

Pattern-oriented modeling of commons dilemma experiments

Running title: Pattern-oriented modeling of human experiments

Marco A. Janssen^{1,2}, Nicholas P. Radtke^{2,3}, Allen Lee^{1,2}

¹ School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287

² Center for the Study of Institutional Diversity, Arizona State University, Tempe, AZ 85287

³ School of Computing and Informatics, Arizona State University, Tempe, AZ 85287

Correspondence to: Marco A. Janssen, School of Human Evolution and Social Change,
Arizona State University, PO Box 872402

Tempe AZ 85287-2402, USA

Phone: 480 965 4193

Fax: 480 965 7671

Email: Marco.Janssen@asu.edu

Abstract. A major challenge in the development of computational models of collective behavior is the empirical validation. Experimental data from a spatially explicit dynamic commons dilemma experiment is used to empirically ground an agent-based model. Four distinct patterns are identified in the data. Two naïve models, random walk and greedy agents, do not produce data that match the patterns. A more comprehensive model is presented that explains how participants make movement and harvest decisions. Using pattern-oriented modeling the parameter space is explored to identify the parameter combinations that meet the four identified patterns. Less than 0.1% of the parameter combinations meet all the patterns. These parameter settings were used to successfully predict the patterns of a new set of experiments.

Keywords: empirically grounded agent-based modeling, commons dilemma, individual decision making, human experiments

1 Introduction

One of the main challenges for the use of agent-based modeling as a scientific method is to create empirically grounded models. Due to the stochastic and non-linear nature of most agent-based models, empirically testing models is a non-trivial task. The challenge of improving our methods for empirical grounding of agent-based models has been addressed in a number of recent papers (Grimm et al., 2005; Janssen and Ostrom, 2006; Robinson et al., 2007). Which approaches are appropriate depend on the specific research questions, available data and the dynamics of the candidate agent-based models.

In the ideal case we would have individual level data across an appropriate interval of time and with the appropriate granularity. However, it is still not clear how we might rigorously test an agent-based model on even such an ideal data set. Simply creating a model that mimics the aggregate behavior is certainly insufficient. Ideally, we should be able to demonstrate the following:

- the proposed model is able to capture important features of the empirical system at different spatial and temporal scales,
- the conceptualization of the model is based on accepted domain knowledge from the relevant behavioral and social sciences,
- simpler models than the one developed are unable to capture the essential features of the empirical system,
- the proposed model can predict important features of an out of sample set of observations. We are using prediction in the sense of distributions of system behavior, not point predictions.

In this paper we use experimental data from group experiments where we placed participants within a commons dilemma. Within the experiment participants make many decisions in a real-time dynamic environment representing a virtual common resource. We have a large amount of highly detailed data that captures each and every in-game decision made by our participants. This presents an interesting case study for exploring the challenge of developing empirically grounded agent-based models.

A typical approach to compare models in behavioral game theory is to estimate parameters of statistical models using maximum likelihood estimation (Camerer, 2003). Models which have a better fit with the data are considered superior to alternative models. Fitting the model, however, is not an adequate approach for evaluating model performance (Pitt and Myung 2002). The main problem is that more complicated models have more degrees of freedom to fit the data. The trade-off is between the fit of the data and the complexity of the model. There are different penalty functions of complexity that can be used to balance the fitness and the complexity of the model. Such penalty functions depend on the number of parameters of the model, and the number of observations used.

Another problem is the multiplicity of empirical patterns at different scales that one may like to fit with the model. The estimation of parameters to fit the model to different patterns requires weighting of the different goals. This still leaves the question of how to decide which pattern is more important to fit well than others though.

A maximum likelihood can be applied to the experimental data of this paper, if optimization routines can provide good solutions in a limited amount of time using the agent-based model without known stochastic properties. Unfortunately, this is not the case. The stochasticity leads to a noisy fitness landscape in which search routines get stuck. Since we don't know the mathematical properties of the fitness landscape, we cannot ascertain whether we are stuck at a local optimum.

If the model of a particular structure is known to be correct, measurement errors are the main source of the imperfectness of the fit between a statistical model and the data. For many behavioral models the actual structural models are not known, and therefore different acceptable representations can exist. This relates to the concept of equifinality where one take into account the possibility of many acceptable representations that cannot be easily rejected (Beven, 2005; von Bertalanffy, 1968). Some models are better than others, but there is not one optimal model. Therefore, we address entertain the questions whether it is possible to find good enough fits between the data and the simulation model? We find that pattern-oriented modeling is an approach from landscape ecology which seems suitable to meet this purpose, and that's why we use it here for testing agent-based models of human behavior (Grimm et al., 2005).

In the next section we will discuss the pattern oriented modeling methodology. We then present the experiments we performed and the patterns from the experimental data. The observed patterns are then compared with some naïve models. In the following section we discuss the agent-based model developed to mimic the observed patterns. We then calculate for a large set of parameter combinations whether the model fits observed patterns and define subsets of acceptable models. Finally, we use the derived models to make predictions for new types of experiments and compare those predictions with new observations.

2 Pattern-Oriented Modeling

Pattern-oriented modeling originates from ecology. A 'pattern' is defined as a characteristic, clearly identifiable structure in the data from the system of interest. The pattern goes beyond random variation and therefore indicates that there is an underlying process that generates this pattern. Such patterns can manifest themselves in spatial and temporal contexts..

Pattern-oriented modeling has been used to test various ecological models, especially related to population dynamics and conservation. Instead of focusing on the best match between data and the model, the approach explores the parameter space and only accepts parameterizations that simultaneously produce simulation results that are close to all the patterns from the data. By focusing on multiple patterns one can look at patterns at different hierarchical levels that simultaneously need to be met to accept the parameterizations. The larger the number of different patterns that are matched at different scales, the less likely it is that the model is structurally wrong (Grimm et al., 2005; Wiegand et al., 2003).

To define when a simulated pattern is close to the data, one defines uncertainty boundaries. Since we use data from experiments for which we have multiple observations for each treatment we can define the boundaries statistically.

The pattern oriented modeling process is iterative. One may find many parameterizations within all the uncertainty boundaries. If that is the case, one may decide to have another look at the data to find more distinct patterns that are harder to produce and be more selective on the set of acceptable models. One may also find that none of the model parameterizations are able to meet all the patterns and provide information to improve the model. Hence, pattern oriented modeling is an iterative process that leads to a set of model parameterizations that are structurally likely to be correct and produce model outcomes that are similar to the empirical observations. We will use the highly detailed data from laboratory experiments to develop a model that meet a set of patterns.

Both statistical estimation of models and pattern oriented modeling explore the parameter space to compare the model outcomes with the data. Whereas statistical estimation is focused on finding the best fit, pattern oriented modeling defines, somewhat subjectively, what is considered to be a good enough fit, identifying the parameter settings that are sufficient. It is possible that quite different parameter settings lead to the same patterns given

the uncertainty ranges, but only one of these parameter settings will be identified when we use statistical estimation. When one uses the model for predicting the behavior of the system in different conditions it would be desirable to include all the model versions that captured the patterns well enough. Finally, using multi-objective optimization, statistical estimation is possible when one is interested in capturing patterns at different scales, but one has to identify how to weight these different objectives. In pattern oriented modeling one defines which uncertainty ranges of the patterns are acceptable, and all patterns need to be met for the parameter setting to be accepted.

Grimm et al. (2005) present pattern oriented modeling as a general strategy to design, test and analyze agent-based models. So far the method has been applied to ecological models, especially landscape ecology. In this paper we will apply this methodology to a different domain, group experiments of human behavior.

