Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds

COMPARING AGENT-BASED MODELS ON EXPERIMENTAL DATA OF IRRIGATION GAMES

Jacopo A. Baggio
Marco A. Janssen

Center for the Study of Institutional Diversity
School of Human Evolution and Social Change
Arizona State University
Tempe, AZ 85287, USA

ABSTRACT

Agent based models are very useful tools for exploring and building theories on human behavior; however, only recently have there been a few attempts to empirically ground them. We present different models relating to theories of human behavior and compare them to actual data collected during experiments on irrigation games with 80 individuals divided in 16 different groups. We run a total of 7 different models: from very simple ones involving 0 parameters (i.e., pure random, pure selfish and pure altruistic), to increasingly complex ones that include different type of agents, learning and other-regarding preferences. By comparing the different models we find that the most comprehensive model of human behavior behaves not far from an ad hoc model built on our dataset; remarkably we also find that a very simple model presenting a mix of random selfish and altruistic agents performs only slightly below the best performing models.

1 INTRODUCTION

Behavior in social dilemmas cannot be explained by the traditional model of rational and selfish decisions makers. Humans cooperate to a much larger extent than can be expected from predictions based on *homo economicus* (or economic human). However, no generally agreed alternative theory has been developed. Experiments on decision making have been used to test hypotheses to build alternative theoretical frameworks (Camerer 2003). Those experiments have revealed important insights into the importance of learning, other-regarding preferences, risk aversion, etc.

Our own research focuses on social-ecological systems and we study especially common-pool resource problems. Experiments performed by Ostrom and others show that participants do not behave as selfish rational actors in common-pool resource dilemmas (e.g., Ostrom et al. 1994). Since the late 1990s agent-based models have been used to explore alternative models of human behavior (e.g., Deadman 1999; Poteete et al. 2010).

Experimental data provides individual level data within a well-constructed problem. Therefore, such data might can be used to develop empirically-grounded agent-based models (Janssen and Ostrom 2006). However, there is the challenge of how to compare the small sample of observations with a stochastic model. Which metrics do we use for our comparison and how do we weight them? It is a common practice to calibrate one structural model to the data (e.g., Arifovic and Ledyard 2012). However, we will test alternative structural models to the data to compare a variety of possible explanations.

The aim of this paper is to demonstrate a systematic comparison of alternative models and the challenges we experienced. We use data from so-called irrigation games since those are the kind of experiments we do ourselves. Not only do we have a good understanding of the data, we have also experienced the limitations in using the traditional econometrics analysis (Rollins et al. in preparation). Our experiments let participants make multiple decisions during each round in a particular order. Since econometric models are based on the independence of the assumptions, there is a limitation in using econometrics to understand the interlinked mechanisms. We first discuss the irrigation games, and the basic results of these experiments. Then we discuss the models we compare and how we will compare them. We then present results and perform some sensitivity analysis. Finally, we conclude with lessons learned on model comparison and provide some suggestions for future work.

1.1 Irrigation Games

Our research focuses on the robustness and governance of social-ecological systems, and we use small-scale irrigation systems as a model system. Over the years we have done experiments in the lab and field on a variety of questions that are beyond the scope of this paper. For the purpose of this paper we use data from laboratory experiments of so-called irrigation games (Janssen et al. 2012).

In the irrigation game participants have positions A, B, C, D or E. A has the first choice to harvest water from the common resource. Then B has the next turn to harvest water from whatever amount was left by A, and so on. The order of the five players is randomly determined before the first round and remains fixed over the rounds of the game. Participants receive an endowment ω of 10 tokens in each round. First each participant makes a decision x_i on how much to invest in a public fund that generates the infrastructure and therefore determines the amount of water available for the whole group to share. In Table 1, the water provision generated is defined as a function *f*() of the total investments of the five participants. This production function is based on the challenge actual irrigation systems face. Upstream participants cannot generate an irrigation system alone, and therefore they will need the contributions of downstream participants.

Experiments used for this paper include variability of water availability given a certain level of infrastructure (Rollins et al. in preparation). The variability is calculated as the probability of having low, medium or high water availability. We distinguish different levels of uncertainty. When variability is low we use either the low, medium or high water availability column of Table 1 with probability 1/6, 2/3, 1/6, respectively, in order to define whole group water availability. In case of high variability, we use the low, medium and high column of Table 1 with probability 1/3, 1/3, 1/3, respectively, in order to assess group water availability.

