Springer Proceedings in Complexity

Zining Yang
Elizabeth von Briesen *Editors*

Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas

Springer

Springer Proceedings in Complexity

Springer Proceedings in Complexity publishes proceedings from scholarly meetings on all topics relating to the interdisciplinary studies of complex systems science. Springer welcomes book ideas from authors. The series is indexed in Scopus.

Proposals must include the following:

- name, place and date of the scientific meeting
- a link to the committees (local organization, international advisors etc.)
- scientific description of the meeting
- list of invited/plenary speakers
- an estimate of the planned proceedings book parameters (number of pages/articles, requested number of bulk copies, submission deadline)

Submit your proposals to: Hisako.Niko@springer.com

More information about this series at http://www.springer.com/series/11637

Zining Yang · Elizabeth von Briesen
Editors

Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas

Springer

Editors
Zining Yang
Computational Analytics
Claremont Graduate University
Claremont, CA, USA

Elizabeth von Briesen
Computer Science
Elon University
Elon, NC, USA

ISSN 2213-8684 ISSN 2213-8692 (electronic)
Springer Proceedings in Complexity
ISBN 978-3-030-77516-2 ISBN 978-3-030-77517-9 (eBook)
https://doi.org/10.1007/978-3-030-77517-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Editors and Contributors

About the Editors

Dr. Zining Yang is Senior Manager at Southern California Edison. She also works as Clinical Professor at Claremont Graduate University and Associate Director at the TransResearch Consortium. She sits on the Board of the Computational Social Science Society of the Americas (CSSSA) and serves as Scientific Advisory Board Member for Human Factors and Simulations. Dr. Yang received her Ph.D. in Computational and Applied Mathematics and Political Economy from Claremont Graduate University in 2015. Her research interests include Data Analytics, Machine Learning, Modeling and Simulation, Complex Adaptive Systems, Agent-Based Models, and Network Analysis. Dr. Yang has published numerous times in the fields of Computer Science, Economics, Public Policy, and Political Science. She has been identified as an outstanding researcher by the government, worked on a National Science Foundation-sponsored project, and won multiple awards from various organizations, including the Ministry of Education of the People's Republic of China; International Social Computing, Behavioral Modeling and Prediction; and the International Institute of Informatics and Systemics.

Dr. Elizabeth von Briesen is Assistant Professor of Computer Science at Elon University and is member of the board of the Computational Social Science Society of the Americas. She received her Ph.D. in Computing & Informatics from the University of North Carolina at Charlotte in 2020. Her research interests are focused on the study of complex adaptive systems using computational techniques, particularly with respect to social systems experiencing identity-based conflict. She primarily works with agent-based models, and performs data mining and sentiment analysis to inform those simulations. Finally, in her current position, Dr. von Briesen strives to contribute toward an evolving undergraduate computer science experience through research, service, and high-quality, innovative teaching.

Contributors

Amira Al-Khulaidy George Mason University, Fairfax, VA, USA

Luis Alonso MIT Media Lab, Cambridge, MA, USA

Georgios C. Anagnostopoulos Florida Institute of Technology, Melbourne, USA

Ian Apperly School of Psychology, University of Birmingham, Birmingham, UK

Chad Atalla University of California, San Diego, CA, USA

Ellen Badgley The MITRE Corporation, McLean, VA, USA

Niloofar Bagheri-Jebelli George Mason University, Fairfax, USA

Nisha Baral University of Central Florida, Orlando, FL, USA

Christina Bouwens University of Central Florida, Orlando, USA

Paul Brenner Center for Research Computing, University of Notre Dame, Notre Dame, USA

Amy Canevello UNC Charlotte, University City Blvd, Charlotte, NC, USA

Yiding Cao Binghamton University, State University of New York, Binghamton, NY, USA

Eric Castellanos Center for Research Computing, University of Notre Dame, Notre Dame, USA

Justin Chin USC Gould School of Law, Los Angeles, CA, USA

Garrison Cottrell University of California, San Diego, CA, USA

John Cox UNC Charlotte, University City Blvd, Charlotte, NC, USA

Andrew Crooks University at Buffalo, Buffalo, USA

Stephen Davies University of Mary Washington, Fredericksburg, USA

Paul K. Davis Pardee RAND Graduate School, Santa Monica, CA, USA

Xinwei Deng Department of Statistics, Virginia Tech, Virginia Beach, VA, USA

Shelley D. Dionne Binghamton University, State University of New York, Binghamton, NY, USA

Yingjun Dong Binghamton University, State University of New York, Binghamton, NY, USA

Gabriel Fair University of North Carolina at Charlotte, Charlotte, USA

Fei Fang Carnegie Mellon University, Pittsburgh, PA, USA

Stephen M. Fiore University of Central Florida, Orlando, USA

Milton J. Friesen University of Waterloo, Waterloo, ON, Canada

Ivan Garibay Complex Adaptive System Laboratory (CASL), University of Central Florida, Orlando, FL, USA

Arnaud Grignard MIT Media Lab, Cambridge, MA, USA

Chathika Gunaratne University of Central Florida, Orlando, FL, USA

Mirsad Hadžikadić UNC Charlotte, University City Blvd, Charlotte, NC, USA

Neda B. Hajiakhoond University of Central Florida, Orlando, USA

Dietmar Heinke School of Psychology, University of Birmingham, Birmingham, UK

Zhihao Hu Department of Statistics, Virginia Tech, Virginia Beach, VA, USA

Jasser Jasser University of Central Florida, Orlando, USA

Chathura Jayalath University of Central Florida, Orlando, FL, USA

William G. Kennedy George Mason University, Fairfax, USA

Minjun Kim Binghamton University, State University of New York, Binghamton, NY, USA

Caroline C. Krejci Department of Industrial, Manufacturing, & Systems Engineering, The University of Texas at Arlington, Arlington, TX, USA

Ankita Kulkarni Binghamton University, State University of New York, Binghamton, NY, USA

Bruce Levin Emory University, Atlanta, GA, USA

Linjie Li University of California, San Diego, CA, USA

Caroline Lourenço Alves University of Sao Paulo, Sao Paulo, Brazil

Neil G. MacLaren Binghamton University, State University of New York, Binghamton, NY, USA

Alexander Mantzaris University of Central Florida, Orlando, USA

Achla Marathe Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

Tim McDonald Pardee RAND Graduate School, Santa Monica, CA, USA

Anuj Mittal Department of Industrial Engineering Technology, School of Engineering, Dunwoody College of Technology, Minneapolis, MN, USA

Srikanth P. Mudigonda Saint Louis University, St. Louis, MO, USA

John T. Murphy University of Chicago, Chicago, USA

Ece Çiğdem Mutlu Complex Adaptive System Laboratory, University of Central Florida, Orlando, Florida, USA

William Nehrboss Lake Anna Homeschool, Bumpass, USA

Olivia B. Newton University of Central Florida, Orlando, USA

Thi Nguyen Internal Revenue Service (IRS), Research, Applied Analytics and Statistics (RAAS), Washington, DC, USA

Santiago Núñez-Corrales Illinois Informatics and National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Champaign, USA

Angela O'Mahony Pardee RAND Graduate School, Santa Monica, CA, USA

Toktam A. Oghaz University of Central Florida, Orlando, USA

Nilufer Oran Gibson Department of Industrial, Manufacturing, & Systems Engineering, The University of Texas at Arlington, Arlington, TX, USA

Osonde Osoba Pardee RAND Graduate School, Santa Monica, CA, USA

Gamaliel Palomo Cinvestav Guadalajara, Zapopan, JA, Mexico

Ann Pendleton-Jullian Pardee RAND Graduate School, Santa Monica, CA, USA

Amirarsalan Rajabi Complex Adaptive Systems Lab (CASL), University of Central Florida, Orlando, USA

William Rand North Carolina State University, Raleigh, NC, USA

Subash K. Ray New Jersey Institute of Technology & Rutgers University - Newark, Newark, NJ, USA

Daria Roithmayr USC Gould School of Law, Los Angeles, CA, USA

Samaneh Saadat University of Central Florida, Orlando, USA

Juan Salamanca School of Art+Design, University of Illinois at Urbana-Champaign, Champaign, USA

Anastasia Salter University of Central Florida, Orlando, USA

Hiroki Sayama Binghamton University, State University of New York, Binghamton, NY, USA

Steven Scheinert University of Central Florida, Orlando, USA

Madeline Schiappa University of Central Florida, Orlando, USA

Chathurani Senevirathna University of Central Florida, Orlando, FL, USA

Samira Shaikh UNC Charlotte, University City Blvd, Charlotte, NC, USA

Peer-Olaf Siebers School of Computer Science, University of Nottingham, Nottingham, UK

Mario Siller Cinvestav Guadalajara, colonia el Bajío, Zapopan, JA, Mexico

John Silver George Mason University, Fairfax, VA, USA

Amanda Song University of California, San Diego, CA, USA

Mel Stanfill University of Central Florida, Orlando, USA

Gita Sukthankar University of Central Florida, Orlando, USA

Samarth Swarup Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA

Seyyedmilad Talebzadehhosseini Complex Adaptive Systems Lab (CASL), University of Central Florida, Orlando, USA

Bartholomew Tam University of California, San Diego, CA, USA

Rajesh Venkatachalapathy Portland State University, Portland, USA

Valentin Vergara George Mason University, Fairfax, VA, USA

Melissa Vigil Internal Revenue Service (IRS), Research, Applied Analytics and Statistics (RAAS), Washington, DC, USA

Elizabeth von Briesen UNC Charlotte, University City Blvd, Charlotte, NC, USA

Anil Vullikanti Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia Department of Computer Science, University of Virginia, Charlottesville, VA, USA

Maria Waldl University of Vienna, Vienna, Austria

Rachel Winter University of Central Florida, Orlando, USA

Hang Xie Department of Computer Science, Wenzhou-Kean University, Wenzhou, China

Francis J. Yammarino Binghamton University, State University of New York, Binghamton, NY, USA

Niloofar Yousefi University of Central Florida, Orlando, USA

Zahrieh Yousefi School of Psychology, University of Birmingham, Birmingham, UK

Xi Zhang Florida Institute of Technology, Melbourne, USA

Social Impression of Faces: From Prediction to Modification

Amanda Song, Chad Atalla, Bartholomew Tam, Linjie Li, and Garrison Cottrell

Abstract First impressions are influential in shaping our personal, economic, and political decisions. We develop a computational framework that can model and modify impressions of faces. First, we use a state-of-the-art predictive model of facial impressions (such as facial attractiveness, trustworthiness, and intelligence) and apply it to a large-scale natural face dataset in order to create a robust facial impression dataset. We validate the augmented dataset with respect to human judgments. Second, we use the new dataset to train a model, ModifAE, that changes face smoothly and effectively in multiple social dimensions. This modification model offers social scientists the ability to manipulate impressions as needed, and it sheds light on both the biases and the visual features underlying first impression formation.

1 Introduction

Humans quickly form subjective impressions of faces, judging traits like facial attractiveness, trustworthiness, and aggressiveness [1]. Despite the continuous scale and subjective nature of these social judgments, there is often a consensus among humans in how traits are perceived; for example, human raters will agree that certain faces appear relatively more trustworthy [2, 3]. Social judgments of faces have a significant impact on social outcomes, ranging from electoral success to sentencing decisions [4, 5]. Modeling is one way to understand these critical split-second impressions. Another way is through explicit human-judged experiments, which require carefully controlled datasets (e.g., building a dataset of faces that vary in "trustworthiness" while remaining consistent across age, gender, and "attractiveness"). In this work, we develop a system to model these impressions, predict human average impressions on facial images, visualize human perceptual biases, and create isolated image modifications for experimental datasets.