3 Spatially explicit commons experiments

3.1 Commons Dilemmas

Many natural resource problems can be classified as commons dilemmas, a dilemma between the interest of the individual and the interest of the group as a whole (Ostrom et al., 1994). A common-pool resource (CPR) is such a dilemma, where a resource is shared by multiple users. CPRs are characterized by the fact that it is difficult to exclude users, and units appropriated by one user are not available anymore for other users. Examples of CPRs include forests, pastures, irrigation systems, and fishing grounds.

When individual interests dominate, a conflict on common resources may lead to the tragedy of the commons in open access situations (Hardin, 1968). However, many empirical studies have shown that people are able to govern common resources effectively (Ostrom et al., 1994). A typical method to study the fundamental processes on how individuals are able to self-govern themselves in such commons dilemmas is via controlled laboratory experiments. In those experiments participants interact with an abstract resource during a number of rounds and derive monetary incentives.

The experimental data used in this paper is part of a series of experiments performed to study how groups develop new institutional arrangements if they share common dynamic resources. We are interested whether different resource dynamics (different regrowth rates in this paper) lead to different type of institutional arrangements. For the modeling exercise described in this paper we only use data from the rounds of the experiments before the participants are allowed to communicate. For more specific information on the experimental design we refer to Janssen et al. (2008) and Janssen and Ostrom (2008).

3.2 The experiments

Participants were recruited from a large database of undergraduate students at Arizona State University during the spring semester of 2007. The average age was 21.4 years and 67% of the participants were male. Data used for this paper is a subset of the actual data, using only round one and two of the experiment. In the experiment, the participants go through a practice round on an individual plot and then go through a real round on an individual plot. In the second round, participants are placed in a common pool resource setting with three other participants, randomly selected from the full group of participants. After the second round, participants are allowed to use text chat to coordinate their actions. In this paper we only use data from round 2¹, before the participants start communicating. Data from 64 individuals in 16 groups is used.

In the individual round, participants collect green tokens on a 14 x 14 spatial grid of cells. In the group experiment, four randomly assigned participants share a renewable

resource that grows on a 28 x 28 spatial grid of cells. Participants implicitly harvest a green token by moving their virtual avatar's location on top of the token by pressing the arrow keys (left, right, up, and down). There are two modes, implicit and explicit, that can be toggled by pressing the 'M' key. In the implicit mode one collects a token automatically when one's avatar is on a cell with a token. In explicit mode one can move around without automatically harvesting tokens. When one wishes to harvest a token, the participant must press the spacebar when their avatar is on a cell with a token.

The resource renewal rate is density dependent. The probability that a green token will appear on an empty cell increases as the number of adjacent cells with tokens increases (Figure 1). The probability p_t is linearly related to the number of neighboring cells with tokens: $p_t = p \cdot n_t / N$ where n_t is the number of neighboring cells containing a green token, and N the number of neighboring cells ($N = 8$ because we use a Moore neighborhood). The parameter p is defined in such a way that the renewal of the resource is quick enough to be observed by the participants, but sufficiently slow that the participants experience a dilemma between immediate, individual benefits and longer-term, group benefits. If participants collect tokens as quickly as they can, there will soon be no tokens remaining on the screen. Once every token has been harvested, no further opportunity exists for any new tokens to be created.

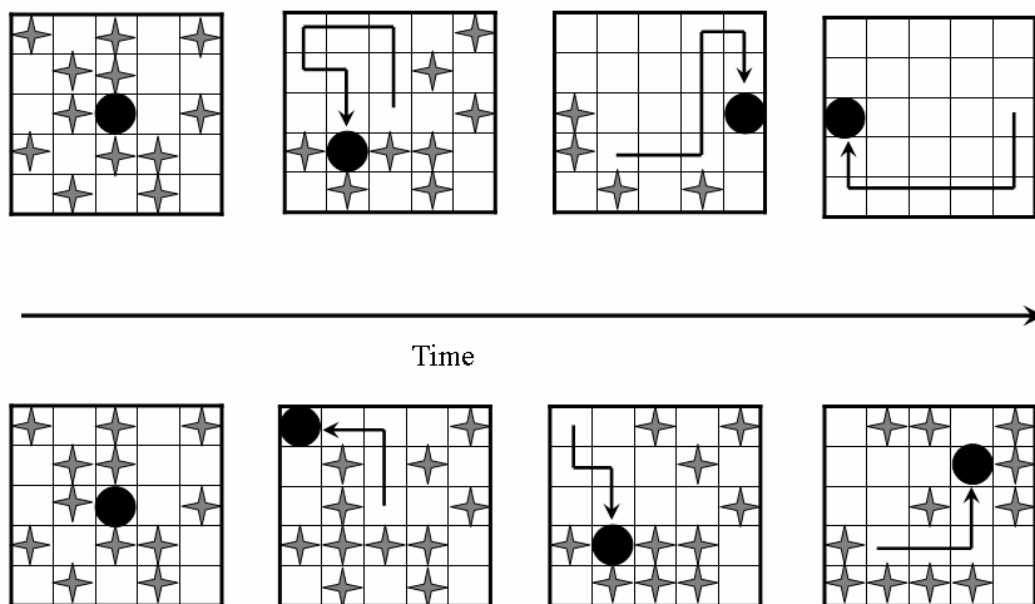


Figure 1: Four snapshots of two harvesting strategies by two different types of subjects in a hypothetical situation of a 5x5 resource, where resource units are depicted by star-shape objects. On the top row in the figure above, the subject moves their avatar (circle) eight steps per time period. On the bottom row, the subject moves their avatar only four steps per time step.

The participants have four minutes to collect tokens and each token is worth \$0.02. Two treatments are considered: A low growth case with p equal to 0.01 and 25% of the cells

initially populated with tokens, and a high growth case with p equal to 0.02 and 50% of the cells initially populated with tokens. We have 6 groups for the low growth treatment and 10 groups for the high growth treatment.

4 Patterns

4.1 Detecting Patterns

Four patterns are distinguished to describe the data. The choice of these patterns is guided by having different scales (group versus individual, average over a round versus time series) and after a thorough analysis of the data for earlier publications. More data analysis may lead to additional patterns, and therefore the choice of patterns is somewhat subjective. However, if the patterns are not distinctive for our data set, we will find many model versions explaining the patterns and such an exercise would not be productive. Since we use experimental data of multiple groups we use for each pattern the standard deviation to define the acceptable range.

4.2 Pattern 1: Resource size over time

The first pattern is the number of tokens on the screen over time. Figure 2 shows the mean as well as the range (\pm standard deviation) for both the low and high growth conditions, for round and 2. For simplicity's sake we assume that a pattern within the range of average \pm the standard deviation is an acceptable reproduction of a pattern.