Table 1: Water production as a function of units invested in the public infrastructure.

Total tokens invested into Public Fund	**Water Available**		
	Low **LowVar: 1/6** **HiVar: 1/3**	**Medium** **LowVar: 2/3** **HiVar: 1/3**	**High** **LowVar: 1/6** **HiVar: 1/3**
0 – 10	0	0	0
11 – 15	2	5	8
16 – 20	8	20	32
21 – 25	16	40	64
26 – 30	24	60	96
31 – 35	30	75	120
36 – 40	34	85	136
41 – 45	38	95	152
46 – 50	40	100	160

Second, each player, in sequential turns from upstream to downstream players decides how much to extract from the water available to her, that is, the water produced minus the water extracted by those before her in the sequence. Each token kept (not invested) in the first stage has a monetary value for the player that is equal to the value of each unit of water extracted in the second stage.

This experiment includes a first dilemma of upstream participants who need the contribution of downstream participants to maintain the structure of their common resource, which is crucial for the production of water in the game. However, the downstream participants can only obtain benefits from the resource if upstream participants avoid the temptation to deplete the common resource and leave little water for downstream players.

Experimental data used in this paper do not include scenarios in which participants can communicate or coordinate. After all five have made their investment decision, the total water available for the extraction phase is announced. Then, participants can only see how much water is available to them before they decide how much to extract. Hence, individual levels of decisions or extractions are not known. The water left by participant E will not be available for future rounds. In practice no E participant ever left any water behind.

The game is asymmetric since upstream participants have first access to the resource compared to downstream participants. Under this asymmetric game, participants first experience a provision dilemma in the contributions stage, and then face a resource appropriation dilemma when they extract from the generated resource. The earnings of the participants are the result of provision – x_i – and extraction – y_i – decisions, and the resulting payoff z_i for player i is defined as

$$z_i = \omega - x_i + y_i \qquad (1)$$

where: $$\sum_{i=j+1}^{5} y_i \leq f(\sum_{i=1}^{5} x_i) - \sum_{i=1}^{j} y_i \qquad \text{for } j = 0, 1, 2, 3, 4$$

Rational, self-interested individuals would not invest in infrastructure provision in the first stage. Since the upstream participant is expected to collect the whole resource, downstream participants will not invest. For participant A there is no benefit to invest when others don't. If this is the reasoning of the participants in the last round of experiment we find via backward reasoning that the same happens for all earlier rounds. Thus, the Nash equilibrium for this game is that no one invests and all receive 10 tokens for group earnings of 50 tokens.

To define the cooperative solution we calculate the maximum amount of the infrastructure plus the tokens not invested. There are multiple social optimum outcomes. For a 41-token investment, a resource of 95 tokens is generated, and for a 46-token investment a resource of 100 tokens is generated in each round. The total earnings of the group in the cooperative solution amounts to 104 tokens, doubling the social earnings of the Nash equilibrium.

1.2 Experimental data

We use data from experiments run at Arizona State University with 22 groups of five individuals each. A detailed analysis of the experimental results can be found in Rollins et al. (in preparation). Results from the experiments are shown in Figure 1.

The inequality can also be quantified by Gini coefficients. We calculate the Gini coefficients for investments as well as extractions. The average Gini coefficient for investments is 0.30. The average Gini coefficient for extractions is 0.38.

Analysis in Rollins et al. (in preparation) shows that the order of the treatments has no significant effect. Figure 1 reports the main results regarding investment and extraction during the experiments. Treatments are not statistically significantly different, thus the order in which the different levels of uncertainty have been introduced did not affect the results. We calibrate models on all treatments simultaneously. This mean that the outcome is one model that aims to explain the observed patterns in all treatments