Choosing a subset of social impressions for modeling, we look to the 10k US Adult Faces Database [6]. Bainbridge et al. [6] investigated what social traits influence the

A. Song (✉) · C. Atalla · B. Tam · L. Li · G. Cottrell
University of California, Gilman Dr, La Jolla 9500, San Diego, CA 92122, USA

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_1

memorability of a face. They compiled a list of 20 spontaneous social judgments and the corresponding opposite traits. Then, they assembled a human-judged dataset of trait ratings on 2,222 faces from the 10k US Adult Faces Database. Among the 40 traits, "trustworthy", "attractive,", "aggressive," and "intelligent" were frequently used in human-written face descriptions, played a significant role in face memorability, and had relatively high rating agreement levels between human judges. Therefore, we choose them as a representative subset of social impressions for modeling in this paper. Motivated by the success of deep learning in modeling visual properties, we use deep learning-based pretrained representations as the basis for learning to predict first impressions on realistic faces, training a predictive linear model that successfully predicts human social perception on faces whenever human have consensus.

To create controlled face datasets and visualize perceptual biases, a generative model is needed. Recent generative image models have been successful in creating high-resolution, high-fidelity, and diverse images [7–9]. However, in the face space, most generative models have focused on editing or modifying categorical and objective attributes, such as expression, gender, hair color, and identity [9]. These categorical changes are referred to as "image-to-image translation." Here, we focus on modifying continuous traits of an image, which we refer to as "continuous image modification" [10]. Regarding continuous image modification, there has been work on modifying the memorability [11] and attractiveness of a face [12], but these models do not generalize to wider sets of social impressions. Also, some researchers have endowed computer-generated faces with particular social impressions, but these models cannot modify real face images [5, 13]. So, no prior work has attempted to automatically modify general continuous social impressions of real face photographs. Part of the difficulty lies in the fact that training a high-fidelity generative model requires a large amount of data, yet there is no pre-existing dataset that has tens of thousands of faces with labeled social impression trait scores. We overcome this difficulty by proposing a cost-effective and easy-to-scale-up method to construct a large-scale facial impression dataset.

Conditional generative adversarial networks (GANs) [14] have become the most popular tool for the image-to-image translation task, so we compare against a recent GAN as a state-of-the-art (SOTA) reference point [10, 15, 16]. StarGAN [9] is a SOTA conditional GAN that can modify multiple binary categorical traits of faces at once, maintaining the identity of the face using a "cycle consistency" loss function, which translates the face back to the original one [17]. StarGAN consists of two networks: a generator and a discriminator. The generator takes an image and a set of desired categorical traits, producing a modified image. The discriminator takes an image and makes a prediction about its realism and categorical traits. By comparing the fake images to genuine images, the discriminator gives feedback to the generator about how to make the image and desired traits appear more realistic.

Despite the success of GANs in categorical image-to-image translation, they cannot perform continuous image modification without binarizing the task. GANs typically have many parameters and long training times. They are also sensitive to hyper-parameter selection and the delicate balance between generator and discriminator training. Therefore, they can be difficult to train compared to a single-network

model. Finally, they suffer from a lack of interpretability, offering no means of visualizing or understanding why the model makes the modifications it does.

In this work, we address these architectural concerns while designing a neural network to model and automatically modify continuous-scale face traits (rated from 1 to 9) in real face images. First, we use our deep learning-based predictive linear model [18] to predict human facial impressions of attractiveness [3, 19], trustworthiness [2, 20], aggressiveness [21], and intelligence, to form an augmented dataset. We validated the effectiveness of this dataset augmentation method with human experiments.

With this large-scale realistic facial impression dataset, we train a deep modifying autoencoder, ModifAE, that can smoothly and naturally modify the first impressions of faces. We evaluate the model performance quantitatively and qualitatively and compare it with StarGAN. Notably, our generative model can modify multiple traits at once and can provide visualizations of group average trait features. It is also easy to train. These capabilities make it a powerful tool, which can, for example, modify multiple traits while controlling other high-level attributes, such as gender. We then quantify the actual changes ModifAE makes to modify perceived impressions, shedding light on what geometric features correspond to social impression dimensions.

2 Predicting Social Attributes of Faces

Previous studies have shown that pretrained deep learning models can provide feature representations versatile for related tasks [22]. After comparing multiple off-the-shelf pretrained neural networks, we find that conv5_2 layer of VGG16 (pretrained for object classification) leads to satisfactory results. After obtaining intermediate representations from the pretrained neural networks, we apply Principal Component Analysis to reduce the dimensionality, then train a ridge regression model to produce predictions on each social attribute, respectively. For "trustworthy", "attractive,", "aggressive," and "intelligent", the predictive model's correlations (Spearman rank correlation) with human averaging ratings are 0.73, 0.75, 0.72 , and 0.62, respectively, on the test set.

3 Creating a Large-Scale Facial Impression Dataset

To train a generative model on continuous face traits, we need a large and diverse dataset. We use images from the CelebA dataset [23], which consists of over 190,000 images of celebrities. The images in CelebA are annotated with binary categorical labels such as "wearing a hat" but not on continuous ratings of social impressions.

To generate continuous social impression traits of these faces, we use our predictive model mentioned above [18] trained on a smaller dataset (approximately 2,000

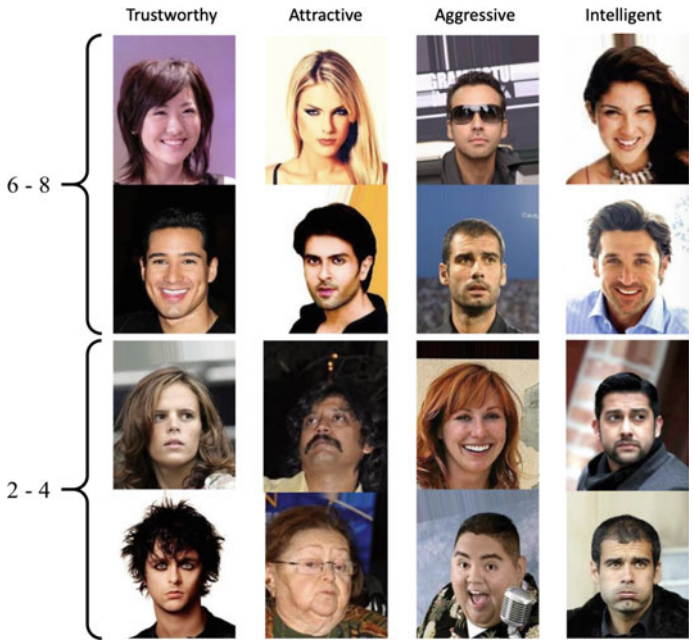


Fig. 1 CelebA faces and their predicted trait scores

faces from the 10k US Adult Faces Database [6]) that have been annotated with continuous ratings by 15 raters for each of the social traits.

We apply this model to over 190,000 faces from the CelebA dataset. The resulting model predictions are highly correlated with human judgments (*** denotes that $p < 0.0001$): trustworthy (0.73***), attractive (0.90***), aggressive (0.76***), and intelligent (0.62***). These correlations are obtained by asking subject to rate faces on these four traits and computing average human ratings' Spearman correlation with the model's predictions. The details of these experiments are given in the Methods.

Example faces and their predicted ratings are shown in Fig. 1. Note here that 6–8 are high ratings, and 2–4 are low ratings.

4 Validating the Algorithm-Augmented Dataset

To evaluate the effectiveness of this algorithm-augmented dataset, we collect human judgments of faces in CelebA and evaluate how model predictions correlate with human judgments. We examined four traits: attractive, aggressive, trustworthy and intelligent because they represent different aspects of first impressions and are of relatively high human agreement. For each trait, we chose 80 faces whose predicted scores are evenly spread across a range of predictions (i.e., from 2 to 8). Every

participant is presented with a random sequence of these 80 faces, and is asked to give each face a rating on a 1–9 scale for the specified trait. Every face is rated by roughly 15 subjects (ranging from 12 to 16), and we compute the average ratings for each face. Lastly, we compute the Spearman rank correlation between the average human ratings and the model's predictions. For all four traits, human average ratings are significantly correlated with model predictions (*** indicates $p < 0.001$): trustworthy (0.73***), attractive (0.90***), aggressive (0.76***), and intelligent (0.62***). We plan to publicly release the large-scale facial impression dataset for future researchers' use.

5 ModifAE: A Modification Model of Social Impressions

With the large-scale validated first impression face dataset, we train our modification model, ModifAE. The network is trained on an autoencoding task (reproducing the input on the output) with an added input corresponding to the trait value (see Fig. 2). By using aggressive dropout on the image side (a technique where a random half of the activations are set to 0), the model implicitly learns to depend on the input trait value to generate the reconstruction of the image. This enables the model, after training, to use different trait values to modify the image.

5.1 Architecture

The ModifAE architecture consists of a single autoencoder with two (image and trait) sets of inputs that pass through an encoding stage and then are fused (by averaging them) in the middle of the network. This latent representation is then fed into an image decoder.

The image encoder and decoder are identical to the encode and decode portions of the StarGAN generator network, scaled to fewer channels [9]. More specifically, the network has two downsampling convolutional layers with stride two, four residual blocks, a bottleneck with 16 channels, four more residual blocks, then two upsampling transposed convolutional layers with stride two [9]. All layers have ReLU activation. We use the first half of this network (including the bottleneck) as the image encoder. We use the remainder of the network as the image decoder. Theoretically, this portion could consist of the encode and decode halves of any other image autoencoder.

The trait encoder takes a one-dimensional set of traits, feeds these into a single dense layer with Leaky ReLU activation, and reshapes the output to create a vector of the identical shape as the image encoder output. The outputs of the trait and image encoders are then combined into a single latent representation of the image and ratings.

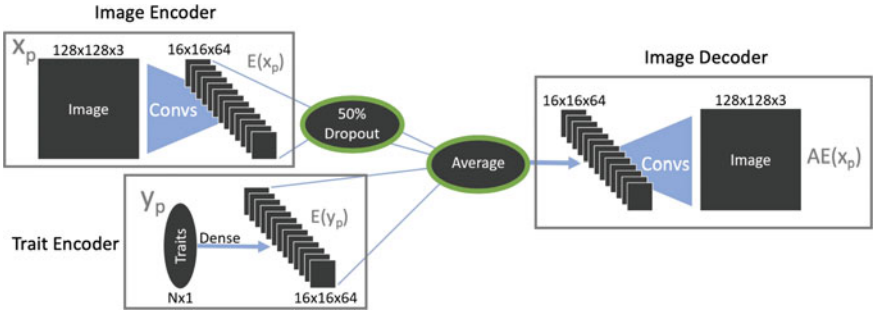


Fig. 2 General illustration of ModifAE architecture

In order to encourage the model to encode the trait information, which is otherwise unnecessary to reproduce the image, 50% dropout is applied to the values from the image encoder. This is then averaged with the trait encoder output to arrive at the combined latent representation. The image decoder projects the representation back into image space, creating the single output image. The architecture is depicted in Fig. 2.

Despite sharing some aspects of architecture with StarGAN's generator [9], ModifAE has over 50 times fewer parameters.

5.2 ModifAE Training Procedure

ModifAE is only trained on an autoencoding task. We train ModifAE using the Adam optimizer [24] and train for 100 epochs on our augmented CelebA images [23]. The objective is to optimize a single loss function based on two terms. We use the L_1 loss on the image autoencoder. We also optimize the L_1 loss between the trait encoder and image encoder. The total loss is

$$L = \frac{1}{N}\sum_{p=1}^{N} |x_p - AE(x_p)| + |E(x_p) - E(y_p)| \tag{1}$$

where x_p is the pth image example, y_p is its trait vector, $E(\cdot)$ is the result of the trait or image encoder, and $AE(\cdot)$ is the output of the full-architecture autoencoder. The second term in this loss function encourages the network to have a similar representation between the trait and the image encodings.