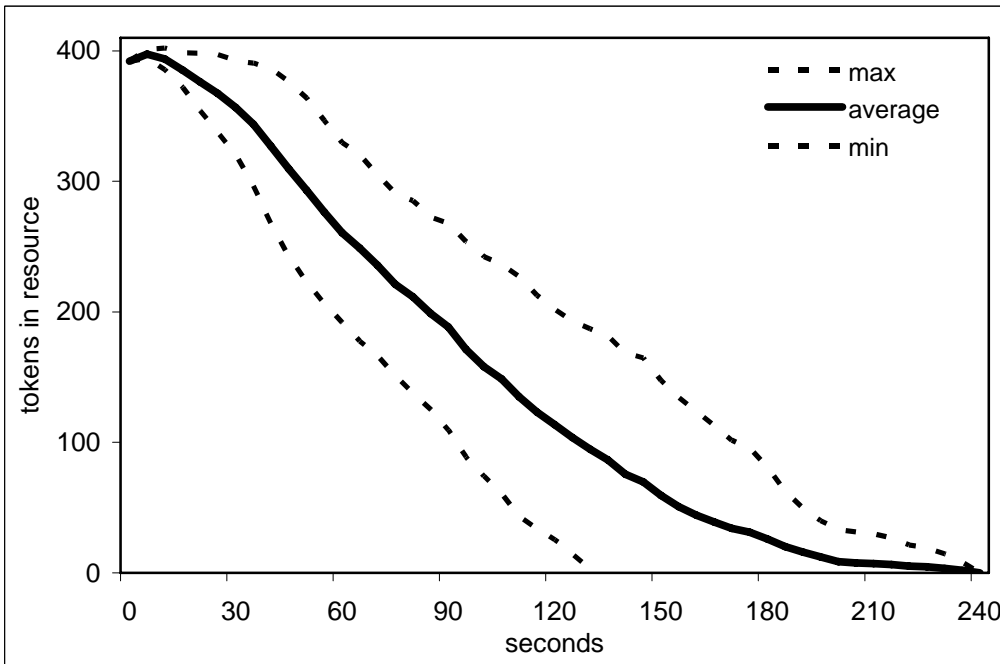
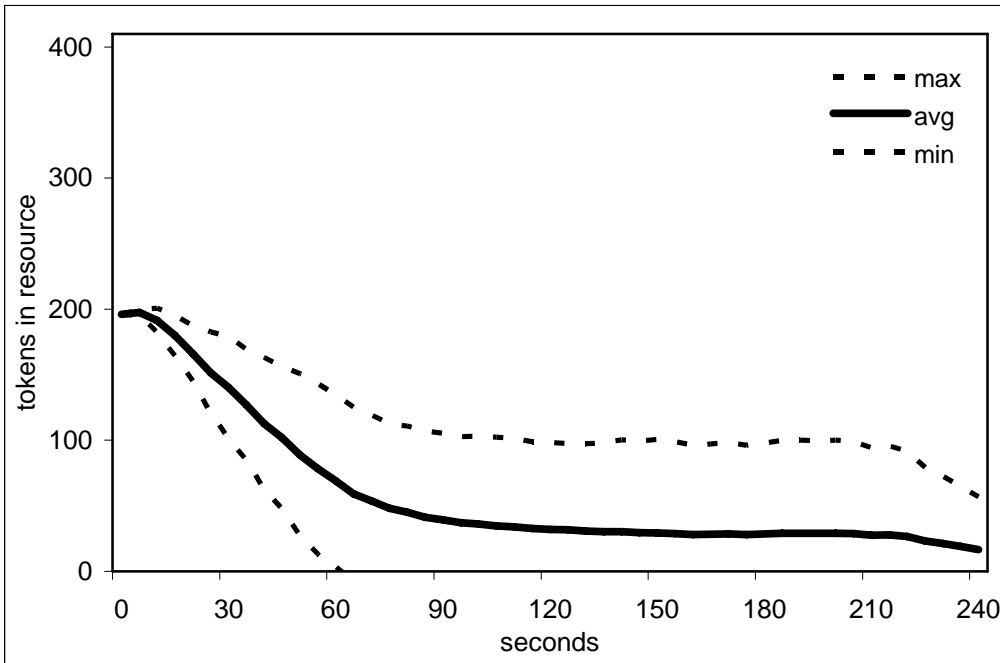


Figure 2. Average number of tokens on the screen every five seconds for round 2. Top is the low growth case and bottom is the high growth case. The max and min refer to the average +/- one standard deviation.

We found it difficult to meet this pattern with our agent-based model and we found that in the actual experiments there is a delay between the start of the experiment and the time when participants start making their decisions that can be attributed to both network and cognitive latency (Figure 3). This led to some extra growth of the resource in the first few seconds. Therefore we will ignore the first 10 seconds in our analysis of the first pattern.

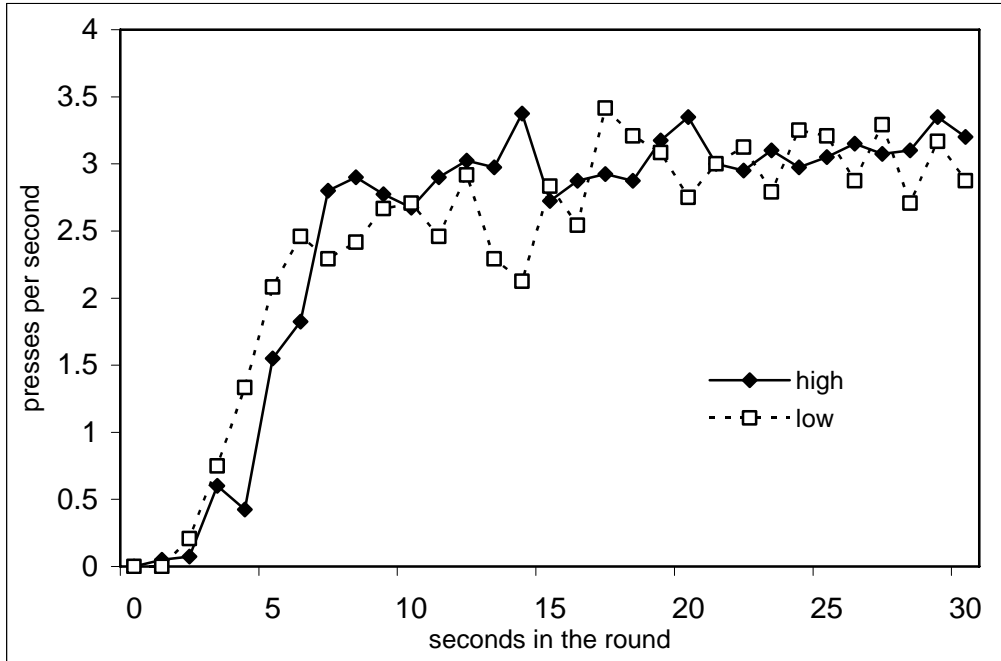


Figure 3. The number of key presses per second over time in round 2.

4.3 Pattern 2: Gini-coefficient

The second pattern is the relative inequality of tokens collected within a round. We calculated the share of tokens collected by participant i compared to the group total. Then we calculated the Gini coefficient for each group (Table 1). A Gini coefficient is a measure of inequality². If all participants collect equal numbers within their group, the Gini coefficient would be 0. If one participant takes everything, the Gini coefficient would be 1. We find that the Gini coefficient for the low growth rate experiments is higher on average compared to the high growth rate experiments: 0.23 versus 0.11. This means that there is more inequality when resource regrowth is low.

Table 1: Average Gini coefficients and standard deviations of groups, as well as the lower and upper boundaries used for the pattern oriented modeling (round 2).

	Mean	Stdev	Lower value	Upper value
Low growth	0.2323	0.1294	0.1029	0.3617
High growth	0.1124	0.0563	0.0561	0.1787

4.4. Pattern 3: Total number of tokens collected

The average number of tokens collected by the groups during the four minutes of the experiment is obviously lower for the low growth rate treatment compared to the high growth rate treatment. This information is used as the third pattern (Table 2).

Table 2: Average number of tokens collected per group, as well as the lower and upper boundaries used for the pattern oriented modeling (round 2).