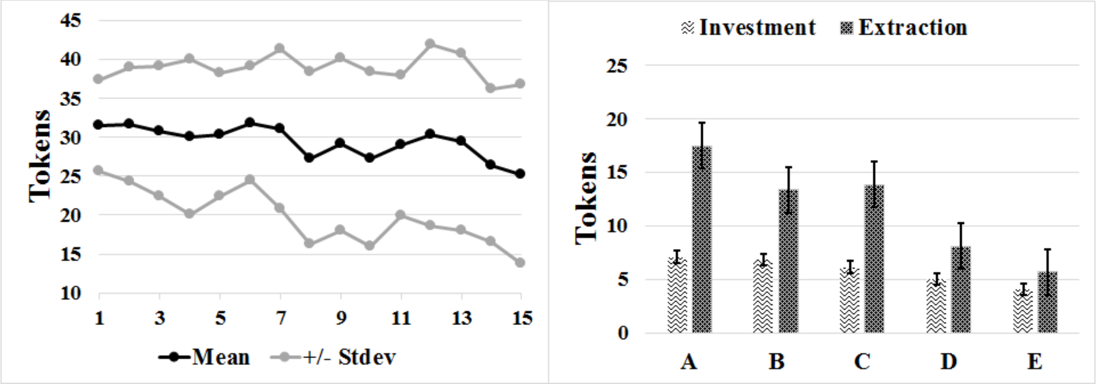

Figure 1: Average group investment per round with +/1 standard deviation and average contribution and collection level for each position with standard errors.

2 MODELS

We compare the outcome of different models grounded in theory and simulated against the experimental data. We start by comparing three null models, acting as benchmarks to more complex models with multiple parameters needing calibration.

The models define agent types where each agent type is a representative agent for that type defined by a set of parameters. Some of the models below define two or three types of agents in the population of players. The calibration procedure estimates the mix of different agent types.

If all agents are of the same type, they do not all make the same decisions. The reason is that decisions of most agent types are probabilistic. We also implement a version of calibrated models that assumes that parameters for the agent behavior varied among the agents using normal distributions among the average parameter values. We do not find this to improve the performance of the models significantly. Hence we only discuss the model with homogenous agent types.

2.1 Null models

<u>Selfish</u>

- Investment: all agents will always invest 0.
- Extraction: all agents will always extract the maximum possible amount.

<u>Altruistic</u>

- Investment: all agents will always invest 10 tokens (= the maximum available).
- Extraction: all agents will take an equal share of the amount of resource available. Hence agent in position *i* will take a share equal to 1/(6-*i*) of the amount available to *i.*

<u>Random</u>

- Investment: agents invest an integer amount uniformly distributed in the interval [0 10].
- Extraction: agents extract an amount uniformly distributed between 0 and the maximum amount available to them.

2.2 Calibrated models

Mixed Strategy (mixedrsa)

- Agents can act according to selfish, altruistic or random behavior according to probability p_s (probability of being selfish), p_r (probability of being random) and $p_a = 1 - (p_s + p_r)$ (probability of being altruistic). After an agent is classified as selfish, altruistic or random, he follows the rules of investment and extraction as described in section 2.1.
- We calibrate the fraction of agents in the population using the selfish, altruistic and random strategy in order to minimize the difference between simulated and observed behavior in investment and extraction.

Base level plus trembling hand (pseudorandom)

- Investment: in the first round inv is derived from a truncated random normal distribution between 0 and 10 with mean m_{inv} and standard deviation $\mathrm{stdev}_{\mathrm{inv}}$. In subsequent rounds (i.e., from round 2 to 15) investment levels are defined by the investment level of the first round plus a noise term drawn from a Gaussian distribution with mean 0 and standard deviation $\mathrm{stdev}_{\mathrm{n}}$ Investment is always bounded between 0 and 10.
- Extraction: agents extract an amount corresponding to the equal share plus a noise term drawn from a Gaussian distribution with mean 0 and standard deviation $\mathrm{stdev}_{\mathrm{n2}}$.Extraction level is bounded by availability. We calibrate mean and standard deviation of the initial investment distribution and the standard deviation of the added noise in the investment and extraction phases.

Simple rules based on statistical analysis (heuristic)

- Initial investment : investment is drawn from a random uniform distribution between 0 and 5 with probability p_s and from a random uniform distribution between 6 and 10 with probability 1-p_s
- Investment: $inv_t = tr * w_i^{\Delta ie_{t-1}}$ where tr = trust parameter in the interval [0,1], w_i = weight given to Δie_{t-1} (difference between extraction and investment in the previous round). Investment levels are defined only in the interval [0,10].
- Extraction: $ext_t = (1/(6-i))^{w_e}$ where i = position of agent assuming value 1, 2, 3, 4, or 5 and w_e = parameter that weighs the importance of position in extraction. Extraction levels cannot exceed resource availability for player i.
- We calibrate the weights given to investment and extraction and the variable representing trust.