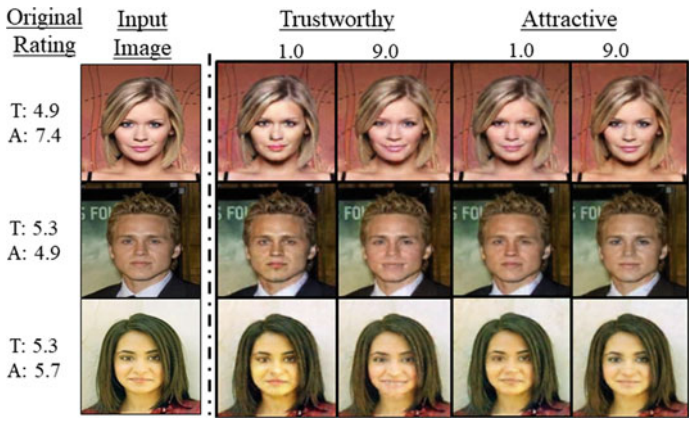


Fig. 3 Image modifications by ModifAE in trustworthy and attractive

5.3 How ModifAE Works

Each image is encoded along with its predicted traits. The image encoder compresses the image down to a bottlenecked latent space, where higher level features about the image are encoded. Simultaneously, the trait encoder projects the given traits to the same latent space, creating an average face representation with those ratings (Fig. 3).

Results

First, we qualitatively examine ModifAE's modifications and visualizations of trait representations and then quantitatively compare the modifications of ModifAE and StarGAN with human behavioral studies.

5.4 Qualitative Evaluation

We obtain visualizations of ModifAE's trait representations by presenting the model with trait values in the absence of any image input. The resulting transformation maps show ModifAE's representation of a trait at different trait values. These transformation maps can be produced from models which were trained on multiple traits, enabling visualizations of how ModifAE perceives some traits to vary independently of others. Figure 5 shows traversals of"attractive," "intelligent," "trustworthy," and "aggressive," while holding gender constant. Through this method, ModifAE addresses the issue of interpretability in generative models. These images provide a

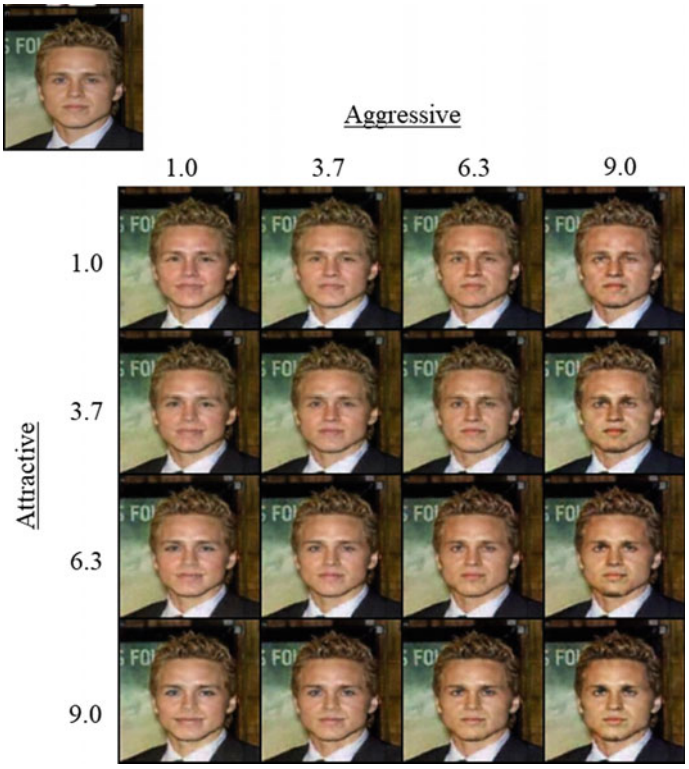


Fig. 4 Multi-trait image modification by ModifAE

window into how ModifAE represents each trait and how ModifAE changes a face to increase or decrease a given trait value.

In general, it appears that faces that subtend a larger visual angle are rated more positively, with the trend for the faces to get bigger from left to right for the three positive traits, and the opposite for aggressive. To our knowledge, this has not been observed previously and hence is a prediction of our model.

Similarly, a larger smile results in more positive ratings, with big smiles on the right for the positive traits and on the left for the negative trait. This accords with our intuition and is consistent with previous research that demonstrates smiling is associated with positive person perception [25].

For attractiveness, in addition to the larger smile corresponding to more attractive, at the unattractive end of the scale, there is lower contrast in the face features.

We also are able to modify two traits at once, by training on both trait values in a single network (see Fig. 4). For this experiment, we trained ModifAE on two traits: "attractive" and "aggressive." The picture in the upper left corner is the original. Looking at the (1, 1) point in Fig. 4 (unattractive and not aggressive), the man's mouth is fairly neutral, and his features are not very pronounced. As attractiveness

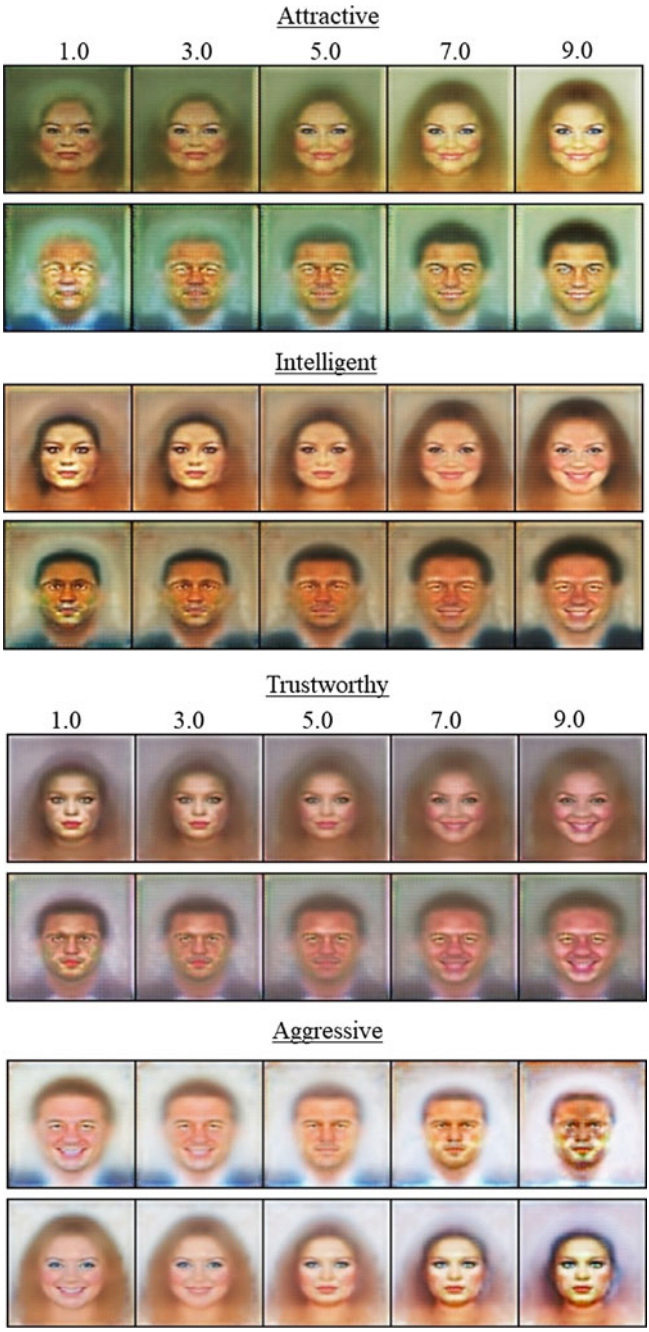


Fig. 5 Visualization of model's internal perception of traits. Each is a traversal of a trait (increasing left to right) while gender is held constant

and aggressiveness increase, the angles of the face become sharper, there is more definition of features like eyes and eyebrows, and the smile shrinks.

In the "attractive" transformation map of Fig. 5, age is a salient factor, with rounder, pudgier, older faces appearing on the left side of the traversal, and faces with sharper features, clearly defined large eyes, and larger smiles appear on the more attractive side. This makes sense and is a confirmation that the model reflects human biases.

In the "intelligent" transformation maps, the degree of smile increases left to right, the hair gets longer, and the eyebrows get smaller, perhaps suggesting that large brows are perceived as less intelligent. Additionally, the head size clearly increases from left to right. This suggests that images in which the head subtends a greater visual angle are perceived as more intelligent, a bias that, to our knowledge, has not been previously observed. Of course, in this case, it is not the real-world size that matters, just the size of the head in the image. This is a prediction of our model.

Similar to the attractiveness traversal, the trustworthiness increases with the degree of smiling. The forehead gets larger, and there is a slight lightening of the hair, suggesting older people are more trustworthy.

For aggressiveness, clearly the bigger the smile, the less aggressive the face appears, which fits well with our intuition. Also, the visual angle of the face becomes smaller, and to us, at least, the eyes appear "beadier." Unfortunately, there also appears to be a racial bias, with darker skin appearing more aggressive.

5.5 Quantitative Evaluation

To evaluate the quality of ModifAE's continuous subjective trait modifications, we performed Amazon Mechanical Turk (AMT) experiments on the four traits we focus on in this article: aggressive, attractive, trustworthy, and intelligent. For each trait, we created 90 image pairs, of which 80 are the same identity modified to be at high and low values of each trait. For StarGAN, we used a median split of low and high-rated traits to train the model, making the transformation binary. ModifAE was trained as previously described. For each model, then, faces were modified to be low or high on each trait. Subjects judged which face had more of the particular trait. 10 pairs were repeats in order to judge subject consistency, and 10 pairs were unmodified CelebA faces with high and low ratings. This latter we called "ground truth" pairs to test whether subjects were paying attention. Subjects whose ratings on these pairs were at chance or below were rejected.

Hence, for each trait, we present participants with a sequence of 100 image pairs, and participants are asked to pick which image most exemplifies the trait in each pair.[1] Each pair was evaluated by 15 subjects.

[1] In a pilot experiment, we asked subjects to rate faces with different identities generated in a fine continuum, but found significant variance with no correlation to the intended scores, presumably because the images were not differentiable at that fine a grain.

Table 1 Comparison of ModifAE with StarGAN

Attribute	ModifAE	StarGAN	"Ground Truth"
Aggressive	0.68***	0.72***	0.90***
Attractive	0.68***	0.51	0.94 ***
Trustworthy	0.63***	0.40	0.87***
Intelligent	0.68***	0.58***	0.81***

*$p < 0.05$, **$p < 0.001$, ***$p < 0.0001$

We calculate the fraction of pairs in which subjects chose the image with the higher modified trait across all participants and all pairs. If they choose the face that was modified to be higher in the trait, then they agree with the model's modifications. The results are shown in Table 1. We perform a binomial test to determine whether each trait's accuracy is significantly below or above chance (***$p < 0.001$). Note that the fourth column "Ground Truth" indicates the overall accuracy of the unmodified "ground truth" pairs. Given the variance in human impression judgments, these numbers serve as a reference ceiling for how well the models can perform.

From Table 1, we can see that for all four traits, ModifAE produces pairs that yield above chance level human agreement. In three out of the four traits, ModifAE significantly outperforms StarGAN; whereas, for the aggressive trait, StarGAN performs only slightly better than ModifAE. StarGAN is good at creating discrete changes in facial expressions, which accounts for this advantage.

Since ModifAE is able to generate continuous modifications, we evaluated this property by creating two more same-face pairs: ones modified to have low values and middle values, and ones modified to have middle values and high values. We obtain human agreement (accuracy) over the Low-Mid and Mid-High pairs for each of the four traits. The results are shown in Table 2.

From Table 2, we find that all the low-mid pairs yield significantly above chance accuracy, yet for mid-high level, only trustworthy pairs have accuracy slightly above chance ($p < 0.05$*). This suggests that human psychological face space is nonlinear and has more differentiation toward the low- to mid-range of social dimensions. Another possibility is that when our model generates faces that are of more extreme scores (e.g., 8 or 9), the model is extrapolating and produces artifacts that lead to that face being rejected. This speculation requires further analysis to be confirmed.

