They highlight the need for papers to have as much detail as possible about model assumptions, within the constraints of length (which has the additional benefit of keeping models simple enough to be explained in a few pages).
4. They point out details of our model that are inadequately explained in the paper.
5. They report monotonic effects of mobility on trust that are actually stronger than our own for the explanation we proposed for cross-cultural differences in trust—that mobility is higher in the U.S. than Japan.

1.2

Their critique grew out of a graduate seminar on agent modeling in which our paper was discussed. From our own experience teaching these courses, we are well aware how students delight in calling every assumption into question, challenging every specification, and suggesting better alternatives. This is a valuable learning experience and it can also lead to the discovery of serious flaws in a model. However, it is also the case that the critical orientation can lead to "false positives" due to strong incentives to find fault. Their criticisms of our model (quoted below in italics, followed by our responses) are almost entirely of that kind.

1.3

A lot is at stake: namely, modern western social life. And it is an agent-based computational model that bears the burden of proof for the threatening cultural diagnosis.

1.4

Their sarcasm is unwarranted. What is at stake is the "embeddedness of social relationships needed to make trust and trustworthiness self-reinforcing" (p. 7220). And as we make clear in the paper, it is empirical research, not an agent-based model, that bears the burden of proof (p. 7220). The computational model can generate hypotheses for empirical testing, but it cannot "bear the burden of proof."

1.5

Obviously the Macy-Sato-model is meant to answer a fundamental question about societal life: Why, when and how is it possible to build trust with distant people?

1.6

We disagree. Our model cannot answer this or any other fundamental question about societal life, nor was it meant to. The model is, however, "useful for identifying possible causal mechanisms for observed cross-societal differences" and "also useful for identifying empirical possibilities that have not yet been observed but which might nevertheless obtain were current trends to continue" (p. 7220).

1.7

... if one succeeds in replicating a model, one has nothing new to tell—and even worse, nothing new to publish. Therefore, investment in one's own models seems more promising than squandering time and effort on other's models! There is evidently an incentive problem.

1.8

We strongly disagree. "Investment in one's own models" can be a dead end as well, as can a laboratory experiment that yields a null result, or an expensive and time-consuming survey that turns out to simply confirm the current theory. There is always this risk in scientific research that one is in a *cul de sac*. For students learning a new research methodology, the effort to replicate previous studies pays an enormous dividend even if the replication turns up nothing interesting. The opportunity to publish a paper is icing on the cake.

1.9

In several emails we asked the authors to deliver more information and the source code but they did not.

1.10

We cannot imagine not providing source code to anyone who asked. We searched our email records and found two references -- a May 21, 2006 email from Hegselmann that mentioned his use of our paper in his modeling seminar, that he was not clear about the signal-detection algorithm in the model, and asked if we had a longer paper in which these details were elaborated. There was no mention of any interest in or attempt to replicate our model nor any request for source code. Macy interpreted the email as simply a friendly acknowledgment that the paper was being used by his students and a general expression of interest in the problem and approach. Macy responded accordingly, mentioning a follow-up project and the possibility for a more elaborate study at some point. Given that this was clearly a misreading of the inquiry, we would expect a follow-up email clarifying more precisely the reason for the request and the urgency. Indeed, Macy received a follow-up email a month later, yet once again there was no mention of replication, nor a request for source code, but the email again pointed to several specific questions about how the model implemented key features, which had come up in the seminar. Given the pressure of competing agenda items, and the considerable time required to refresh memory of a four-year-old project, Macy put this item at the back of the queue, and unfortunately, it never made it the front.

The failure to follow up on this was a serious oversight. In retrospect, Macy regrets not taking a more proactive stance by at least asking if the students might like the source code. That is one of the most important lessons we take from this experience.

1.11

There is no information on whether or not a parochial newcomer trusts its new neighbours or not. Our experiments indicated that we come a bit closer to Macy and Sato's results if we assumed that they are distrusting. This is plausible to some extent, in the same way as newcomers are strangers to the neighbours, the newcomer's new neighbours are strangers to the newcomers.