Other-regarding preference model with two types of agents (utilitarian):

This is the most complex and the most comprehensive model based on findings from behavioral economics (e.g., Charness and Rabin 2002; Camerer 2003; Arifovic and Ledyard 2012). We assume two types of agents. The model (from here on, utilitarian) has a probability p_s to have complete selfish agents (defined as agents having $\alpha=\beta=0$). We assume that agents maximize their utility. This utility u_i is formalized in a general way to include different types of other-regarding preferences:

$$u_i = z_i - \alpha_i \cdot \max(z_i - \bar{z}_{-i}, 0) + \beta_i \cdot \max(\bar{z}_{-i} - z_i, 0) \qquad (2)$$

where:

- α_i and β_i are drawn from the interval [-1,1].

z_i is agent i's earnings, and $\bar{z}_{-i}$ is the average earnings of the other agents in the group. α can be regarded as the strength of an individual's aversion to exploiting others, and β can be regarded as an individual's degree of altruistic tendency. A lower value of β compared to α implies that a player gives a larger weight to his own payoff when his payoff is smaller than the average payoff of others compared to when it is larger. In line with Charness and Rabin (2002), we can define the following cases for $\beta \leq \alpha \leq 1$: Case 1: When $\beta \leq \alpha \leq 0$, the player is highly competitive. Case 2: When $\beta < 0 < \alpha \leq 1$, the player prefers payoffs among all players to be equal. Case 3: When $0 < \beta \leq \alpha \leq 1$. The player feels guilt earning more than others, and gains a sense of pride in acting altruistic. Case 4: if $\alpha = \beta = 0$, we have the condition in which a player cares only about his or her own welfare.

In order to define the investment decision, agents are assumed to estimate the expected utility based on the expected behavior of others.

The expected investment level of others is equal to

$$\hat{x}_{-i} = \omega \cdot 4 \cdot \eta_i \tag{3}$$

where η_i is the expected level of cooperation by other agents, which is within the interval [0,1].

This enables each agents to estimate an expected level of the public infrastructure, $\hat{p}_i$. For each level of investment x_i, the expected level is

$$\hat{p}_i = x_i + \hat{x}_{-i} \tag{4}$$

Agents predict how much of the resource would be available to the group using the production function of Table 1 with the expected value $\hat{p}_i$.

How much is expected to be available to agent *i* depends on how much upstream agents have taken from $\hat{p}_i$. The lower the level of cooperation they expect from the other participants, representing here the upstream participants, the less she expects to receive from the resource before it is her turn. Hence agents assume that an amount $\hat{y}_i^A$ is available for agent *i*.

$$\hat{y}_i^A = \hat{p}_i \cdot (1 - (\frac{i-1}{5})^{(2-\eta_i)}) \cdot \tag{5}$$

If agent *i* expects that other agents are cooperative, $\eta_i = 1$, they will take an equal share from the resource. If they expect the other agents will be less cooperative, it is then expected that they will take more than an equal share.

In rounds 2 to 10 a simpler estimation technique is used by the agent to determine $\hat{y}_i^A$. The agents are assumed to expect the upstream participants take a share s_i from the expected resource size.

$$\hat{y}_i^A = \hat{y}_i^A \cdot s_i \tag{6}$$

We use the values of α_i and β_i to define how much the agent takes from the share that is expected to be available to her. Agents who are selfish are expected to take the whole amount of available resources, but those with other-regarding preferences are expected to take a lower level.

$$\hat{y}_i = \hat{y}_i^A \cdot (1 - \alpha_i) \cdot (1 - \beta_i) \tag{7}$$

Now the agent can define her utility of investing x_i and receiving $\hat{y}_i$ from the resource. Using the expected earnings, we can estimate the expected utility for agent *i* for each level of investment. Based on the expected utility levels, agents make a probabilistic choice of how much to invest

$$\Pr(x) = \frac{\exp(\lambda \cdot u(x))}{\sum_X \exp(\lambda \cdot u(X))} \tag{8}$$

where Pr(x) is the probability of investing an amount x in the public fund and λ is the weight given to the utility values. If λ is 0 all options have an equal probability, while if λ is equal to infinity the agents choose the option with the highest expected utility. In the numerical calibration of the model we use an upper bound of 5.

Based on the investment decisions of the agents the actual level of the public infrastructure *p* can be determined. Now, each agent decides how much to collect based on the available resource during the turn in which she can make the decision. Similarly to the investment decisions, the expected utility for each level of collection is determined, and decisions are made from upstream to downstream.