	Mean	Stdev	Lower value	Upper value
Low growth	274.0	63.75	210.25	337.75
High growth	743.1	114.1	629.0	857.2

4.5 Pattern 4: Direction of movement

The fourth pattern originates from our observation that participants have a tendency to continue in the direction they are already moving instead of changing direction (Figure 4). There is a slight cost of changing direction (hitting a different key on the keyboard). Since we record every in-game action of every participant we can analyze how often a participant moves in the same direction until (s)he changes direction. In both treatments we observe that participants move in straight lines more often than they would if they were simply making random directional choices. In our pattern oriented analysis we only consider the number of moves in the same direction up to ten straight moves, so any additional moves past ten straight moves are disregarded.

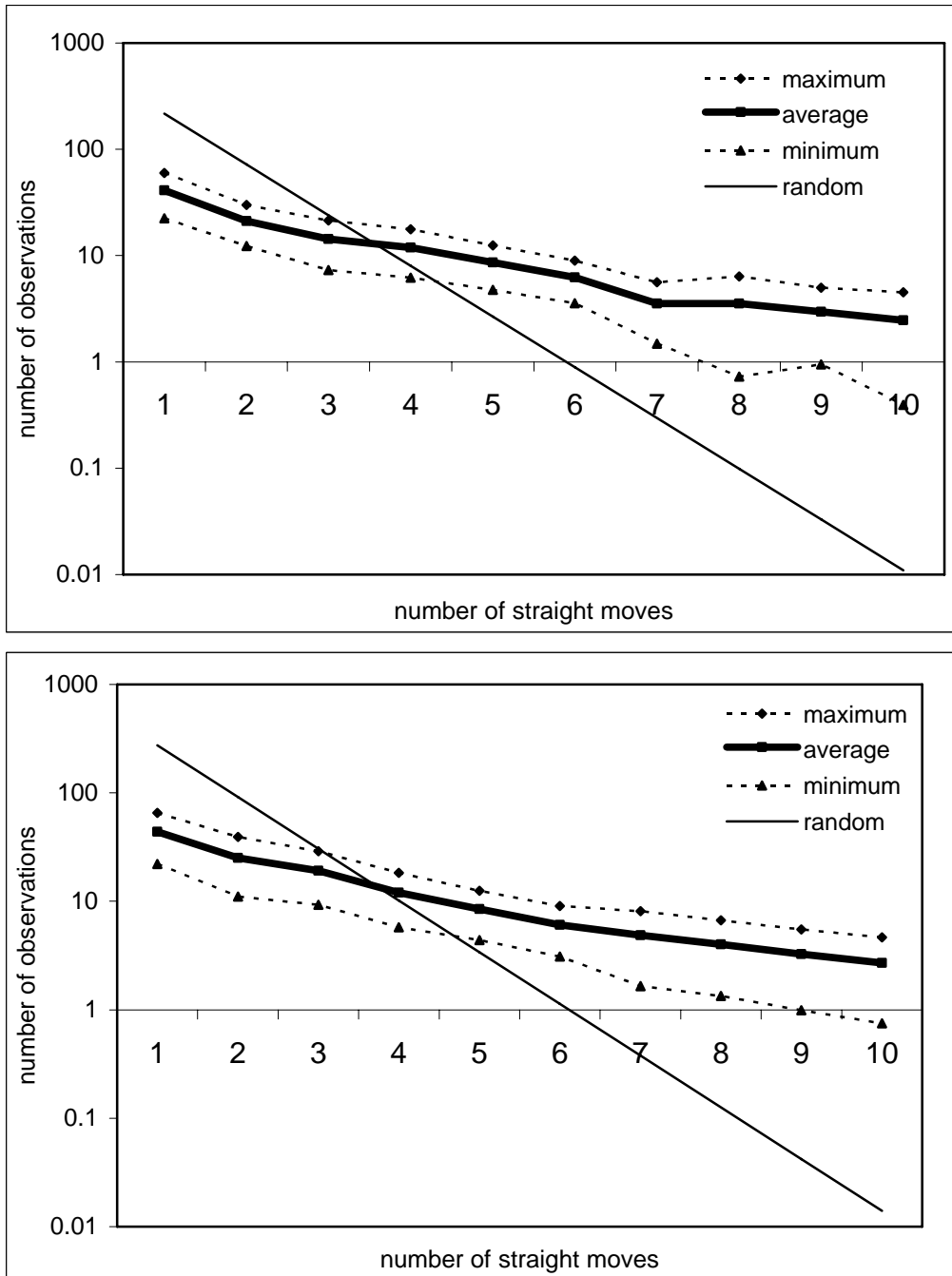


Figure 4. Round 2 data for low growth rate (top) and high growth rate (bottom) for number of straight moves. As a reference we include the relationship when participants make random moves.

4.6 Parameters: Speed and mode

Besides these four patterns we use to test our model, we extracted two patterns from the data that we will use as inputs: the use of explicit mode and speed of movement.

First, we determine the percentage of participants that use the explicit mode. With the higher growth rate fewer participants use the explicit mode. In the model we assume an exogenous fraction of agents use the explicit mode during the round. More in depth analysis shows that participants rarely change their mode during the round. We will use 30% explicit mode for the low growth treatment, and 15% for the high growth treatment.

We did not find a significant relation between speed, the number of key presses per second, and the proximity of other avatars nor between speed and the size of the resource. Therefore we assume that speed is an individual attribute independent of other avatars. To determine the base speed of the participants, we calculated the average speed per second between 10 and 70 seconds. We ignored the first 10 seconds, since participants needed to get up to speed when the round started, and after 70 seconds some resources started to be depleted. There is a substantial difference in the speed distributions for the two treatments. Figure 5 shows that the average speed in the low growth case is lower than the high growth case, especially since a few individuals do not move around to let the resource replenish. The speed distribution of the high regrowth case is similar to the group experiments.

Given the mean number of key presses per second, we assume a normal distribution of speed variation per second (Figure 5). The data shows a higher standard deviation when the number of key presses per second is higher: $\text{stdev} = 0.598 * \text{Mean} - 0.062 * \text{Mean}^2$.