Table 2 ModifAE Low-Mid-High level self-comparison

Attribute	Low-Mid	Mid-High	Low-High
Aggressive	0.60***	0.52	0.68***
Attractive	0.59***	0.52	0.68***
Trustworthy	0.61***	0.53*	0.63***
Intelligent	0.60***	0.50	0.68***

*$p < 0.05$, **$p < 0.001$, ***$p < 0.0001$

5.6 Qualitative Interpretations

With a hypothesis-driven approach, psychologists have identified certain visual features that contribute to specific impressions. The symmetry of the face [26] can explain why certain faces look more attractive. Other global face features such as femininity, babyfaceness [27], typicality [28], and facial width-to-height ratio (fWHR) [29] drive different aspects of social impression perception (warmth, honesty, submissiveness, dominance, etc.). Emotions such as perceived anger and happiness drive aggressiveness and trustworthiness perceptions, respectively. Using morphing and averaging methods, studies [30] have established that age also serves an important role in social perception of attractiveness, trustworthiness, and dominance.

6 Discussion

We have shown that a deep network can be used to predict the human social perception of faces, achieving a high correlation with the average human ratings. As far as we know, this is the widest exploration of social judgment predictions, showing human-like perceptions on 40 social dimensions. By predicting this as a continuous value, rather than categorical, the subjective nature of human judgment is modeled smoothly, along with the subjective face trait landscape.

Of greater significance is our model's correlations with human judgments for traits such as trustworthiness, responsibility, confidence, and intelligence, which correspond to more static features of the face. In this area, the deep network, which responds to facial textures and shape, has superior performance. While these judgments do not correspond to the traditional notion of "ground truth", they are descriptions for which humans have a fair amount of agreement, suggesting the presence of a signal to be recognized. Furthermore, we have shown that our prediction model can generalize reasonably well to an entirely new dataset, making it widely applicable to real-world scenarios.

We further develop a generative model, ModifAE, which can modify a face's social impressions while preserving its realism. ModifAE can change a face's perceived social features (e.g., make a face look more sociable, trustworthy). It can also produce transformation maps that elegantly summarize the average opinions and biases of a group of raters who have created a dataset. This functionality enables psychologists to quantify human biases during the formation of social impressions in a precise and systematic manner. Psychologists could generate variants of a real face differing in age, gender, and race while holding other traits constant. This controlled dataset could be used to explore how various factors separately and jointly affect the social impressions of faces.

Our computational models make predictions and modifications regarding the first impressions of faces, and such first impressions are indicative of implicit bias toward different social groups [31]. With knowledge of people's first impressions, along

with the embedded potential bias, we have a chance to analyze the perceptual and social interactions that are fundamental to humans.

These results are also significant for the field of social robotics and the fight against discrimination. Predictive models like this can bring empathy to robotics, where technology can help us bridge the emotional and social divide and promote social equality. Empathetic technology can benefit people who are implicitly discriminated against based on social impressions. While a robot should not purely judge a human on appearance, much of human interaction is dictated by the underlying fabric of social impressions. Thus, it is important for a robot to be aware of the subjective social fabric, opening the door to useful knowledge such as whether humans might judge a person to be trustworthy. These judgments may happen subconsciously for humans, while a robot can be more objective, predicting these judgments and objectively choosing when to consider them in a decision. A robot need not treat an attractive or unattractive person differently for its own purposes, but this knowledge could affect how interactions are made for the sake of the human, knowing in advance how that person may feel that they fit into the social landscape. These applications can have significant societal effects.

References

1. Todorov, A., Olivola, C.Y., Dotsch, R., Mende-Siedlecki, P.: Social attributions from faces: determinants, consequences, accuracy, and functional significance. Annu. Rev. Psychol. **66**(1), 519 (2015)
2. Falvello, V., Vinson, M., Ferrari, C., Todorov, A.: The robustness of learning about the trustworthiness of other people. Soc. Cogn. **33**(5), 368 (2015)
3. Eisenthal, Y., Dror, G., Ruppin, E.: Facial attractiveness: beauty and the machine. Neural Comput. **18**(1), 119–142 (2006)
4. Dumas, R., Testé, B.: The influence of criminal facial stereotypes on juridic judgments. Swiss J. Psychol. **65**(4), 237–244 (2006)
5. Oosterhof, N.N., Todorov, A.: The functional basis of face evaluation. Proc. Natl. Acad. Sci. **105**(32), 11087–11092 (2008)
6. Bainbridge, W.A., Isola, P., Oliva, A.: The intrinsic memorability of face photographs. J. Exp. Psychol.: Gen. **142**(4), 1323 (2013)
7. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis (2018). arXiv:1809.11096
8. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation (2017). arXiv:1710.10196
9. Choi, Y., Choi, M., Kim, M., Ha, J., Kim, S., Choo, J.: Stargan: unified generative adversarial networks for multi-domain image-to-image translation. CoRR (2017) arXiv:1711.09020
10. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CoRR (2016). arXiv:1611.07004
11. Khosla, A., Bainbridge, W.A., Torralba, A., Oliva, A.: Modifying the memorability of face photographs. In: International Conference on Computer Vision (ICCV-2013), pp. 3200–3207. IEEE (2013)
12. Leyvand, T., Cohen-Or, D., Dror, G., Lischinski, D.: Data-driven enhancement of facial attractiveness. ACM Trans. Graph. (TOG) **27**(3), 38 (2008)
13. Vernon, R.J., Sutherland, C.A., Young, A.W., Hartley, T.: Modeling first impressions from highly variable facial images. Proc. Natl. Acad. Sci. **111**(32), E3353–E3361 (2014)

14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
15. Mirza M., Osindero, S.: Conditional generative adversarial nets. CoRR (2014). arXiv:1411.1784
16. Lee, M., Seok, J.: Controllable generative adversarial network. CoRR (2017). arXiv:1708.00598
17. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
18. Song, A., Li, L., Atalla, C., Cottrell, G.: Learning to see people like people (2017). arXiv:1705.04282
19. Gray, D., Yu, K., Xu, W., Gong, Y.: Predicting facial beauty without landmarks. In: Computer Vision–ECCV 2010, pp. 434–447. Springer, Berlin (2010)
20. Todorov, A., Baron, S.G., Oosterhof, N.N.: Evaluating face trustworthiness: a model based approach. Soc. Cogn. Affect. Neurosci. **3**(2), 119–127 (2008)
21. Mignault, A., Chaudhuri, A.: The many faces of a neutral face: head tilt and perception of dominance and emotion. J. Nonverbal Behav. **27**(2), 111–132 (2003)
22. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)
23. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (2015)
24. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR (2014). arXiv:1412.6980
25. Reis, H.T., Wilson, I.M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., Franco, M., Gioioso, E., Freeman, L., Radoane, K.: What is smiling is beautiful and good. Eur. J. Soc. Psychol. **20**(3), 259–267 (1990)
26. Scheib, J.E., Gangestad, S.W., Thornhill, R.: Facial attractiveness, symmetry and cues of good genes. Proc. R. Soc. London. Ser. B: Biol. Sci. **266**(1431), 1913–1917 (1999)
27. Berry, D.S., Zebrowitz-McArthur, L.: What's in a face? facial maturity and the attribution of legal responsibility. Pers. Soc. Psychol. Bull. **14**(1), 23–33 (1988)
28. Sofer, C., Dotsch, R., Wigboldus, D.H., Todorov, A.: What is typical is good: the influence of face typicality on perceived trustworthiness. Psychol. Sci. **26**(1), 39–47 (2015)
29. Haselhuhn, M.P., Ormiston, M.E., Wong, E.M.: Men's facial width-to-height ratio predicts aggression: a meta-analysis. PLoS One **10**(4), (2015)
30. Sutherland, C.A., Oldmeadow, J.A., Santos, I.M., Towler, J., Burt, D.M., Young, A.W.: Social inferences from faces: ambient images generate a three-dimensional model. Cognition **127**(1), 105–118 (2013)
31. Stanley, D.A., Sokol-Hessner, P., Banaji, M.R., Phelps, E.A.: Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proc. Natl. Acad. Sci. **108**(19), 7710–7715 (2011)

Corruption and the Effects of Influence Within Social Networks: An Agent-Based Model of the "Lava Jato" Scandal

Amira Al-Khulaidy and Valentin Vergara

Abstract Corruption, and more specifically corruption in Latin America, is a complex phenomenon that is affected by politics, social structures, and institutions, as well as individual behaviors. The Lava Jato scandal was a large-scale example of corruption in Brazil. Advances in data analysis, computation, and social networks have allowed progress to be made with these types of investigations. The Lava Jato case has been a clear example of how breaking up social networks and understanding the extent of crime and individual corruption have revealed webs of corruption that have influenced politics, as well as hindered economic development in Brazil. The several layers of interactions between individuals and institutions can be difficult to grasp and understanding the patterns and relationships within complex large-scale phenomena such as corruption can seem impossible. Agent-based models can help with understanding these complex behaviors and systems. By capturing the patterns and gaining a better understanding of how corruption emerges and is manifested, we can help inform policy, as well as create better tools and methods for crime prevention and detection.

Keywords Corruption · Influence · Agent-based modeling · Social networks · Brazil · Lava Jato

1 Introduction

Corruption in Latin America has been a topic of interest across many fields. The historical, social, and political roots of Latin America are deeply tied to the origins of the socio-political systems that govern many Latin American countries [25]. Latin America has been at the forefront of political and economic discussions for decades;

A. Al-Khulaidy (✉) · V. Vergara
George Mason University, Fairfax, VA, USA
e-mail: aalkhula@gmu.edu

V. Vergara
e-mail: vvergara@gmu.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_2

however, more recently, several webs of events have emerged, and the effects are long-reaching with more complex ramifications. Corruption tends to focus on financial and political power, but in Latin America, there are several aspects that make the topic of corruption more complex in terms of the financial and legal intricacies [11]. This paper is focused specifically on corruption in Latin America and more specifically on the *Lava Jato* corruption scheme in Brazil.

The reach and intermingled web of the *Lava Jato* scheme have had global implications, with many contributing factors. The complex interactions of individuals, institutions, and transactions can be better explored in an agent-based model (ABM) with a focus on the social networks of the agents. ABMs and social network analysis (SNA) allow us to capture the patterns from data to build a model and then run a simulation to understand the patterns and results of the phenomenon we are investigating. From the model, we can better capture the more prominent patterns to gain a better understanding of how the different processes interact [10].

By building an artificial social network and simulating the transactions through time, we can derive a better understanding of how the *Lava Jato* corruption scheme occurs and how the impact of the various factors give rise to corrupt agents within a network. To better understand these factors, we will look at the categories of different players and how their relationships enable this level of corruption based on whether agents engage in corruption, as well as their risk aversion, which represents whether or not they engage in criminal behavior. We aim to look at how the social networks are formed and if different social networks have different outcomes in terms of the frequency of corrupt behavior.

The importance of understanding corruption in a dynamic way is that law-enforcement entities can experiment with different ways of targeting corruption. Ultimately, models and experiments that can be done for corruption detection, as well as understanding patterns for prevention, can help with the social inequities that arise from misdistribution of funds and abuses of power. The following sections explore the different aspects of the model. In Sect. 2, we provide background and describe the area of study in more detail. Section 3 is an overview of the model with Sect. 3.1 exploring generation of networks, Sect. 3.2 exploring agents, and Sect. 3.3 showing the model initialization and the interactions between agents. Section 4 shows our results, and Sect. 5 is our conclusion with some thoughts on areas of further work.

2 Background

Operacao Lava Jato, or "operation car wash", was an investigation by the Federal Police of Brazil in the Southern part of Brazil [7]. The scale of corruption ranged from elected officials to private companies, as well as cartels and off-shore banks, which has provoked a public uproar, protests, and political upheavals in Brazil [11]. The initial investigation, which gave origin to the title of the operation was set to focus on *doleiros*, black market money dealers who used petrol stations and car washes

to launder the profits of crime. The interactions of individual agents with criminal activity from the "bottom-up" [6] gave rise to the emergence of one of the largest, if not the largest, corruption scandals in Latin America, if not the world [1].