1.12

Their Figure 2 diagram is equivalent to the rule: "Distrust unless you use signs and your partner signals trustworthiness or you use relationship and your partner is not a stranger." Our rule is: "Distrust unless you use signs and your partner signals trustworthiness or you use relationship and your partner is not a newcomer." While both versions assume locals distrust newcomers, W&H also assume newcomers distrust locals, which we omitted as redundant since both sides must trust in order for exchange to be consummated. Thus, the two rules are behaviorally equivalent. We thought it sufficient that we specified that "parochial agents distrust strangers" (p. 7216) but in retrospect we recognize that we should have said "newcomers" instead of "strangers" to be more precise.

1.13

... an agent with $P_C = 0.01$ that ends up in mutual defection cooperates in the next time step with a probability of 0.208... Thus, the agent in mutual defection refines its propensities in a way that is likely to make its state even worse. This implication does not sound plausible anymore.

1.14

Not true. First, P_C (probability of cooperation) is also influenced by the behavior of neighbors who are more successful, as our paper makes very clear (p. 7216) and W&H explicitly acknowledge elsewhere. An agent with $P_C = 0.01$ who has a successful neighbor also with $P_C = 0.01$ can be influenced to remain non-cooperative. Second, W&H are assuming the agent will remain paired with a defector. As we explain in the paper (p. 7216), agents are randomly paired for only a single exchange, which W&H also explicitly acknowledge elsewhere. By "state" that is "even worse" they mean the payoffs in a Prisoner's Dilemma, which are $T > R > P > S$, to which we add the "X" payoff for choosing to "exit" (due to distrust). An agent who earns the "P" payoff (for mutual defection) may be paired next time with a partner who prefers to cooperate, in which case, switching to cooperation will not "make its state even worse" (since $R > P$). And even though the agent might have earned T (by defecting on a naive trustworthy partner), there is the possibility that the partner might have detected the intention to defect and chosen to exit, inflicting a payoff of $X < R$. Third, we object to their unsupported assumption about what behaviors are plausible. In claiming that a 0.208 probability of cooperation following punishment for mutual defection is "implausible," W&H imply that humans always play myopic best-response strategies and never explore ways to avoid repetition of aversive stimuli. They ignore 50 years of behavioral research to the contrary, including studies of reinforcement learning, probability matching, and cooperation in prisoner's dilemma. Evolutionary theory also points to the mate-attraction advantages of risk-taking cooperative behavior. The rationality of occasionally extending an olive branch, given the long-term benefits of living in a highly cooperative neighborhood, are also noted throughout the New Testament and in the writings of Ghandi, Martin Luther King, and other advocates of pro-social behavior. That W&H believe this behavior is implausible makes us glad that we do not live in Bayreuth, which appears to be a very nasty place. Finally, we are curious why W&H now insist on the empirical plausibility of agent based models. Does this mean we will no longer find their own agents living on brightly colored checkerboards?

1.15

It seems natural to expect agents to rate payoffs in an inversely proportional relation to the amount of time that has gone by since they were gained. This is especially true because otherwise present payoffs would have less impact with every iteration ticking by.

1.16

Not true. W&H are assuming that all payoffs are positive, which would indeed cause present payoffs to have declining impact over time. As we explain very clearly in the text (p. 7216), the S and P payoffs are fixed at negative values and R and T depend on O (the opportunity cost of choosing a partner from a smaller pool, as given in the original equation 1). Each agent's running total can increase or decrease over time or remain roughly stable, depending on the sign of the payoffs. Hence, present payoffs could have less impact, more impact, or equal impact with every tick. We also object to yet another unsupported assertion about what is "natural" or "plausible" behavior. If a neighbor has a more expensive house and car, do we discount their comparative advantage by taking into account the year in which they earned the money they used to make these purchases? Do we keep track of the time stamps on each of our neighbor's earnings? In calculating the distribution of wealth, why do statisticians use the running total instead of discounting by the time that has passed since the money was earned?