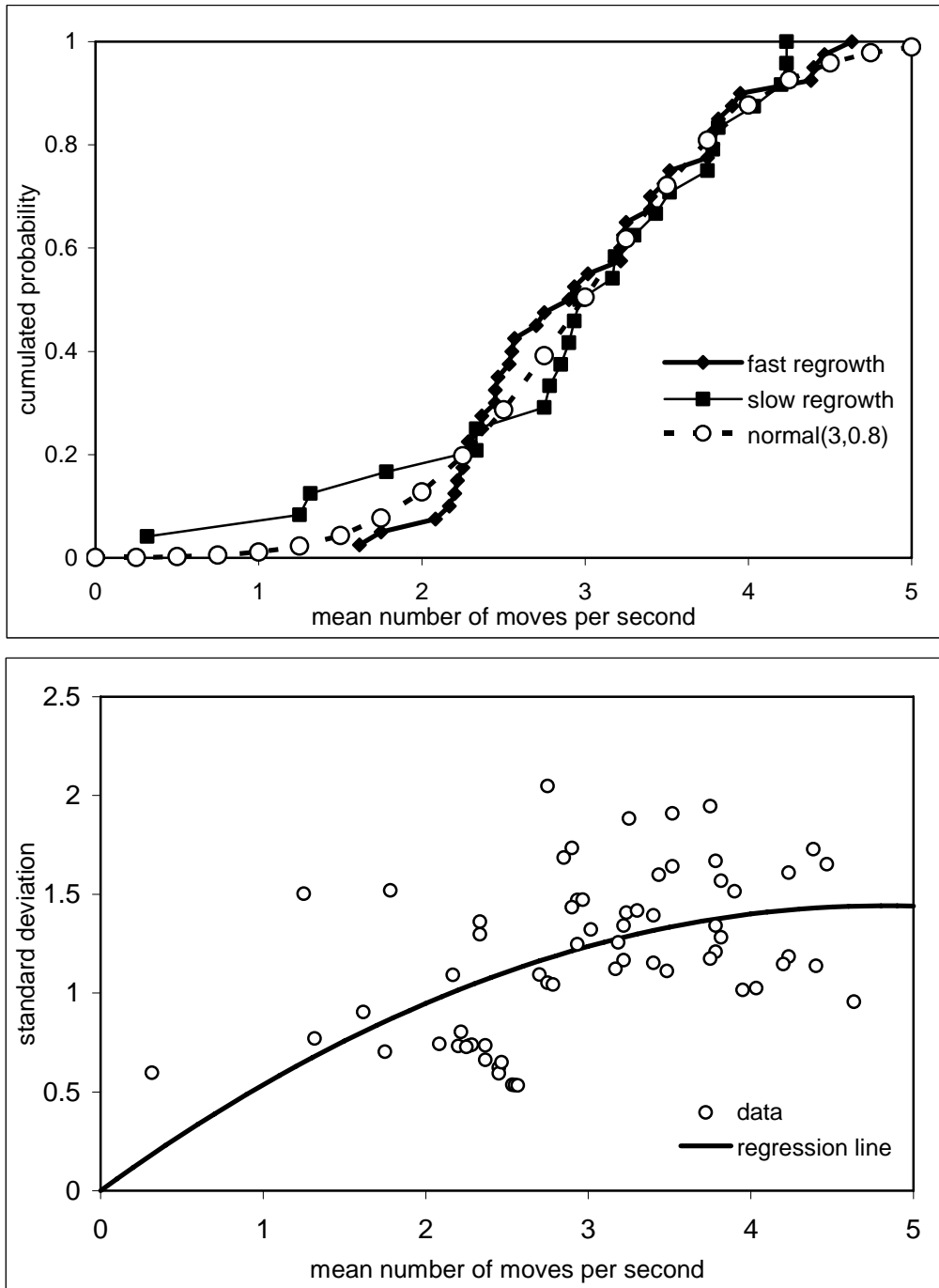


Figure 5. Distribution of average key presses per second in Round 2(top). Relation between standard deviation and the number of key presses per second (bottom).

5 Naïve models

We will present now two simple, naïve models and show that they are not able to capture the important empirical patterns. All models discussed in this paper simulate the token regeneration as has been implemented in the experimental software using time steps of 1 second. For high growth rate cases p is equal to 0.02 and for low growth rate cases p is equal to 0.01. Agents cannot pass through the borders. In the initialization, we allocate 392 (50%) or 196 (25%) tokens randomly on the unoccupied cells, for high and low growth respectively.

For the simple models, we use the speed distributions as derived from the data and use some simple assumptions of how the agents move. If naïve models can explain the observed patterns well, there is no need for a more complicated model. The naïve models we use are:

- Random walk: agents are in implicit mode and randomly move around on the resource.
- Greedy agents: agents always move towards one of the nearest tokens.

We see in the Figure 6 that the random walk agents lead to a much high number of tokens in the resource, which means a much lower number of tokens collected. On the other hand, greedy agents rapidly reduce the resource size, much faster than within the range of observed patterns.

The naïve models show that the participants in the experiment did not pursue each opportunity to collect a token like with the greedy agents, but were more directed in their movement compared with a random walk. Our comprehensive model will therefore focus on how agents make decisions which token to pursue.

The naïve models cannot explain the other patterns either. The random agents collect too few tokens (73 for low growth case and 173 for high growth case). The greedy agent is within the observed pattern of the low growth case (223), but too low for the high growth case (513). The agents change direction more frequently for both the random and the greedy agents compared to observed distributions. A similar analysis is performed for data from the first round of our experiments and demonstrates the same inability to explain the observed patterns by the naïve models.

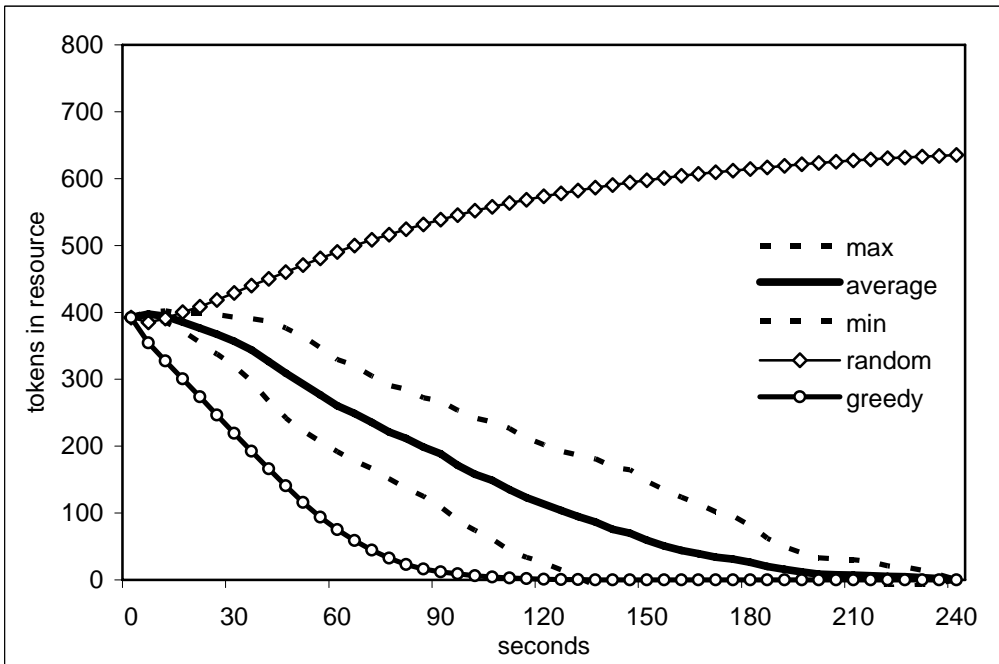
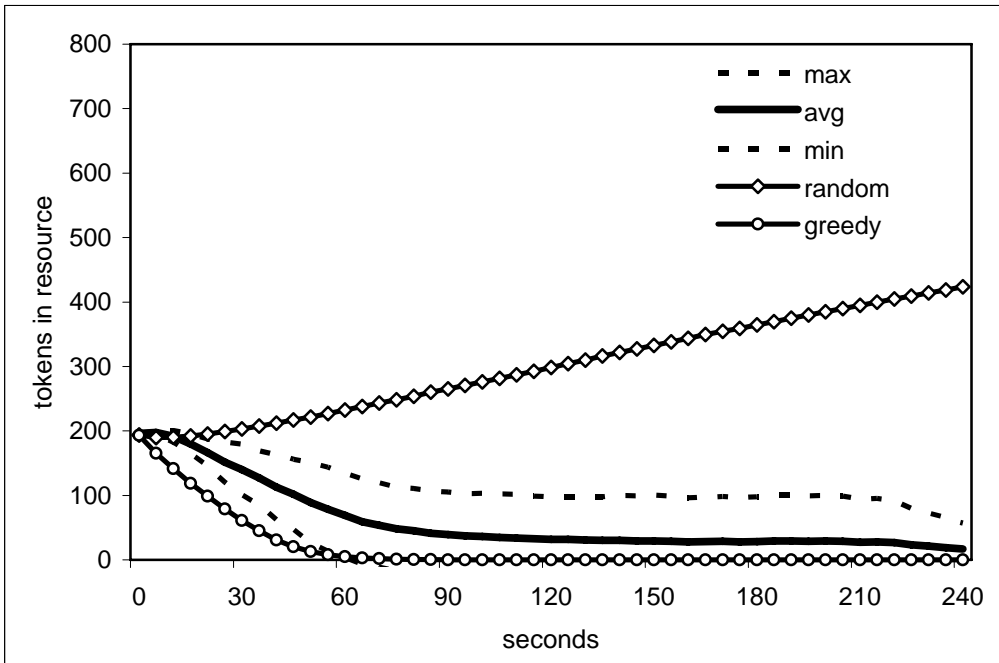


Figure 6. Number of tokens in the resource for naïve models (random walk & greedy) in relation with the empirical patterns. On the top the low growth case and on the bottom the high growth case.