Several other investigations related to *Lava Jato* came to the forefront, including the beef and poultry industries. Sanctions were placed by other countries on imports from Brazil, which led to a cycle of effect that affected consumer options and choices in homes around the world, which in return has resulted in an impact on the Brazilian Stock Market [2]. Sergio Moro, the federal judge responsible for the persecution, states that these types of crimes "represent a challenge for law-enforcement agencies" and that the level of sophistication of the people engaged in corruption and the lack of tools and resources to deal with these types of crimes have allowed for years of corruption to go undetected and unpunished for decades [19].

There has been controversy surrounding the methods used for investigation and non-traditional means of persecuting these crimes [16]. Previous scandals have also come to light, such as the *Mensalao* (also known as Case 470) in 2002, but the results of previous investigations have not resulted in the scale and outcomes of the *Lava Jato* investigation. The reason has been a delay in the justice system which makes it harder for people of power to be tried [19]. The defendants from this case have been politicians, CEOs, judges, cartels, and large corporations. Politicians who engage in corruption through their individual actions affect policy and political outcomes. "Corruption can change the alignment of citizen and leader preference" [3].

Corruption in Brazil has its origins in the history and political narrative of the country [16]. With the *Lava Jato*, there have been changes to the justice system, focusing on speedier trials, repatriation, as well as quicker arrests of those involved in corruption and crimes, which has also proved controversial [5]. However, the openness and publicity involved in these entails, as well as the open resources online for denouncing crime, have allowed us to have more data in terms of who the corrupt people are, as well as their network and how all the people connect to one another [1].

The major players in the *Lava Jato* case can be divided into three main categories: "criminal" which covers those agents who represent full-time criminals, "public" which includes public employees of any government institution, and "private" referring to those people who carry out licit commercial activities outside the public sphere. The distribution of agents is as follows: criminal = 98, political = 193, and private = 608 agents (so far) in each respective category [8]. It is important to note that in this case, the larger number of private agents may have contributed to the agents being caught. These classifications allow us to better understand the organizational and institutional role of each agent, in addition to the particular characteristics of the structure of the network [8]. The individual choices from employees in the private industry acting from the "bottom-up" have created large-scale corruption. The agent's individual choices in the *Lava Jato* case and the decision to engage with crimes or not help our understanding of human psychology and sociology [3].

Why and how people choose to commit or not commit crimes and engage in corruption is a complex issue, but many psychologists [18, 22] have looked at group behavior and power dynamics to better understand how people act and react within

a social context, as well as tipping points of when people decide to adopt a behavior or not [4]. The "infectious" nature of how behaviors are adopted in social settings is the aim of this paper and we hope to expand some of these thoughts to examine the corruption of individuals involved in the *Lava Jato* scheme through modeling this behavior.

The reach of corruption has been extensive, and the effects on political parties and campaigns have also tainted Brazilian politics and the trust of the people in the government. It would be impossible to reflect all of these different aspects in a model, but the focus of this paper is to understand the patterns of the social network and the influence [4] of people on their peers, and whether they are more or less likely to commit crimes.

3 The Model

The model we developed recreates how corrupt behavior spreads in a network structure and how the cumulative effects of corruption cases that come to the public eye behave over time. We based many of our assumptions and the model on work being done so far to understand the key players and their networks, from data gathered from the trials [8].

Past models have looked at corruption within government with an analysis of corruption and the effects on the effectiveness of rule of law [12]. Others have examined the micro- and macro-level interactions of corruption with mathematical models of corruption and social evolution [15], agent-based models within a game theory context [13], as well as corruption risks in contracting markets [23]. Our model hopes to look at the spread of corruption, specifically through a lens of the influence of corruption on the agents within the *Lava Jato* network. In the following subsections, we model the spread of corruption in a network structure and describe in more detail the different networks that we will create and the processes underlying their changes. We then describe what our model does and what the outcomes tell us about the *Lava Jato* corruption scheme.

3.1 Generation of Networks

For this model, we used custom code, since we had to create networks where the nodes are the agents in our model. Examples of the networks chosen are in Fig. 1. The first network is a Random Network [9]. In an undirected network with n nodes, for every one of the possible $\binom{n}{2}$ links, there is a probability p of creating that edge. The density of this network approaches p. We then look at the *small-world network* [24]. We started with a ring network where every node has degree 4. Then, for every link (i, j), we break it and create another one (i, k) with probability p where k is a node chosen at random.

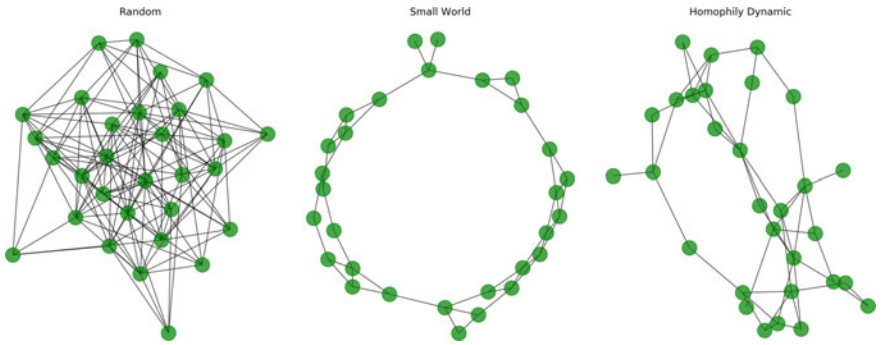


Fig. 1 Three network types used in our model

The network on the right of Fig. 1 is a Homophily Dynamic network, which means that starting from the same ring network (with degree 4) as the small-world networks, we created new links based on attributes of the nodes, where every node is connected to the adjacent 2 in the "left" and in the "right" relative to its position in the ring [20]. It is assumed that every node has some number of attributes $f \in \{1, 2, \ldots, F\}$. These attributes can take q different values, where $2 \leq q \leq Q$. As F is increased, the social system becomes more complex, and as Q increases, the individuals become more heterogeneous. At every time step, the edges in the network may change based on the following three processes.

- Global Attachment (GA). A node selects one of its F attributes and chooses at random from the subset of nodes with the same q value in that attribute. They establish a link.
- Local Attachment (LA). A node i with $k_i > 0$ chooses one of its neighbors, j, who in turn, chooses one of its neighbors, k such that $f_j = f_k$. Then, if i and k are not connected, a node is created.
- Link Deletion (LD). Every element in the edge set is removed with probability p_d.

3.2 Agents

For this model, we only used one attribute of the nodes as the homophily variable: the political party of every node, chosen at random at the initialization of the model with a uniform distribution between A, B, and C. The reason for this is that political parties and the networks constructed around these parties play a big role in how the *Lava Jato* scandal developed [14]. Allowing agents to connect with other agents who have similar political attributes (in a simplified political system) provides a way to aggregate agents and their social networks. It also reflects Brazil's clientelistic model where there is minimum disruption of the exchange network due to the rigid

structures of the networks [17]. This allows for the parties and thus the networks to generally keep the same configuration over time.

In this artificial social system which we base on the categories of agents from the *Lava Jato* case, agents can have one of three possible roles [8, 21]: they can belong to the private, public, or criminal sector. We chose the proportion of agents in each of the roles as 0.7, 0.19, and 0.11, respectively, as these proportions are based on empirical data gathered from the numbers of individuals caught from each sector [8].

During the model, we let the corrupt status of every agent be a random variable resulting in a Bernoulli trial, with two possible outcomes, **corrupt** or **not corrupt**, with probability c of being corrupt, which means agents engage in an illicit transaction where they get a payoff.

Since every agent can be, at different parts of the model, in both ends of an illicit transaction, we created two attributes for this. The first one is a **transaction** size value when the agents propose the illegal exchange. This is the result of a random discrete variable with uniform distribution with domain [100, 1000]. What this variable represents is how big the transaction is, in terms of an abstract currency unit. The second attribute puts the agent at the other end and corresponds to the **risk aversion**, which is a random variable that draws from different distributions depending on the argument passed. The possible arguments are:

- `u`. Uniform risk aversion. Uniform distribution between 0 and 1.
- `l`. Low risk aversion. $\beta(2, 5)$ distribution. Biased to the left.
- `m`. Medium risk aversion. $\beta(2, 2)$ distribution. Symmetrical centered around 0.5.
- `h`. High risk aversion. $\beta(5, 2)$ distribution. Biased to the right.

Finally, agents keep track of their **political party**, a random variable that chooses between A, B, or C, which is kept at one attribute to better capture the behavior of this network using the more salient attribute of this case [17], which was abstracted due to the many different political parties in Brazil, whether they are under **investigation** for corruption or not, and the total amount of **payoff** obtained from illegal transaction since the first time step.

3.3 Initialization of the Model and Agent Interactions

When the model initializes, we create a list of the agents, with the proportions described in Sect. 3.2. Every item in the list is a node in a network using either a random, small-world, or homophily dynamic structure. At the beginning of every time step, we update the network using a random rewiring or a homophily-based one (with the GA, LA, and LD processes in sequential order).

Every time step is a week w, which is chosen due to the nature of the transitions carried out and captures the frequency of the transactions based on all the interacting agents. Before any interactions, we choose a random number of transaction attempts τ. For every one of the random transactions, we choose a random node i that will

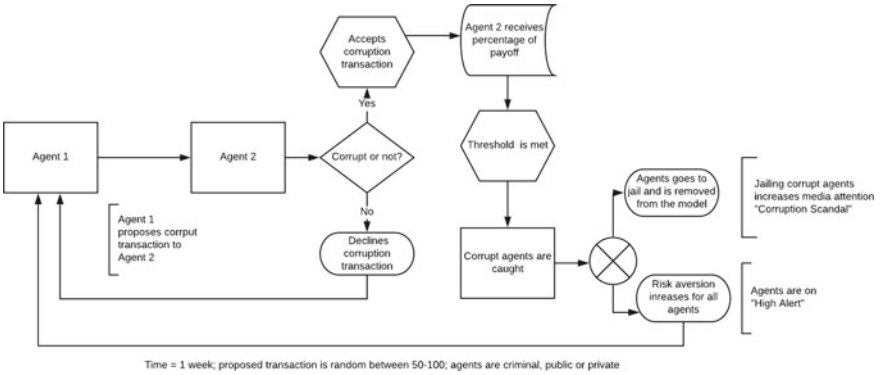


Fig. 2 Overview of the model over the span of one week. The model is run for a year. Agent 1 and Agent 2 can be criminal, private, or public

Table 1 Baseline parameters of the model

	Initial values
Number of nodes	500
Number of transaction attempts per week	[20, 30]
Amount of payoff until agent gets investigated	100
Probability of getting caught	0.1
Proportion of corrupt agents	0.2
Distribution of risk aversion	'u'
Network update process	GA, LA and LD

try to forge an illegal transaction with node j. In case j is corrupt, the transaction takes place and j gets as a payoff of the transaction value (transaction size). In case j is not corrupt, it can accept the exchange with probability (1 risk aversion). If the transaction is successful, j changes its status to corrupt and updates its payoff.

After all the transaction attempts have been made, the model checks the current value of the payoff for every agent. If it exceeds some value p^*, the agent gets caught with a probability p_c. When this happens, the caught agent goes to jail (is removed from the agent set and from the network) and all of its neighbors change their status to "under investigation" and increase their risk aversion (Fig. 2).

We replace every agent removed from the model, keeping the same proportion between criminal, public, and private sectors. Since at the beginning of each week w we will update the network, these new agents get connected with another node depending on the specific network structure for the update.

In Table 1, we established a set of baseline parameters with enough variability that will produce noticeable results, against which we will compare different parameter values.