1.17

Therefore, it is an a priori truth that the levels of cooperation and market size, understood as means over the population, cannot fall to near zero... if one wants to show that a population can move to a state of high trust, cooperation and market size, it does not seem to be a very good idea to make use of assumptions that exclude situations without substantial amounts of these.

1.18

Not true *a priori*. There is also an equilibrium in which every agent trusts and cooperates parochially and no one ever enters the market, hence market size remains near zero (the initial value). Moreover, we did not "want to show that a population can move to a state of high trust..." as W&H incorrectly suggest. As we make abundantly clear throughout the paper, our interest is in the effects of mobility on trust, not in the emergence of trust regardless of mobility. Since agents learn to cooperate and to enter the market based on behavioral assumptions that are held constant across all levels of mobility, the effects of mobility cannot be attributed to these behavioral assumptions.

1.19

"...we learn that they "decide whether to cooperate" and "give the appropriate signal" (Mac2002, p.7217[1]). It is not clear, what "appropriate" means in this context..."

1.20

True, we should have used "corresponding" instead of "appropriate." Nevertheless, W&H interpreted "appropriate" appropriately. They are correct that defectors are not able to mimic cooperators, and an arms race between detection skills and deception skills cannot take place, even though this is an important possibility for which future research is needed, as we note in the discussion (p. 7220). Clearly, the ability for trust to be established among strangers depends on the ability to detect cheaters with sufficient accuracy (whether based on an involuntary cue, or the ability to detect better than others can deceive, or some kind of reputation system that tags agents). The "imperfect vision" in our model is intended to capture a range of ways in which those who rely on "telltale signs" are vulnerable, for one reason or another. But for simplicity, we do not model each of the possible reasons.

1.21

...it would be perfectly plausible that their decision on cooperation can depend on whether or not they know their partner. Unfortunately Macy and Sato exclude this possibility ...

1.22

We do not exclude this possibility, we simply let the agents decide whether they would cooperate more frequently in local neighborhoods than in the global market. In the footnote on p. 7216 we elaborated our rationale for allowing partner familiarity to inform trust but not trustworthiness. The purpose of this modeling exercise was to find out if local neighborhoods can become "classrooms" in which agents acquire the ability to successfully navigate the open market (p. 7214). This ability includes not only learning to trust strangers based on "telltale signs of character" but also having the character traits on which the signs are based. Accordingly, we posited trustworthiness as an acquired character trait that others can learn to detect, rather than a situation-dependent conditional strategy based on familiarity.

1.23

There are hints in the text that it is the "ongoing" character of local interaction that causes the minimization of transaction costs (Macy and Sato 2002, p.7216[3], p.7219[5]). But how can a relation in the model be ongoing if in "a new iteration" agents "randomly choose a partner from the available pool (either the market or the neighborhood)" (Macy and Sato 2002,p.7216[last])?

1.24

This is a matter of definition. By "ongoing," we mean a relationship that is "stable," which we define on p. 7214 as what Axelrod calls "the shadow of the future"—a high probability to re-encounter the same partner. An alternative definition is that agents have only one partner for the duration of the experiment. The latter is clearly not possible if partners are re-assigned at each iteration, as W&H point out. However, re-encounters are highly likely in small neighborhoods, as are pairings with the partners of one's partners (p. 7215).

1.25

Passages, describing relations in the neighbourhood as "embedded", point in the same direction (Macy and Sato 2002, p.7214[3,4,5,footnote1], p.7215[2,4,8], p.7216[5],p.7217[6], p.7219[2,3,7], p.7220[1]). This embeddedness is described as being characterized by "stability and transitivity" (Macy and Sato 2002, p.7214[footnote1]) but there are no hints on how this could have been implemented.