6 Model description

We will now discuss the more comprehensive model. The speed of the agents is defined by the empirical distributions. For every move the agent must decide where to go to. We develop our model in line with the model of Roberts and Goldstone (2006) who developed an empirical model for a similar experimental environment. The basic idea is that each agent defines the value of each token on the screen as a target to move to. This value is established via three components.

- the closer a token is to the agent, the more valuable the token
- the more competing agents that are close to a token, the less valuable the token
- tokens that are straight ahead in the current directional path of the agent are more valuable.

As such we formulate the value of a token at location (i,j) as follows:

$$V_{i,j} = P_1 \cdot \left(\frac{1}{\text{tokendistance}} \right) + P_2 \cdot I - P_3 \cdot \text{agentdensity}$$

where agent density is the number of agents in radius R_{P2} around the potential target. I is an indicator function and is 1 when the location of the token is in the straight path of the agent, and 0 otherwise. Furthermore the sum of the weights P_i is equal to 1.

One of the tokens will be drawn based on the relative value among all the tokens. When the agent reaches this position, it will select a new target. Agents also update their target once per second, or less if their speed is slower than one move per second. An agent in explicit mode does not consider a token in its same cell to be part of the eligible set of tokens. Using probabilistic choice, the probability of having a token at location (i,j) as the target is defined as

$$P[T(i, j)] = \frac{e^{\beta V_{i,j}}}{\sum_{k,l} e^{\beta V_{k,l}}}$$

where β is the parameter that defines how sensitive agents are to differences in the value of the tokens.

Based on the chosen target, the agent defines the direction of the target and decides to go up, down, left, or right. When a move is made, and the agent is on a cell with a token, the agent automatically collects the token when in implicit mode. However, when in the explicit mode, the agent needs to decide whether or not to collect the token. We assume that agents are more tempted to take the token when more tokens are around the cell, and the probability to collect the token is defined as

$$P[\text{collect}] = \frac{x^b}{x^b + a^b}$$

where x is the number of tokens in the eight cells of the Moore neighborhood divided by 8. The parameters a and b define at what density the probability is 50% ($x=a$) and how steeply the probability increases with higher values of x .

Next to a few parameters we directly relate to the observations (speed and mode) we have seven parameters that we use for the calibration: P_1 , P_2 , P_3 , R_{P3} , β , a and b .

7 Models that meet patterns

Pattern-oriented modeling argues that acceptable models are those that are able to match all observed patterns. Given the stochastic nature of the models, matching observed patterns is defined as generating average statistics that are within defined ranges for the patterns. Since we performed a number of experiments, averages and standard deviations of the patterns can be determined as described above. We simply assume that the acceptable level of generated averages of patterns lies between the average +/- of one standard deviation.

Our model is implemented in MASON (Luke et al., 2005) and was run one hundred times for each parameter combination³. Seven parameters were varied for the round 2 version of the model. The weights of the values of the tokens were varied in 66 different ways by varying P_1 , P_2 and $P_3 = 1 - P_1 - P_2$ systematically with steps of 0.1. Furthermore, we varied β , b and radius for 5 values and a for 4 values. This leads to 33,000 parameter combinations and 6.6 million model runs (Table 3).

Table 3: Parameter values used in exploring the parameter space for matching the four patterns.

Parameter	Description	Parameter values
P_1	Weight of tokens nearby	66 combinations
P_2	Weight of clicking straight direction	
P_3	Weight of other agents nearby token	
β	Agent sensitivity to differences in values for tokens	3; 6; 9; 12; 15
a	[explicit] parameter affecting threshold	1; 3; 5; 7;
b	[explicit] Steepness of curve	3; 6; 9; 12; 15
Radius	Radius to determine the number of agents nearby token	3; 6; 9; 12; 15

Table 4 shows the results of the parameter explorations. Patterns 2 and 3 are not discriminating. However, pattern 4, the straight line movements of the avatars is met by only about 10% of the parameter combinations (Table 4). For the low growth case 2103 parameter combinations meet all four patterns, while only 116 parameter combinations meet all four patterns for the high growth case.

15 parameterizations meet all patterns in both data sets. In those cases P_1 is 0.4 or 0.5, P_2 is 0.3, P_3 is 0.2 or 0.3, the radius is 9, 12 or 15, β is 12 or 15, and a and b vary among all possible combinations. This indicates that agents tend to select targets nearby and avoid having more than one other agent in the neighborhood. The agents put less weight on whether the tokens are in a straight direction. Since explicit mode is used by only a small fraction of the agents, the parameters a and b are not found to be discriminating. When 2 or more agents are in the radius of 10 around a token, this token is not appealing as a target, except when the token is one or two cells from the avatar. Given the high value of β the agents discriminate substantially between tokens with different scores.

Table 4: The number of parameter combinations that meet the patterns for the low and high growth conditions. Between parentheses the percentage of the total number of parameter combinations.

	Pattern 1 Tokens left	Pattern 2 inequality	Pattern 3 Tokens collected	Pattern 4 Direction	All patterns
Low growth	8635 (26.2%)	25573 (77.5%)	32313 (97.9%)	4614 (14.0%)	2103 (6.4%)
High growth	697 (2.1%)	17570 (53.2%)	144988 (45.4%)	4773 (14.5%)	116 (0.4%)

8. Predictions of new experiments

How well is the suite of model versions able to predict results of new experiments? Colleagues at Indiana University performed experiments with the same software in the Spring and Fall of 2008 with a different treatment¹. It must be noted that these experiments were not designed to test our particular model, but performed to test hypothesis that will be reported in a different study. We will use here only the first open access round of the experiment. The Indiana University experiment was different in the following ways:

- The participants in this experiment did not have an individual round after the practice round.
- The group consisted of 5 people instead of 4 people.
- The participants could only use the explicit mode.
- The size of the resource was 29 x 29 cells.
- The regrowth probability was $p=0.01$, the low growth condition
- The initial number of tokens on the screen was 210 (25%)

To predict the results for the Indiana University experiments, we will change the number of agents from 4 to 5, assume a 100% use of the explicit mode, change the size of the resource to 29x29 cells, and use the low growth case.

Before we present the results of the prediction in relation to the data, we first present how well we would predict the results of the low growth experiments of the ASU data set, which we used to identify the parameter settings. For each of the 15 parameter combinations found in the previous section we ran 100 runs of 6 experiments. This will generate an expected value for the average set of 6 experiments.

Figure 7 shows that the predicted pattern is close to the observed patterns. This is not true for the number of tokens in the resource over time after 90 seconds. One of the six experiments in the ASU data contained an experiment where the participants did not collect all the tokens. This is the only time out of about 50 experiments done with a low growth condition in which this happened. This indicates an outlier emergent group behavior which is not captured by the current model version. We do not have enough observations to test the occurrence of such events.

¹ We thank Robert Holahan for deriving the data from the experiments at Indiana University.