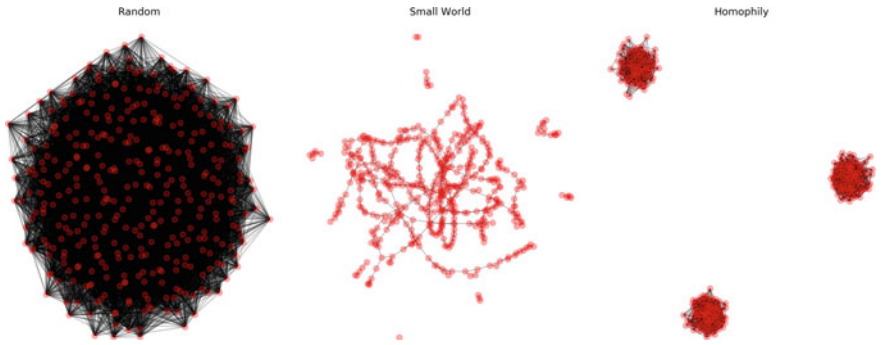


Fig. 3 Networks produced with baseline parameters at $w = 52$

The parameters are based on the data we have so far. In some cases, we have made assumptions based on our understanding of the percentages of people who actually were caught in the *Lava Jato* scandal. As it is difficult to know the actual probability of getting caught when committing a crime or how many people committed crimes that have gone undetected, we have done our best to approximate the values to realistic values that would map to the real world.

Based on these parameters, we created three networks, with an initial configuration of random, small-world, and homophily network. The model is run for 52 weeks to understand the cycle of different transitions that occur within the private, public, and criminal sectors, and with an update process that uses GA, LA, and LD [20] each week. In the end, the resulting networks are in Fig. 3. The random network looks like a hairball and depicts an unlikely scenario where every node has random links. Given that corruption in the *Lava Jato* case is coordinated and not random, this scenario does not apply, so we move on to the next structure.

We ran the model with different values for the probability of agents being corrupt and different proportions of agents from the criminal, public, and private categories. Because a large proportion of the agents from the corruption scandal belonged to the private category, we have varied the "probability of getting caught" and "proportion of corrupt agents" across all three categories. This is due to the reality of the situation where despite having career "criminals", only a small number were actually caught or were captured in the data we have of the scandal [8].

The network in the middle, the small-world, has some characteristics that could be useful in the *Lava Jato* case. For example, two nodes at random have the shortest path L that is proportional to $\log n$, where n is the number of nodes. In our example, this means that one node can easily reach any other node in the network.

The three distinct subgroups in the plot are the three political parties used in our model. Besides the differences in the shape of the three networks, they also have different values in some of their attributes. The detail of these differences is in Table 2.

Table 2 Differences in network measurements at $w = 52$, baseline parameters

	Random	Small World	Homophily
Number of nodes	330	491	306
Number of links	9525	741	2467
Average Degree	57.7273	3.0183	16.1242
Average Clustering	0.1741	0.3492	0.1644
Average Path Length	1.8246	6.2859	1.9136

4 Results

We then ran the model for 208 weeks to understand the number of illegal transitions and see how many corrupt agents are caught after the end of 52 weeks. There are a few things to note. First, the number of illegal transactions depicted in Fig. 4 (top) decreases in all three networks. Since the nodes react to their neighbors and general corruption decreases as a result of being on high alert, there are fewer proposals for corrupt transactions. Agents will also propose illegal transactions to fewer people at every time step since jailing corrupt agents decreases the corruption in the system.

All plots show that regardless of network type, agents will decrease their activity by observing their neighbors. In the arbitrary threshold we set, there is also evidence to support that the homophily network will lead to more corruption scandals than other networks. The homophily network in this model shows that nodes tend to influence others in their political party, so that when the corrupt agents are caught, it is easier to uncover the corruption acts of their neighbors due to their proximity and similarity of attributes.

The second aspect is the number of corrupt agents that get caught and are sent to jail. Figure 4 (lower) shows these results. The reason is that in a given week, if the threshold of people caught is met and agents and are sent to jail for corruption, the whole social system will be on high alert with additional attention of the media (for example) and could represent the beginnings of what could end in a corruption scandal. Our results show the behavior of the homophily networks, because they capture more closely how people create and maintain political contacts. Figure 5 shows the effects of different distributions of risk aversion for the small-world network.

In the baseline model, at every time step (every week), the network gets updated by a sequence of the GA, LA, and LD processes. We tried running the three models in Fig. 4 by changing the network update process while all the other conditions remained the same. Instead of the sequence GA, LA, and LD, we updated the networks through a random process, where at every time step, we broke all links and created new ones with probability p.

The results in Fig. 4, as expected, show little difference between the three types of networks. All three are affected by the risk aversion increase at more or less the same rate because every week new networks are created.

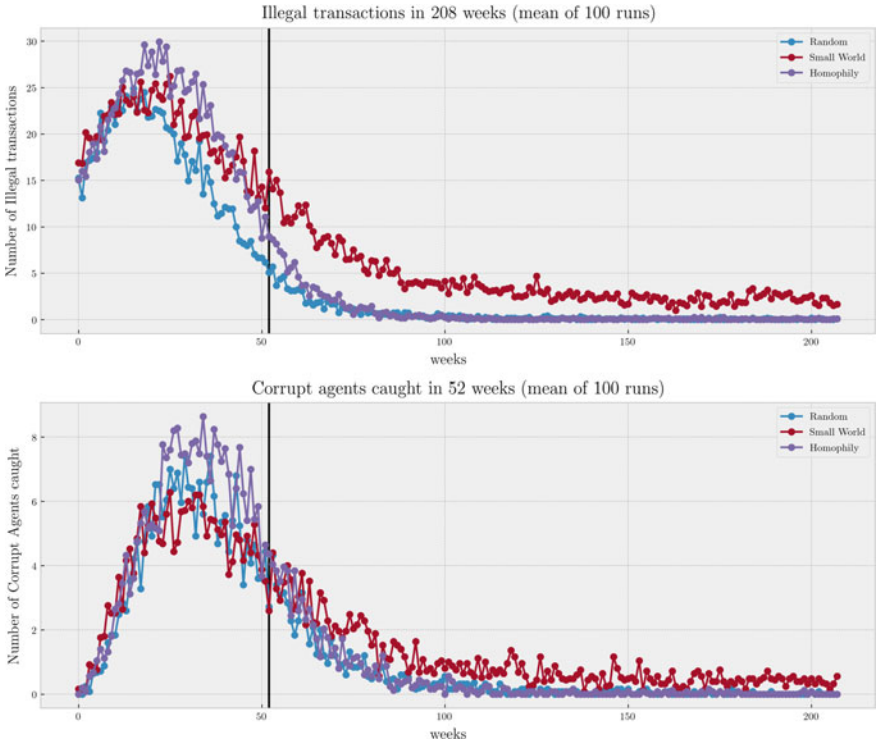


Fig. 4 Differences in network measurements, with variable initial configuration, from illegal transactions over 208 weeks (top), with a mean of 100 runs. Corrupt agents over 52 weeks, with a mean of 100 runs (lower). The results are for the random, small-world, and homophily networks

The final test we performed on the parameters was to see how different proportions of corrupt agents affect the outcome of the model. We used the number of illegal transactions per week as our values, with 1, 20, 40, and 60% of corrupt agents. The results in Fig. 6 reproduce what was found by modifying the risk aversion distribution. However, in the case where the proportion of corrupt agents is 60%, there is a spike in the number of illegal transactions, regardless of how much time has passed, and corruption also appears to decrease at the same rate.

5 Conclusion

This is a theoretical model based on the details we have of the *Lava Jato* investigation thus far, and it does show similar behavior to the known networks that have already been revealed [8]. Despite what we know of the investigation, there is limited data for us to thoroughly validate this model. Our results are robust and conceptually represent

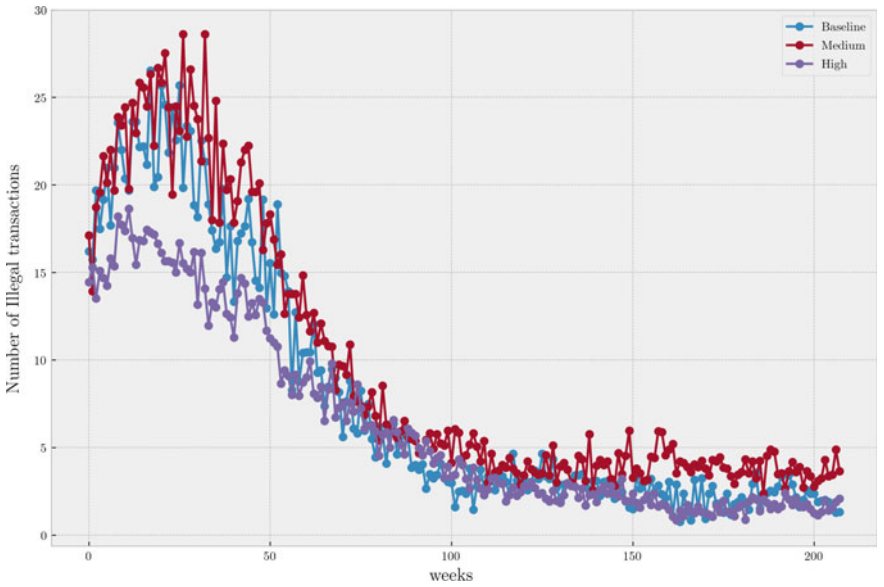


Fig. 5 Small-world initial configuration, with different initial risk aversion values

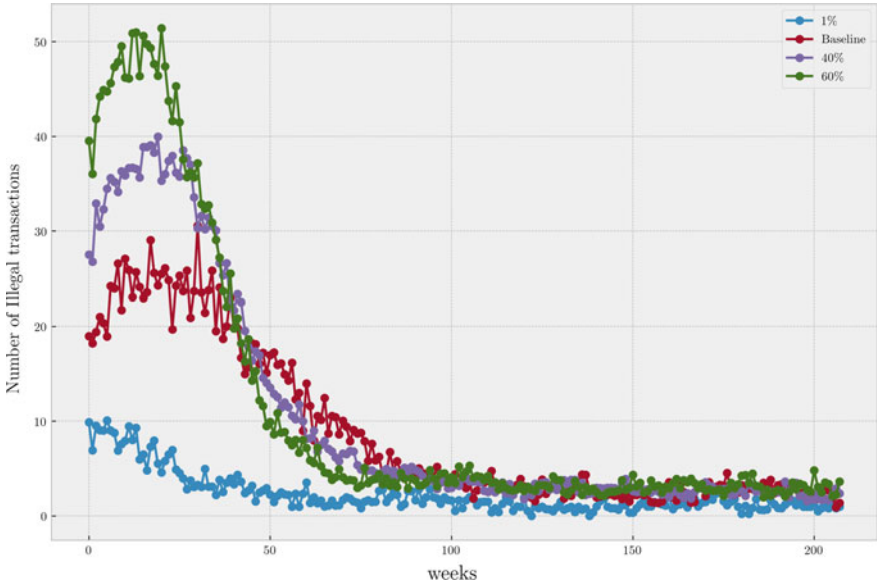


Fig. 6 Small-world initial configuration, with different initial risk aversion values

the way in which corruption occurs and has an effect on others within a network, but real data could provide more information about the structures of the networks and the motivations and decision-making processes of the individuals engaged in corruption. However, even with additional data, we would at most have the illegal transactions that occurred, but not the attempted ones, which may be another benefit of simulating these types of behaviors to understand corruption.

As we mentioned earlier, a closer look at the linkages between the major players in the *Lava Jato* case allows us to better understand the patterns of corruption in this case and how individual choices from the bottom-up created large-scale illegal activities. We learned from the results of our model that small-world networks are the ones where the illegal transactions are low and stable, where the persistence of illegal activities is ongoing, even if agents are linking with new separate nodes. Therefore, what happened in Brazil was not a small-world phenomenon, and it also was not random, as corruption on such a grand scale requires multiple interactions with the same people over a period of time [3]. This is further evidence for believing that there was a homophily process in the formation of the corruption network in the Lava Jato case. As more arrests and progress are made in the Lava Jato investigation, further work can be done to better understand how the homophily developed within this corruption scandal.