1.26

The paper is very explicit that stability and transitivity are an immediate consequence of neighborhood size. On p. 7215, we explain how small neighborhoods increase the stability and transitivity of relations and how stability creates a "shadow of the future," or prospect of future interaction. The logic is straightforward. The smaller the neighborhood, the fewer the neighbors, and the fewer the neighbors from whom a partner is randomly chosen, the higher the probability of being randomly paired with a previous partner and the higher the probability of being randomly paired with a partner of a previous partner.

1.27

... the replicated model lacks the implementation of transaction costs. Thus it is unlikely that we simply failed to understand how they are caused by the other elements of the model.

1.28

In the text, we defined transaction costs as the likelihood of being cheated by one's partner (p. 7215). We did not wire in a pre-determined difference in transaction costs (or propensity to defect) between neighborhood and market. Rather, this difference is an emergent property of agent interaction. It depends in part on whether agents learn to cooperate or not. This in turn is a function of neighborhood size and agent mobility (the two parameters we manipulate in our experiments). The larger the neighborhood, and the greater the mobility, the lower the probability that agents will coordinate a self-reinforcing equilibrium based on high trust and trustworthiness, as we explain on p. 7215. The size of the market is also endogenous and depends on whether agents learn to leave the neighborhood, with its limited opportunities for finding an optimal partner. As the market grows in size, the opportunity cost of local interaction increases while the transaction cost of

global interaction also increases. Our model thus captures the trade-off between opportunity costs and transaction costs in Yamagishi's theory of trust. Because our primary theoretical interest was in the effects of mobility on trust, we chose not to directly manipulate transaction and opportunity costs but instead to manipulate mobility and to then observe the effects on opportunity and transaction costs as these influence the emergence of trust and trustworthiness.

1.29

...where the size cannot be represented by an integer, we ran a share of simulations with the next smaller and a share with the next larger whole number of neighbourhoods.

1.30

Our implementation was similar but a bit simpler. We simply rounded down non-integer values for the number of neighborhoods. We should have indicated this in the text.

1.31

The term 'equilibrium' indicates that in their simulations the simulations end in a situation in which the agents' action vectors no longer change. Unfortunately the combination of the chosen Bush-Mosteller learning algorithm and payoff matrix only allows for three stable situations:

1.32

This definition of equilibrium as a condition of stasis applies to a Nash equilibrium, but not to the dynamic equilibrium of the Bush-Mosteller model. In dynamic models, equilibrium is characterized by balance rather than stasis. Individual agents continue to change behaviors yet the population mean settles into a meta-stable state. We thought the dynamic understanding of equilibrium was evident from the context, but it appears that might not be the case for readers who are unfamiliar with previous papers by Macy and Flache explaining the equilibrium dynamics of reinforcement learning, including a paper on this topic in the same issue of *PNAS* (a [replication](#) of which appears in the March 2008 issue of *JASSS*).

1.33

Macy and Sato's main results stem from simulations with heterogeneity of one. For this amount of heterogeneity, we find that in comparison to the possible sizes of the global market, the applied neighbourhood sizes between 10 and 100 have little influence on the opportunity costs...Thus Macy and Sato's conclusion can only be justified if they give reasons why the chosen formula is the most adequate.

1.34

Their Figure 9 is misleading because it omits the size of the alternative—a market in which all $N=1000$ agents participate. On p. 7216 we explain that the opportunity cost O decreases linearly from 1 to 0 as n (the size of the pool) increases from 1 to N . (In a lottery with N tickets, the chance of winning increases linearly from 0 to 1 as the number of tickets purchased increases from 0 to N .) If $N=1000$, O decreases by 0.1 for every 100-agent increase in n . The justification for the formula is given in the text (p. 7215-6), in which we postulate that the chances to find an optimally complimentary exchange partner increase linearly with the size and heterogeneity of the pool from which the partner is chosen, assuming maximum heterogeneity ($h=1$).