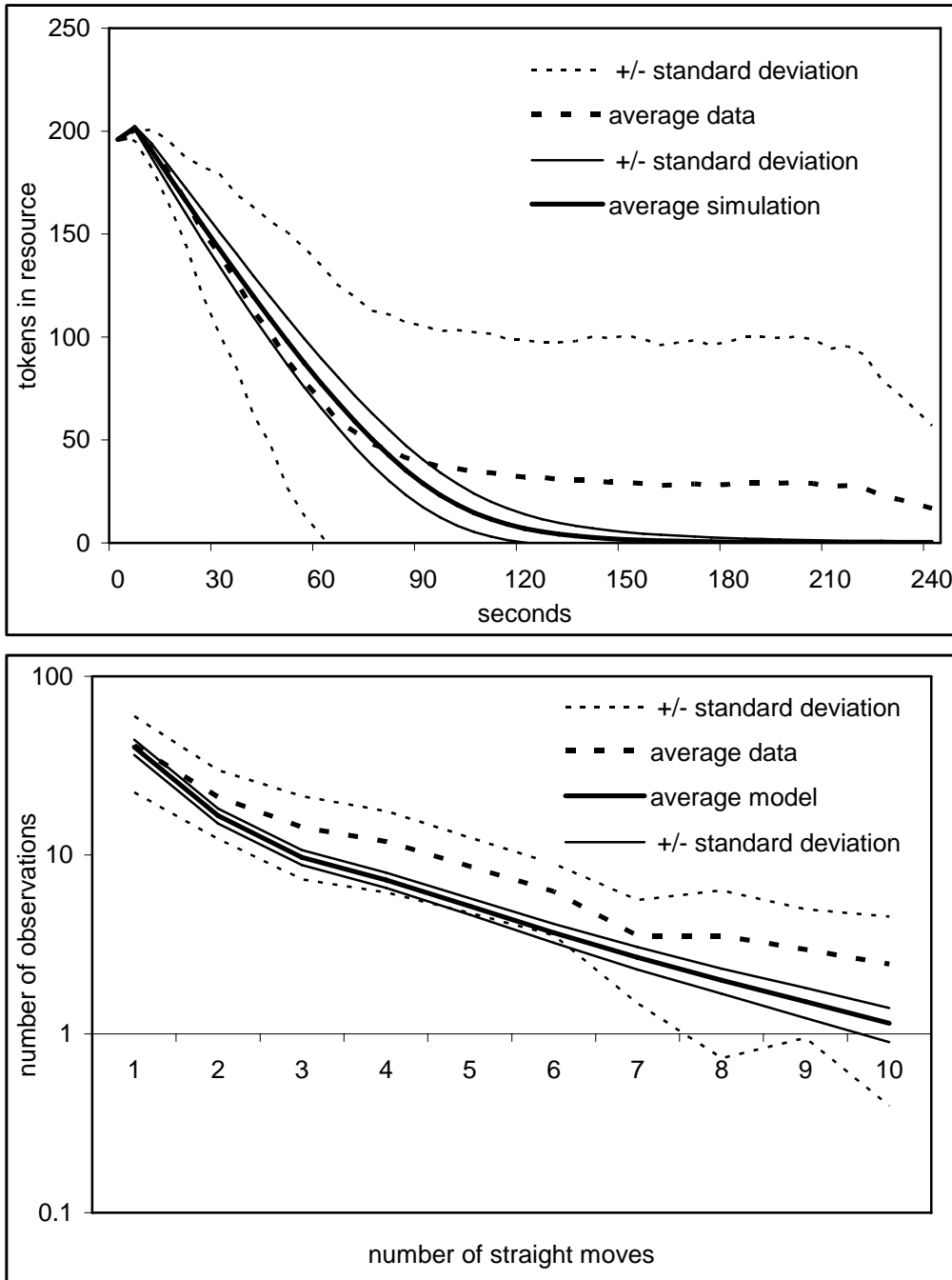


Figure 7: The predicted range of the ASU experiments (straight lines) and the average of actual observations (dotter line) with standard deviation of observations. The top figure present the average number of tokens in the resource, and the bottom figure presents the distribution for the number of straight moves.

We use the data of the ASU experiments as the best guess for determining the speed for the IU experiments. The average number of movements per second is assumed to be represented by a normal distribution with a mean of 3 and a standard deviation of 0.8. Furthermore, we assume that distribution of key presses per second is distributed by the drawn mean and the estimated standard deviation: $stdev = 0.598 * Mean - 0.062 * Mean^2$.

We use 16 experiments from the Indiana University experiments². We use the model version based on the ASU experiments to calculate distributions of the four patterns of the average of the 16 experiments from Indiana University. For each of the 15 parameter combinations that met all the patterns (as found in the previous section), we ran the model 100 times 16 (the number of experiments) = 24000 times. This provides us with 1500 simulations of the average of 16 experiments. This information can then be used to define distributions of the patterns to test whether the predicted patterns were similar to the observed patterns.

We predicted that the average number of tokens collected by the groups would be 20 tokens higher in the IU experiments than in the ASU experiments (Table 5). The actual data indicates an increase of 25 tokens which is within the uncertainty range. The gini coefficient in the IU data set was predicted to be higher in the ASU experiment, while the data suggested the opposite (Table 6).

Table 5: Predicted average number of tokens collected with standard deviation between parentheses, and the actual average number of tokens collected for the 6 experiments at ASU and the 16 experiments at IU.

	Predicted	Actual
ASU data set	271.8 (8.5)	274
IU data set	291.8 (8.0)	299.7

Table 6: Predicted average gini coefficient of tokens collected with standard deviation between parentheses, and the actual gini coefficients for the 6 experiments at ASU and the 16 experiments at IU.

	Predicted	Actual
ASU data set	0.133 (0.027)	0.232
IU data set	0.141 (0.016)	0.169

The predicted number of tokens in the resource over time follows closely during the first 80 seconds, but remains higher afterward (Figure 8). The observed distribution of the number of straight moves falls within the uncertainty margins of the predictions. There is a slightly higher count of short number of straight moves in the observed behavior than predicted.

In summary, the predicted patterns fit very well, though not perfectly, with the observations. In the observations participants slowed down when the resource started to become depleted. A possible explanation is that the participants were only able to use the explicit mode which may have caused the participants to be more restrained.

² Other treatments in this study included communication and/or costly sanctioning in the first round.