The agents update the expected level of cooperation η_i based on the information they received on the average investments of the other agents. The learning parameter τ_1 defined the speed of learning. If τ_1 is equal to 1, η_i remains the same, and agents are therefore assumed not to learn, while if τ_1 is equal to 0 agents assume that the level of cooperation in the next round is the same as observed in the current round.

$$\eta_i = \eta_i \cdot \tau_1 + (1-\tau_1) \bar{\cdot} \frac{\bar{x}_{-i}}{10} \tag{9}$$

Similar to the share that upstream agents are expected to extract, we assume that agents update the value of s_i based on the observed share, where τ_2 is a learning rate.

$$s_i = \cdot s_i \tau_2 + (1-\tau_2)\frac{y_{i,t-1}}{p_{t-1}} \cdot \tag{10}$$

2.3 Defining the fit

We calibrate the models described on the experimental data. We use the standard genetic algorithm of *BehaviorSearch.org* for the model that is implemented in Netlogo 5.0.3. The model code and documentation can be found at http://www.openabm.org/model/3854/version/1/view. For the fitness evaluation of each parameter configuration we run the model 100 times the number of groups present in each treatment (6, 5, 5 = 16 in total). We run each model 1600 times. For each of the groups we compare simulated and actual data of group investment per round, investment per position, extraction per position and Gini coefficient for investment and extraction.

The fit between the model and the data is based on the normalized squared difference between simulated and observed data. We scale all data used for calculating the metrics to be between 0 and 1 and define the fit as 1 minus the squared difference between simulated and observed metrics. Hence for each of the metrics included, we calculate the fitness score between 0 and 1, using

$$f_i = 1 - (d_e - d_s)^2 \tag{11}$$

Where the data of the experiments, d_e, and simulations, d_s, are scaled to values between 0 and 1. Then the fitness values of all 5 metrics are aggregated to derive the final fitness score used in the calibration. We assign equal weight to all 5 fitness measures.

The metrics used to evaluate the performance of the model include (please see online supplementary material for more details on how fitness measures are calculated):

- Average group level investments in the public infrastructure level over the 10 rounds (f_1).
- The average contribution per position (f_2).
- The average collection per position (f_3).
- The average Gini coefficient of contributions (f_4).
- The average Gini coefficient of collected tokens (f_5).

Aggregating fitness measures:

There are different ways to aggregate the individual fits with the indicators. However, we use the most conservative approach (i.e., the approach that, in value, gives the lowest fitness, thus penalizing more heavily low fitness in one of the five metrics used to evaluate the performance of the models. Therefore we use a multiplication function between the 5 fitness values. Janssen (in press) compared different fitness measures and did not find a qualitative difference in the results. Formally, (12) reports the equation by which the fitness measure is calculated:

$$f^{mlt} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5 \tag{12}$$

Calibration process

Null models are not calibrated (as they have no parameters), and their fitness is calculated in order compare them to the calibrated models. While selfish and altruistic models are non-stochastic (i.e., results are always the same no matter how many times we run the model) the random behavior model is run 100 times and averages are taken for comparison.

Calibration for the calibrated models is performed using a genetic algorithm with 10*100 individual randomized starting conditions per model run. Fitness for each model run represent an average of 100 runs with the same initial parameters. The calibration process allows for 10,000 different permutation per run. We present parameter values and the best fitness score given by f^{mlt}.We use the standard GreyBinary Chromosome genetic algorithm from *BehaviorSearch 1.0*, with population size = 50, mutation rate = 0.02, and cross-over rate = 0.7. We stop the genetic algorithm after 2,000 fitness evaluations. It is important to note here that the fitness is only calculated for new parameter combinations. We recheck our best fitness 20 times in order to avoid results driven by chance as all calibrated models have a certain degree of stochasticity. We perform 10 different searches per model.

3 RESULTS

We first begin assessing the fitness of the different models proposed. As Table 2 and Figure 2 show, more complex models (i.e., models with more parameters) do not always lead to better results. The best fitness is reached through the heuristic model (our model based on statistical analysis). All other models lead to statistically significantly lower performance (based on Wilcoxon-Mann Whitney test).