1.35

Macy and Sato's use of the term social mobility is rather unusual. It generally refers to intergenerational mobility up or down the class hierarchy or income scale and not to mobility among different spatial partitions of the population.

1.36

W&H's use of the term "social mobility" is rather unusual. Intergenerational changes in income are generally referred to as "economic mobility," not social. According to Anthony Giddens, "The term 'social mobility' means the movement of individuals and groups between different socio-

economic positions" (Sociology, 5th edition, p. 328). When people change positions, they also change the set of others with whom they regularly interact. In contrast, the Japanese lifetime employment system greatly restricts the set of ongoing relationships that an individual can expect to have (p. 7219). We believe the text is very clear that we are not referring to the effects of intergenerational changes in income, and that the mobility we refer to is within generations and involves changes of network location rather than income. Moreover, we refer not to "social mobility" but to "social and spatial mobility" (p. 7214).

1.37

The replicated model qualitatively reproduces the positive effect of moderate rates of mobility on trust in strangers...There is, however, one major difference. While the plot from the original data indicates a break down of trust in strangers if mobility is set to 1, this is not the case in the replicated model.

1.38

In retrospect, we should have disaggregated hypothesis 1 (on the effects of mobility) to parallel the disaggregation of hypothesis 2 (on neighborhood size). That would have allowed us to separately test the positive and negative effects of mobility on trust. It would also more closely capture the theoretical motivation of the paper—to test an extension of Yamagishi's theory of why trust in strangers is lower in Japan. Our hypothesis is that the lower trust in Japan is due to lower mobility. The breakdown of trust we observed at high levels of mobility contradicts our explanation. Although one would never suspect this from their paper (and W&H never point it out), their monotonic results provide stronger support for our explanation of cross-cultural differences than do our own, because their results show only the positive effect of mobility on trust. We are certainly very pleased to see that an independent team, working with a large number of possible implementations, found the positive effect of mobility to be highly robust. As our results also make clear, the negative effect of high mobility is only apparent at the upper limit. When we experimented with various parameter settings, we also found that the negative effect does not always obtain—nor for that matter does the positive effect. These computational experiments should be regarded only as existence proofs that show how mobility could explain higher trust in the US than Japan, under specified scope conditions. Even so, it is equally important to also recognize the possibility that there might also be too much mobility at some point. This possibility seems highly intuitive and gets some support—albeit limited—from our results. When all agents are mobile, there is little difference between the local neighborhood and the open market—no opportunity to coordinate a high-trust local equilibrium and no opportunity to spread that equilibrium to other neighborhoods through the imitation of successful role models.

1.39

To conclude, we applaud the effort W&H invested in replicating our model, and regret the misunderstanding in our correspondence with Hegselmann. The lesson we take from this is the need to post source code for readers to directly download, which we have done for our model: ftp://hive.soc.cornell.edu/mwm14/webpage/Trust_PNAS.pas. We also appreciate their identification of places where we could have elaborated the verbal description of the model. However, we did not find as much merit in the long list of objections they raised to our definitions (e.g. mobility, equilibrium, on-going, appropriate), behavioral assumptions (e.g. reinforcement learning, cumulative payoffs), and experimental design (e.g. endogenizing opportunity and transaction costs and local trustworthiness rather than wiring these in). Their analysis of our model might have been more useful had they focused more carefully on the central question: whether — and under what conditions — the effects of mobility on trust are or are not monotone.



References

WILL, Oliver and Hegselmann, Rainer (2008). 'A Replication That Failed — on the Computational

Model in 'Michael W. Macy and Yoshimichi Sato: Trust, Cooperation and Market Formation in the U.S. and Japan. Proceedings of the National Academy of Sciences, May 2002'. *Journal of Artificial Societies and Social Simulation* 11 (3) 3 <http://jasss.soc.surrey.ac.uk/11/3/3.html>.

[Return to Contents of this issue](#)

© [Copyright Journal of Artificial Societies and Social Simulation, \[2008\]](#)