While our results are consistent, further experimentation with empirical data and different model assumptions may produce variations of these results which would help with the validation of this model. Gaining a better understanding of the linkages between the major players can also help with targeted corruption detection and prevention. In further works, we can experiment with targeting links and targeting nodes to understand what types of strategies can be effective in cases of large-scale corruption, as well as experimenting with removing nodes from the network in more strategic ways, targeting specific types of agents with attributes such as highest degree and centrality, for example, to understand the underlying influence and relationship between the nodes. By building an agent-based model to look at the frequencies of agents committing crimes and engaging in corruption, we can explore the different thresholds and understand what types of effects result in higher proportions of agents committing crimes, as well as understanding if some types of social networks are more or less vulnerable to crime detection.

Overall this model serves as a theoretical representation of corruption within the case of the *Lava Jato* scandal and understanding these networks and effects can help enhance our understanding of individual behaviors with large-scale impacts. Modeling can show clearer patterns and underlying networks, which can aid in a better understanding of the structure of large-scale crime and help to combat corruption in the early stages. By understanding this specific example of corruption, we can apply similar methods to other corruption scandals which may reveal even more about human behavior in regard to corruption, influence, and the likelihood of people following the footsteps of those in their networks when it comes to committing crimes.

References

1. Angelico, F.: Brazil: Open data just made investigating corruption easier. https://www.transparency.org/news/feature/brazil_open_data_just_made_investigating_corruption_easier (2017)
2. Araújo, E.C.C.D., Rodrigues, V.R.D.S., Monte-mor, D.S., Correia, R.D.: Corrupção e valor de mercado: Os efeitos da operação lava jato sobre o mercado de ações no brasil. REVISTA CATARINENSE DA CIêNCIA CONTÁBIL **17**(51), nil (2018). https://doi.org/10.16930/2237-7662/rccc.v17n51.2626
3. Besley, T.: Principled Agents?: The Political Economy of Good Government. Oxford University Press, Oxford, New York (2006)
4. Centola, D., Willer, R., Macy, M.: The emperor's dilemma: a computational model of self-enforcing norms. Am. J. Soc. **110**(4), 1009–1040 (2005). https://doi.org/10.1086/427321
5. Economist, T.: Brazil's biggest anti-corruption investigation is at a turning point. https://www.economist.com/the-americas/2019/04/11/brazils-biggest-anti-corruption-investigation-is-at-a-turning-point (2019)
6. Epstein, J., Axtell, R.: Growing Artificial Societies: Social Science from the Bottom Up. Brookings Institution Press, Washington, DC (1996)
7. Federal, P.: Operacão lava jato. http://www.pf.gov.br/imprensa/lava-jato (2018)
8. Garay-Salamanca, L. J., Salcedo-Albarán, E., Macías, G.: Macro-corruption and the lava-jato case: a criminal networks analysis. https://smallwarsjournal.com/jrnl/art/macro-corruption-and-lava-jato-case-criminal-networks-analysis (2018)
9. Gilbert, E. N.: Random graphs. Ann. Math. Stat. **30**(4), 1141–1144 (1959). https://doi.org/10.1214/aoms/1177706098
10. Gilbert, N., Troitzsch, K.: Simulation for the Social Scientist. Open University Press, Maidenhead, England New York, NY (2005)
11. Goldstein, D.M., Drybread, K.: The social life of corruption in Latin America. Cult. Theory Crit. **59**(4), 299–311 (2018). https://doi.org/10.1080/14735784.2018.1531816
12. Guerrero, O.A., Castañeda, G.: Does better governance guarantee less corruption? evidence of loss in effectiveness of the rule of law. CoRR. http://arxiv.org/abs/1902.00428v1 (2019)
13. Hammond, R.: Endogenous transition dynamics in corruption: an agent-based computer model. Tech. rep., Center on Social and Economic Dynamics Working Paper No. 19 (2000)
14. Katz, A.S.: Making Brazil work? Brazilian coalitional presidentialism at 30 and its post-lava jato prospects. Revista de Investigações Constitucionais **5**(3), 77 (2018). https://doi.org/10.5380/rinc.v5i3.60965
15. Lee, J.H., Iwasa, Y., Dieckmann, U., Sigmund, K.: Social evolution leads to persistent corruption. Proc. Nat. Acad. Sci. **116**(27), 13,276–13,281 (2019). https://doi.org/10.1073/pnas.1900078116
16. Lopes, A.S.: The hermeneutics of Brazil's impeachment. Harvard Int. Rev. **37**(4), 11–14 (2016)
17. Lyne, M.M.: Proffering pork: how party leaders build party reputations in Brazil. Am. J. Polit. Sci. **52**(2), 290–303 (2008). https://doi.org/10.1111/j.1540-5907.2008.00313.x
18. Milgram, S.: Behavioral study of obedience. J. Abnorm. Soc. Psychol. **67**(4), 371–378 (1963). https://doi.org/10.1037/h0040525
19. Moro, S.F.: Preventing systemic corruption in brazil. Daedalus **147**(3), 157–168 (2018). https://doi.org/10.1162/daed_a_00508
20. Murase, Y., Jo, H.H., Török, J., Kertész, J., Kaski, K.: Structural transition in social networks: the role of homophily. Sci. Rep. **9**(1), 4310 (2019). https://doi.org/10.1038/s41598-019-40990-z
21. Parsons, T.: The Social System, new edn. Routledge Sociology Classics, Routledge (1991)
22. Raven, B.H., French, J.R.P.: Group support, legitimate power, and social influence1. J. Person. **26**(3), 400–409 (1958). https://doi.org/10.1111/j.1467-6494.1958.tb01595.x
23. Wachs, J., Fazekas, M., Kertész, J.: Corruption risk in contracting markets: a network science perspective. CoRR. http://arxiv.org/abs/1909.08664v1 (2019)

24. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998). https://doi.org/10.1038/30918
25. Weyland, K.G.: The politics of corruption in Latin America. J. Democracy **9**(2), 108–121 (1998). https://doi.org/10.1353/jod.1998.0034

Resistance of Communities Against Disinformation

Amirarsalan Rajabi, Seyyedmilad Talebzadehhosseini, and Ivan Garibay

Abstract The spread of disinformation is considered a big threat to societies and has recently received unprecedented attention. In this paper, we propose an agent-based model to simulate the dissemination of a conspiracy in a population. The model is able to compare the resistance of different network structures against the activity of conspirators. Results show that *connectedness* of network structure and *centrality* of conspirators are of crucial importance in preventing conspiracies from becoming widespread.

1 Introduction

We define conspiratorial thinking as a belief held by an individual or a group of individuals, while there is enough evidence and information to undermine or totally refute the belief. If some conspiracy theories get popular support, it may cause serious concerns. This is particularly true of conspiracies over scientific and medical issues where conspiracy theories can result in rejection of the scientific method [20].

Several underlying reasons have been proposed to explain the existence of conspiracy theories. According to Barkun, conspiratorial thinking exhibits three characteristics: Firstly, nothing happens by accident. Secondly, things are not as they seem on the surface. And thirdly, things are highly connected [6]. All three characteristics mentioned by Barkun refer to a special cognitive function of the conspiracy theorist. Indeed, a great deal of literature on conspiracy theory associates conspiratorial thinking with a special and different heuristic of the conspiracy theorist. On the other hand, Sunstein and Vermeule claim that many of those who hold conspiracy theories

A. Rajabi · S. Talebzadehhosseini · I. Garibay (✉)
Complex Adaptive Systems Lab (CASL), University of Central Florida, Orlando 32816, USA
e-mail: igaribay@ucf.edu

A. Rajabi
e-mail: amirarsalan@knights.ucf.edu

S. Talebzadehhosseini
e-mail: talebzadeh@knights.ucf.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_3

do so not as a result of a mental illness of any kind, or of simple irrationality, but as a result of crippled epistemology (knowing very few things, which are indeed wrong) [22]. For most of what they believe and know, human beings lack direct information; they must rely on what others say and think. Hardin argues many people suffer from a crippled epistemology, meaning they only get their information from a few incorrect sources [10]. Crippled epistemology usually takes place in echo chambers. Echo chambers are communities in which individuals merely communicate with each other and rarely seek information from entities outside the community. The advent of social media platforms has resulted in the rise of echo chambers [3]. Echo chambers play an important role in political and social polarization [5]. The negative effects of polarization in social networks are studied in [16]. Bauman states that individuals who are embedded in isolated groups or small, self-enclosed networks who are thus exposed only to skewed information will more often hold conspiracy theories that are justified, relative to their limited informational environment [7]. The study of the dynamics by which the echo chambers form can therefore shed light on the mechanisms by which conspiratorial thinking forms and thrives in a community.

Traditional studies of conspiratorial thinking assume that conspiratorial thinking extinguishes in large network structure [17] and that conspiratorial ideation is because of flawed reasoning and biased heuristics [6]. This study challenges both of these claims. Contrary to a large body of literature on the conspiracy theory that studies the cognitive function of isolated individuals [22], this paper takes into account systemic belief dissemination as a result of interactions between individuals.

2 Opinion Dynamics

Opinion formation is a complex process that is formed by the interaction of multiple underlying elements. People tend to form their opinion on a wide variety of subjects through the process of learning. "Social learning" is a term referring to the process of learning through the communication of individuals with each other, their own experience, and their observations of others' experiences, media sources, propaganda, and indoctrination from political leaders and the state [1]. In this paper, we refer to models of opinion dynamics as mathematical models that aim at capturing the dynamics of social learning, opinion spreading, collective decision making, and so on from a mathematical point of view.

Models of opinion dynamics can be divided into two categories: Bayesian models of opinion dynamics and non-Bayesian models of opinion dynamics. Bayesian models rely on the Bayes rule [9]. These kinds of models assume that an individual (agent) is Bayesian rational and update their belief optimally with respect to the Bayes rule, given an underlying model of the world. One problem with Bayesian models of opinion dynamics is that these models make a lot of assumptions. One demanding assumption of these models is that an agent must have a reliable prior assumption about the world, an assumption that might be unrealistic in many cases. Additionally, it is assumed that they then go on and update their prior beliefs based

on the new information that they get from others. Bayesian models also put too much structure on updating by ruling out "zero probability events" [1]. The aforementioned problems make Bayesian models unfavorable to be incorporated in our study of the dissemination of conspiracy theories.

Non-Bayesian approaches and models on the other hand try to avoid some of these problems. Non-Bayesian approaches are believed to be more effective in modeling belief manipulation and the spread of disinformation. The simplest of these models start by specifying rules of thumb [1]. Several different non-Bayesian models exist. Classical models of interacting particle systems which in inspired by statistical mechanics inspires many of these models (see for example [11, 12, 15, 18, 19]).

One noteworthy non-Bayesian model is DeGroot (1974) [13]. In this model, a set of interacting agents start by an initial belief about an underlying state held by each agent and exchange information about their beliefs with their neighbors and update their beliefs at discrete time instances, with respect to a weight matrix that represents the social network structure of interactions. This model captures the "imitation" aspect of non-Bayesian models. While notably innovative, this model suffers from duplication of information [2], meaning that agents in this model might interact endlessly with their neighbors that hold unchanging opinions in each timestep.

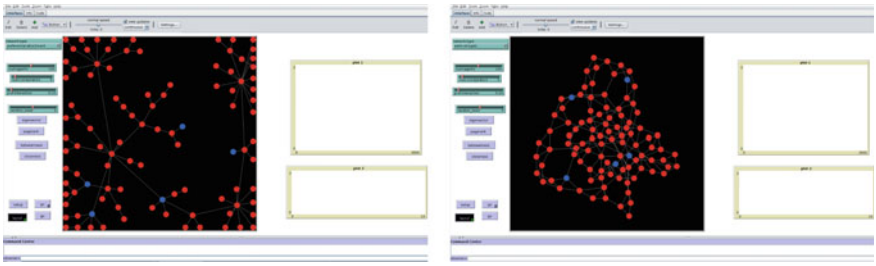
3 Model of Conspirators

Our agent-based model is inspired by the variation of Acemoglu et al. on the DeGroot model [2]. In our model, two types of agents exist. The majority of agents are *susceptible* agents and a minority of agents exist that are called *conspirators*. Conspirators deliberately disseminate false information to susceptibles.