Table 2: Maximum fitness (f^{mlt}) reached by the different models (null and calibrated)

Model	***fmlt***			
	avg	**stdev**	**min**	**max**
selfish	0.013	0.000	0.013	0.013
altruistic	0.052	0.000	0.052	0.052
random	0.408	0.001	0.405	0.411
mixedrsa	0.489	0.000	0.488	0.490
pseudorandom	0.315	0.001	0.313	0.315
heuristic	0.577	0.019	0.539	0.588
utilitarian	0.549	0.007	0.532	0.557

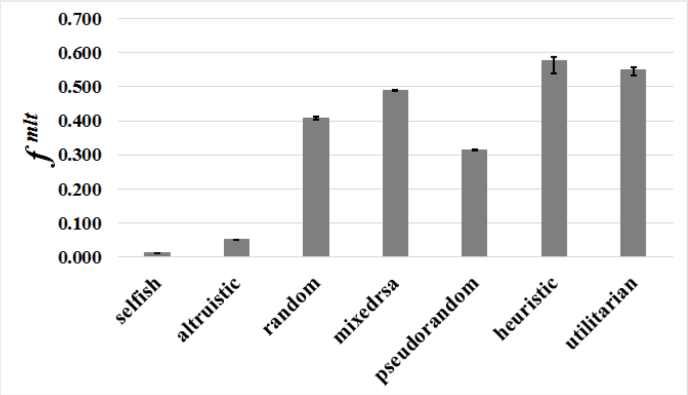

Figure 2: Average (column), minimum and maximum (bars) fitness (f^{mlt}) of different model types. Models are in order of complexity, from least complex models (selfish) to the more complex ones (utilitarian).

Table 3 shows parameter configuration as defined by the genetic algorithm. The parameter configuration represented in Table 3 are the configurations leading to max(f^{mlt}) for each model type. Mixedrsa shows that the estimated share of selfish agents is zero, the share of random agents is 75% and the share of altruistic agents is 25%. It is remarkable that this simple model performs only slightly worse compared to the best performing models. The parameters concerning the pseudorandom model seem to maximize heterogeneity regarding extraction (stdev(n_2)=1), while investment is closer to 0 than expected (i.e., baseline investment = 0 and noise added to investment equals to a normal distribution with mean 0 and standard deviation = 0.49). The heuristic model shows a parameter configuration that is different, but not dissimilar to the coefficient of the statistical model on which it is based (where tr = 0.55, w_i = 1.03 and w_e = -1.85).

Table 3: Parameter configuration leading to the best (max) fitness (i.e. fitness = max in Table 2). Only calibrated models are taken into account. The last two columns represent percentage loss in fitness when specific parameters are increased/decreased by 10%.

Model	**Parameter**	**Max(f^{mlt}) values**		**Δf^{mlt}**	
		Value	**StDev**	**Parameter +10%**	**Parameter – 10%**
mixedrsa	p_s	0.00		na	na
	p_t	0.75		-1.905	-2.248
	1- $(p_s + p_t)$	0.25		na	na
pseudorandom[a]	inv	0.00		-76.420	na
	n	0.00	0.49	-0.009	-0.005
	n_2	0.00	1.00	na	-1.991
heuristic	p_s	0.40		-0.149	0.000
	tr	0.63		-0.901	-0.842
	w_i	1.30		-3.448	-11.112
	w_e	-1.40		-5.835	-8.547
utilitarian	p_s	0.00		-5.087	na
	α	0.96		-1.493	-7.891
	β	0.57		-1.913	-3.652
	η	0.41		-0.489	-0.218
	λ	0.61		-0.326	-0.210
	τ_1	0.90		-2.368	-0.482
	τ_2	0.26		-0.071	-0.015

Notes: $\Delta f^{ml} = f^{mlt} - \max(f^{mlt}) / \max(f^{mlt}) * 100$. Δf^{ml} indicates percentage change in fitness compared to the maximum fitness as reported in Table 2. na = non available as calibration did not take parameter value into account (parameter value falls outside the interval given in the calibration process). Standard deviations are always kept between 0 and 1 for maximization purposes. a = for the pseudorandom model, mean investment in the sensitivity analysis is augmented by 1 unit (i.e., inv = 1) and sensitivity for the noise parameters is calculated by increasing/decreasing the standard deviation by 10% as the mean is fixed at 0.

The parameters of the utilitarian model require more explanation, and allow us to infer some general behavioral traits. According to the definitions put forth by Charness and Rabin (2002), we can affirm that individuals take social welfare in serious consideration being $0<\beta<\alpha<1$. Further, individuals have moderate but noticeable expectation of others cooperating (η). Finally, the parameters of the utilitarian model shown in Table 3, allow us to draw a conclusion about learning. Specifically, learning tends to be slow with regard to expected cooperation (i.e., τ_1=1 = no learning), at the same time, learning regarding the expected share of extractions by upstream users is faster ($\tau_2 < \tau_1$).