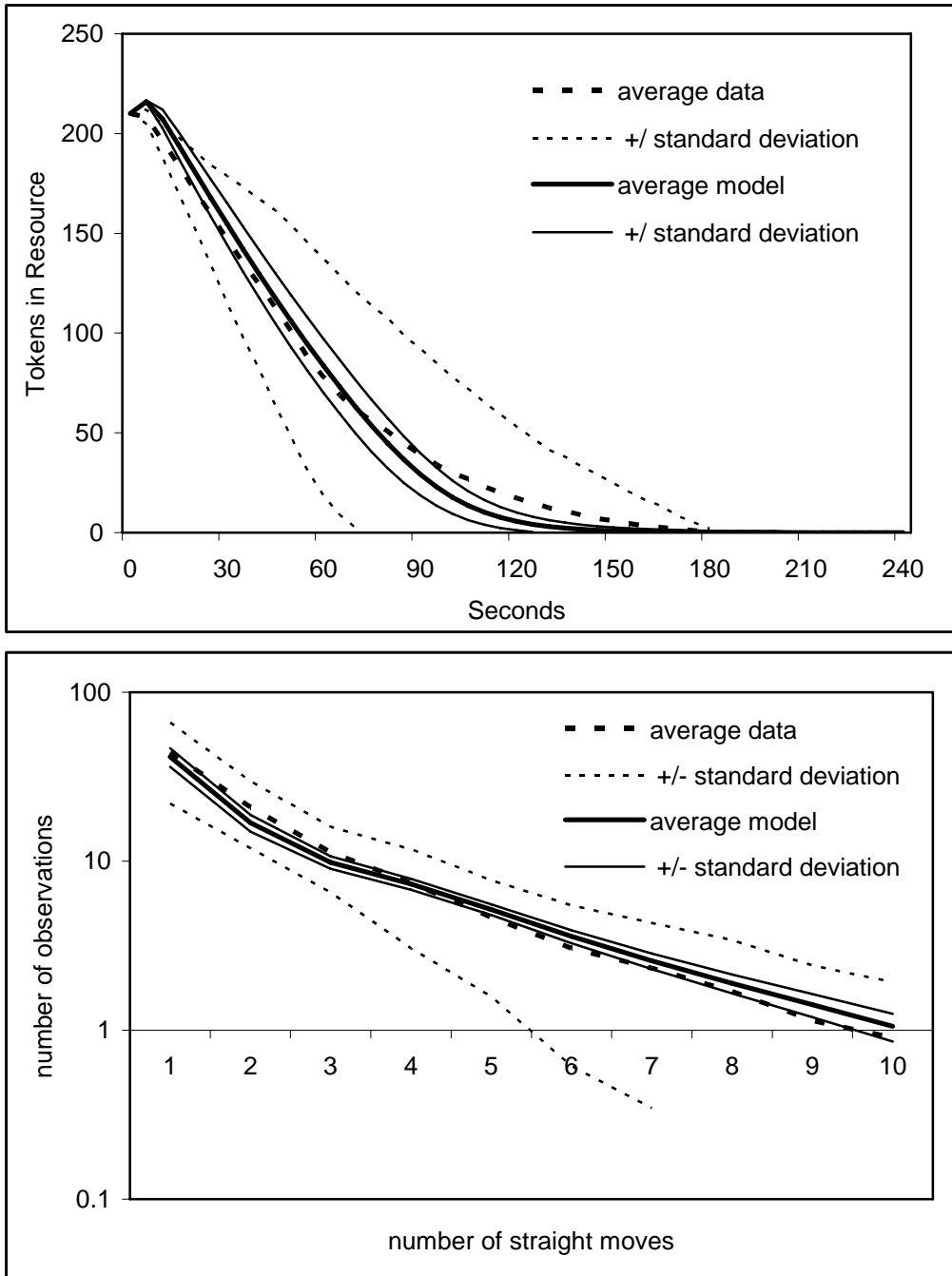


Figure 8: The predicted pattern distributions of new experiments with different design and the actual observed patterns. The lines in the bottom figures represent the average (thick line) of the simulations, and the data. The other lines represent the average plus and minus the standard deviation. The dotted lines present the average and plus and minus the standard deviation of the observed patterns.

8 Conclusions

In this paper we apply pattern-oriented modeling to identify parameter combinations that meet empirical patterns from a number of human subject experiments. We find that less than 0.1% of the parameterizations explored meet all of the patterns, and as such it helped to successfully narrow down the parameter space of acceptable parameter combinations.

Instead of finding *the* model that fit the data the *best*, pattern oriented modeling is successful in identifying model versions that are able to explain observed patterns of the data

at different scales. The model versions that meet all the patterns of two treatments let agents weight more nearby tokens and whether tokens were in a straight line of the current direction, than whether another participant is nearby the token. Unlike simple naïve models such as greedy agents and random walk, the comprehensive model is able to explain the observed patterns. Moreover, we were successful in predicting the range of outcomes of new experiments.

The primary goal of this paper is to show the possible use of pattern oriented modeling for behavioral research in the social sciences using a data rich example of micro-level decisions. Due to the lack of having an exact structural model the model testing process has to deal with equifinality, where different models can lead to the same patterns. The pattern oriented modeling approach allows for finding those model versions which are acceptably matching the observed patterns according to criteria of the investigators. Since we frequently interested in detecting model versions that are able to explain patterns at different levels of spatial and temporal scale, we are less concerned of tuning parameter at high accuracy. Deriving a set of parameter combinations that match the patterns enables us to use this for making predictions of distributions of new situations using a set of models that explained the earlier observations. This can be done for new experiments, but if applied to models of, for example, pedestrian behavior, the simulation of behavior for different spatial architectures.

This study shows with confidence that the pattern-oriented modeling approach is a promising tool for testing agent-based models on empirical data in the social sciences and we hope that this study stimulate the applications to other studies of adaptive behavior.

Notes

¹ We have done the same analysis for round 1, the individual round, and did the same insights as for round 2 of the experiment. Therefore we only report on round 2 of the experiment, the group experiments.

² <http://mathworld.wolfram.com/GiniCoefficient.html> accessed May 19, 2008

³ We systematically explore the parameter space for two reasons. First, we are interested in parameter settings that satisfy the pattern boundaries, not to maximize the fit with the data. Second, since the agent-based model is stochastic, using optimization methods to find parameter settings that match the patterns, it may get stuck in noisy fitness landscapes. The systematic analysis lead to a low resolution of the parameter settings, but since we are interested in the areas of parameter settings that match the patterns instead of the best fit, this is acceptable for our purposes.

⁴ We thank Robert Holahan for deriving the data from the experiments at Indiana University.

Acknowledgments.

We acknowledge support from the National Science Foundation (BCS-0432894) and comments received from Robert Tobias on an earlier version of this manuscript.

References

- Beven, K. (1998). A manifesto for the equifinality thesis, *Journal of Hydrology*, 320(1-2): 18-36.
- Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., & DeAngelis, D. L. (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science*, 310, 987-991.
- Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162, 1243-1248.
- Janssen, M. A., Goldstone, R. L., Menczer, F., & Ostrom, E. (2008). Effect of rule choice in dynamic interactive spatial commons. *International Journal of the Commons*, 2(2): 288-312.
- Janssen, M. A., & Ostrom, E. (2006). Empirically based agent-based modelling. *Ecology and Society*, 11(2), 37.
- Janssen, M.A., & Ostrom, E. (2008). TURFs in the lab: Institutional Innovation in dynamic interactive spatial commons, *Rationality and Society*, 20: 371-397.
- Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., & Balan, G. (2005). MASON : A Multi-Agent Simulation Environment. *Simulation*, 81(7), 517-527.
- Ostrom, E., Gartner, R., & Walker, J. (1994). *Rules, Games and Common Pool Resources*, University of Michigan Press.
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences* 6(10):421-425.
- Roberts, M. E., & Goldstone, R. L. (2006). EPICURE: Spatial and Knowledge Limitations in Group Foraging. *Adaptive Behavior*, 14, 291-313.
- Robinson, D. T., Brown, D. G., Parker, D. C., Schreinemachers, P., Janssen, M. A., Huigen, M., Wittmer, H., Gotts, N., Promburom, P., Irwin, E., Berger, T., Gatzweiler F., & Barnaud, C. (2007). Comparison of empirical methods for building agents-based models in land use change. *Journal of Land Use Science*, 1(1), 31-55.
- von Bertalanffy, L. (1968). *General System Theory: Foundations, Development, Applications*. New York: George Braziller.
- Wiegand, T., Jeltsch, F., Hanski, I., & Grimm, V. (2003). Using pattern-oriented modelling for revealing hidden information: a key for reconciling ecological theory and conservation practice. *Oikos*, 100, 209–222.