Table 2 displays the most performing models. We take the three best performing models (mixedrsa, heuristic and utilitarian) and show a detailed comparison between them and our original data. Figures 3 and 4 portray detailed results for investment per round, and investment and extraction per person. The results show that different models seem to have different strengths. For example the utilitarian model is the most accurate in predicting inequality in distribution for extraction (i.e., 0.38 for simulated and original data) and extraction levels per position (Figure 4). On the other hand, the heuristic model performs best in predicting average investment level per position (Figure 4) and inequality in investment levels (0.30 for simulated and original data). Accurate prediction of investment per round are very difficult to infer with the model proposed and analyzed (Figure 3).

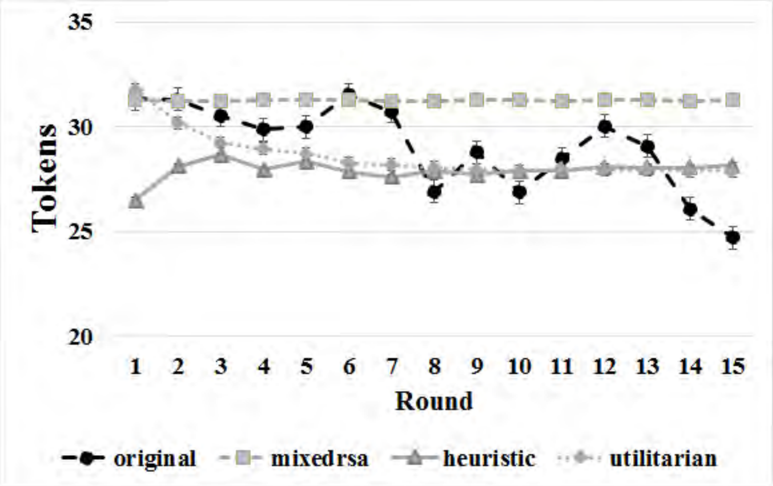

Figure 3: Investment per round for the simulated and original data

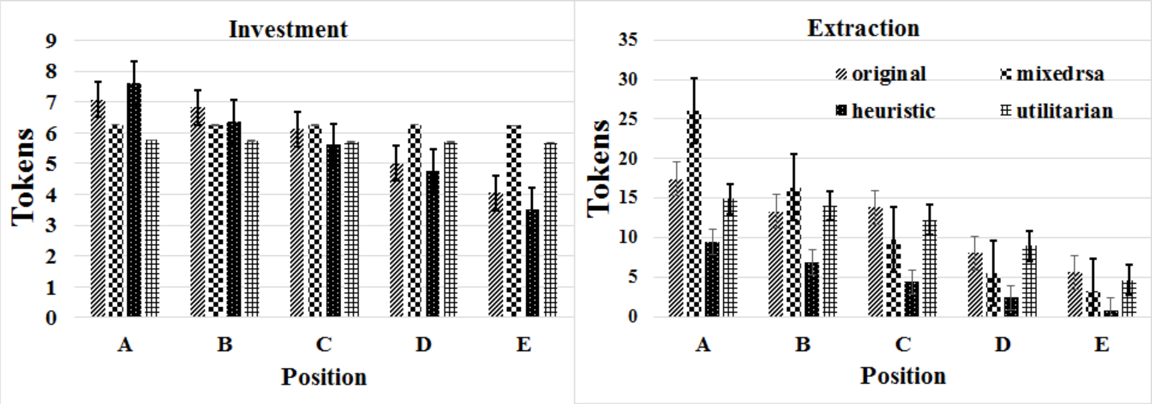

Figure 4: Average investment and extraction per position (original and simulated data)

3.1 Sensitivity analysis

Sensitivity analysis is performed only for calibrated models (as null models do not have parameters). We increase/decrease the parameters' values by 10% and rerun the model in order to assess sensitivity of the fitness value on the parameters of interest. We are not only interested in the best performing model, but also on the most robust one. Results of fitness sensitivity are reported in Table 3.

Our analysis show robust estimates for fitness to moderate changes in parameter values. The largest fitness decreases occur when we increase the mean of investment for the pseudorandom model. However, we do not take into consideration changes of multiple parameters. Maximum decrease in fitness is higher if we change multiple parameter at the time (i.e., -2.235% for mixedrsa, -78.032% for pseudorandom,-19.222% for the heuristic, and -19.748% for the utilitarian model). From our sensitivity analysis we can conclude that more parsimonious models (i.e., models with less parameters) are more robust to changes in parameters.

4 CONCLUSIONS

Lately there is an increased interest in empirically calibrating ABM. Here we try an approach that is based on theory and relies less on actual data mining. We avoid the use of machine learning algorithms (as in Wunder et al 2013) because to advance social science, the ABMs need to have been theoretically grounded in order to provide insights and theory development that go beyond a specific dataset. We tested various models that make assumptions according to specific behavioral theories (i.e., selfish agents, altruistic agents, mixed agents, and agents that maximize utility having other-regarding preferences), we compare different theoretical models to a high quality dataset stemming from experiments we have performed. Further, we compare these theoretical models with a null random and a pseudorandom model and with a model that is based specifically upon the dataset (heuristic).

Although models that are tailored to a specific dataset (and thus are built upon data-mining) perform better, the most complex behavioral models do not lag far behind in terms of performance. Further, models that assume agents who behave altruistically and randomly reproduce data quite well, leaving open questions on how individuals behave during controlled experiments in a laboratory setting.

Finally, based on the parameters presented in Table 3 we can draw tentative conclusions on the behavior and characteristics of individuals. Individuals seem to be socially concerned (i.e., concerned with social welfare), do expect an intermediate level of cooperation, and have a hard time in learning expected cooperation but are faster learner when it comes to predicting the behavior of others based on past instances.

ACKNOWLEDGEMENTS

We would like to thank Irene Perez and Nathan Rollins for running the experiments. We acknowledge financial support for this work from the National Science Foundation, grant numbers SES-0748632 and BCS-1115054.

REFERENCES

Arifovic, J. and J. Ledyard. 2012. "Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism." *Journal of Public Economics* 96: 808–823.

Camerer, C. F. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton, New Jersey: Princeton University Press

Charness, G., and M. Rabin. 2002. "Understanding social preferences with simple tests." *The Quarterly Journal of Economics* 117:817-869.

Deadman, P. J. 1999. “Modelling individual behavior and group performance in an intelligent agent-based simulation of the tragedy of the commons.” *Journal of Environmental Management* 56: 159-172.

Janssen, M. A., F. Bousquet, J. C. Cardenas, D. Castillo, K. Worrapimphong. 2012. “Field experiments of irrigation dilemmas.” *Agricultural Systems* 109:65-75.

Janssen, M. A. and E. Ostrom 2006. “Empirically based agent-based models.” *Ecology and Society* 11: 37.

Janssen, M.A. In press. “An agent-based model based on field experiments.” In *Empirical agent-based-modelling*. A. Smaigl and C. Barreteau, eds. New York: Springer.

Ostrom, E., R. Garder, and J. Walker 1994. *Rules, games, and common-pool resource*. Ann-Arbor: University of Michigan Press.

Poteete, A.M., M. A. Janssen, and E. Ostrom, 2010. *Working Together: Collective Action, the Commons and Multiple Methods in Practice.* Princeton, New Jersey: Princeton University Press.

Rollins, N., J. A. Baggio, I. Perez and M. A. Janssen in preparation. “Lab experiments on irrigation games under uncertainty.”

Wunder, M., Suri, S., and Watts, D. J. 2013. “Agent based models of cooperation in public goods games.” *forthcoming*

AUTHOR BIOGRAPHIES

JACOPO A. BAGGIO is currently a postdoctoral research associate at the Center for the Study of Institutional Diversity, School of Human Evolution and Social Change, Arizona State University, with a Master and a PhD in International Development from the University of East Anglia. His current research touches upon two main areas: 1) the analysis of social-ecological systems from a structural perspective (i.e., the study of social-ecological networks) and 2) understanding decision making, cooperation and uncertainty in complex ecological systems.

MARCO JANSSEN is trained as an applied mathematician with a Master in Operations Research from Erasmus University Rotterdam, and a Ph.D. in Mathematics from Maastricht University. Currently he is an associate professor of modeling social and social-ecological systems in the School of Human Evolution and Social Change, and directs the Center for the Study of Institutional Diversity, both at Arizona State University. As a transdisciplinary scholar he studies collective action and the commons using experimental and computational methods.