Let's consider a conceptual underlying state of the world and call it Θ. We assume that the true value of Θ is 1, and the discussion between agents in the model is on the true value of Θ. $x_i^s(k)$ and $x_i^c(k)$ represent the opinion of susceptible agent i at time k and the opinion of conspirator agent i at time k, respectively. At first, each susceptible agent holds an initial belief about the underlying state. N shows the total number of agents in the model. The initial belief of each susceptible agent is a randomly generated float number between zero and two, the initial belief of each conspirator agent is 0, and the average of initial beliefs of all susceptible agents is very close to one:

$$x_i^s(0) \in [0, 2] \qquad x_i^c(0) = 0 \qquad \frac{1}{N}\sum_i x_i^s(0) \approx 1 \tag{1}$$

Hence although at first, each susceptible agent has their own initial belief on the underlying state Θ, there is a consensus on this underlying state between the susceptible agents. On the other hand, the initial belief of conspirator agents is 0 and remains 0 during the simulation, irrespective of their interactions. These could be thought of as individuals, entities, media, or propaganda outlets that deliberately and constantly disseminate false information throughout the population.



(a) Scale Free (Barabasi-Albert) (b) Small World (Watts-Strogatz)

Fig. 1 NetLogo environment. **a** Shows a network generated with Barabasi-Albert algorithm. **b** Shows a network generated by Watts-Strogat algorithm

There are two main differences between our proposed model and the work of [2]: (1) Agents are not able to freely communicate directly with any other agents. Rather, each agent is only able to communicate with the agents that she is connected to by a link. This means that a network structure regulates the communication in the population. (2) In the variation of [2] on the DeGroot model, each agent meets and communicates with other agents at instances defined by a rate one Poisson process independent of other agents. In our model,, agents are connected to each other by undirected links, and together, they form a network in each timestep of the simulation, an agent will choose another agent to which it is connected and may or may not communicate with them. In this model, the capacity of an agent for communication in each timestep is the number of her links. For example, an agent with 5 links will communicate with 5 or less other agents in each timestep.

Following the interaction of two agents i and j, there is a potential exchange of information between them with probability $p_{interaction}$. Agents update their beliefs according to one of the following possibilities:

If i and j are both susceptibles:

$$\begin{cases} x_i(k+1) = x_j(k+1) = \frac{1}{2}[x_i(k) + x_j(k)] & \text{with probability p} \\ x_i(k+1) = x_i(k) \ \& \ x_j(k+1) = x_j(k) & \text{with probability } 1-\text{p} \end{cases}$$

If i is susceptible and j is conspirator:

$$\begin{cases} x_i(k+1) = \frac{x_i(k)+0}{2} & \text{with probability p} \\ x_i(k+1) = x_i(k) & \text{with probability } 1-\text{p} \\ x_j(k+1) = x_j(k) = 0 & \text{with probability 1} \end{cases}$$

If i and j are both conspirators: no opinion sharing.

The underlying network structure of the model plays an important role in determining the behavior of the model, and it is itself determined by the algorithm by which the model forms the network. It determines which agents are connected to each other, how easily the information would propagate throughout the network, and so forth.

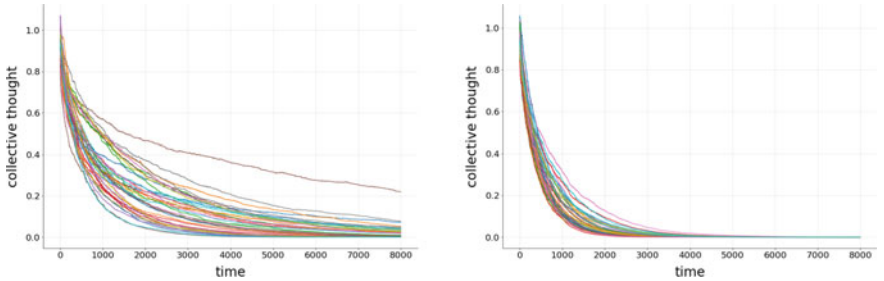
Watts and Strogatz proposed a model to capture the small world, high clustering, and low average path properties of real complex networks [24]. On the other hand, Barabasi and Albert showed that power-law degree distribution is the property of many real-world networks and proposed an algorithm that can capture this phenomenon [4]. Indeed, it is believed that the majority of complex networks exhibit small world and scale-free properties [23]. In order to capture all of these essential properties, the model generates the network of agents using both Watts-Strogatz and Barabasi-Albert algorithms.

4 Results

We developed our agent-based model in NetLogo [22]. Figure 1 shows two instances of the model with Watts-Strogatz and Barabasi-Albert model. For our purposes, we defined a variable named collective-thought. This variable is simply the average of the belief of all susceptible agents and represents the collective belief of the population about the underlying state Θ. Collective-thought starts with a value close to one and always converges to zero (Fig. 2).

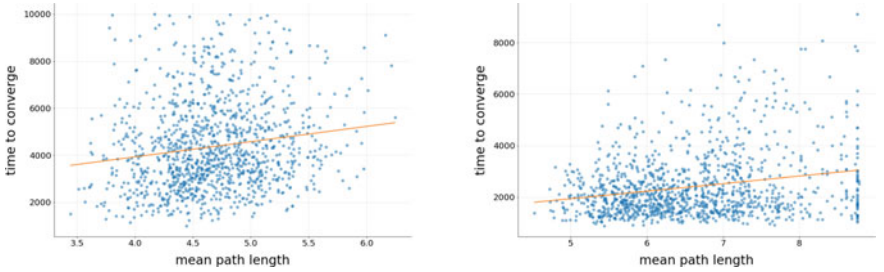
The desired population is one that resists the activity of conspirators. Numerous individuals and groups deliberately disseminate false information in a society. As we know the collective-thought of the model always converges to zero after some timesteps. We call this state the convergence. In order to quantify the resistance of a network against the activity of conspirators, we record the number of timesteps required for collective-thought of the network to converge. This number can be thought of as symbolically representing how long a society resists the effort of conspirators before it is nearly deceived into believing that the underlying state Θ is zero.

We argue that contrary to the conventional belief that conspiratorial beliefs are untenable with a larger network structure (e.g. [17]), a large network cannot ensure the eradication of conspiratorial beliefs. Indeed, the important aspect of a network structure that can ensure resistance against conspiracies is the *connectedness* of the network; that is, how easily an agent can send information through the network to any other agent. A network with high connectedness means that information propagates more easily and echo-chambers are less likely to form. We needed to come up with a benchmark that enables us to quantitatively compare different networks by our objective of connectedness. For this purpose, the model records the mean path length of the network in each instance of the simulation as follows [14]:



(a) Scale Free (Barabasi-Albert) (b) Small World (Watts-Strogatz)

Fig. 2 Figure shows the required timesteps required for the network's collective-thought to converge. Small world network tends to be deceived more quickly than a scale-free network



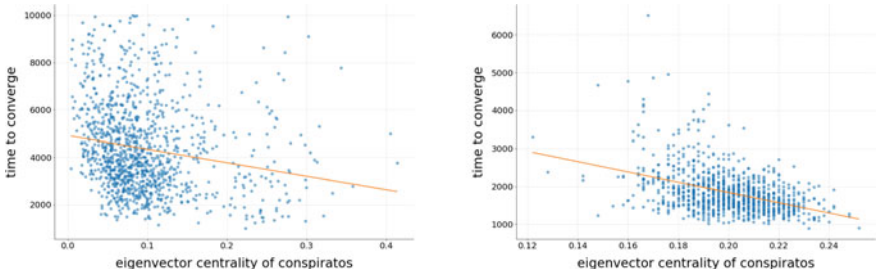
(a) Scale Free (Barabasi-Albert) (b) Small World (Watts-Strogatz)

Fig. 3 The relationship between mean path length and the required timesteps for a population's collective-thought to converge in **a** scale-free and **b** small world networks

$$\begin{aligned}\text{mean path length} &= \text{average shortest path between} \\ &\quad \text{all distinct pairs of nodes in network}\end{aligned} \tag{2}$$

The model was run 1000 times for each of the Watts-Strogatz and Barabasi-Albert networks. In each run, the required timesteps for the collective-thought to converge and the mean path length of the network were recorded. Figure 3 shows that there is a positive relationship between the mean path length of the network and required timesteps for the collective-thought to reach convergence. With the Barabasi-Albert algorithm, the Pearson correlation coefficient between the mean path length and the required timesteps was 0.18, with a p-value less than 0.00001 which shows the correlation is statistically significant. The corresponding coefficient for Watts-Strogatz networks was 0.39 with a p-value less than 0.00001, which shows this correlation is also significant. These results prove our hypothesis that the connectedness of a network improves its resistance against conspirators.

Besides the connectedness of a network, the structural position of conspirators in a network determines the reach of their disinformation and enhances their ability to disseminate their message through the whole population more effectively. To show



(a) Scale Free (Barabasi-Albert) (b) Small World (Watts-Strogatz)

Fig. 4 The relationship between sum of the eigenvector of all four conspirators and the required timesteps for a population's collective-thought to converge in **a** scale-free and **b** small world networks

that our model captures this phenomenon, we assigned to each agent in the network their eigenvector centrality. Eigenvector centrality is a measure of centrality in a network [21]. The more central a node is, the closer it is to all other nodes.

The model was run 1000 times for each of the Watts-Strogatz and Barabasi-Albert networks. For each run, the required timesteps for the collective belief to reach convergence and the sum of the eigenvector centrality of conspirator agents were recorded. Figure 4 shows the negative relationship between required timesteps for collective-thought to reach convergence and the sum of the eigenvector centrality of conspirators. With the Barabasi-Albert algorithm, the Pearson correlation coefficient between the mean path length and the required timesteps was -0.19, with a p-value less than 0.00001 which shows the correlation is statistically significant. The corresponding coefficient for Watts-Strogatz networks was -0.42 with a p-value less than 0.00001, which shows this correlation is also statistically significant.

5 Discussion

The model shows promising results in that its vulnerability is correlated with the connectedness of the network and the importance (eigenvector) of conspirators. Considering the minimum required time, the resistance of a population against conspiracies was imitated. The results show that a network that is built by the Watts-Strogatz algorithm was slightly more vulnerable to conspiracies than a network that is built by Barabasi-Albert algorithm (scale-free). The reason for this difference might be that a Watts-Strogatz network has a high local clustering, while holding a short average path length like random networks; therefore, the propagation of (dis)information is easier and faster in them. On the other hand, while noting that we used only four conspirators in our model and the total number of agents was 100, the probability that conspirators become a hub was low, and therefore their ability to propagate their conspiracy was slightly smaller than in Watts-Strogatz network. A probable misinterpretation here

might be that a less connected network is better against conspiracy theories. This is not correct because, in our model, only conspirators were acting to deceive the population and were trying to propagate their disinformation in the whole network. In fact, the results of this study can be interpreted in a content-agnostic manner. In reality, both sides of the discussion try to influence the network. Nevertheless, both networks show that the connectedness and the eigenvector centrality of conspirators in a network is highly correlated with the network's vulnerability to conspiracies.

This was the first step to make a framework for the computational study of conspiracy propagation. Further research must be conducted using other network formation algorithms. Other ranges of conspirators' ratio and population sizes should also be experimented. Another extension to this research might be to test the effect of multi-dimensional opinion-space in agents. Finally, other non-Bayesian models of opinion dynamics could be tested.

Acknowledgements This research was supported by DARPA program HR001117S0018.

References

1. Acemoglu, D., Ozdaglar, A.: Opinion dynamics and learning in social networks. Dyn. Games Appl. **1**(1), 3–49 (2011)
2. Acemoglu, D., Ozdaglar, A., ParandehGheibi, A.: Spread of (mis) information in social networks. Games Econ. Behav. **70**(2), 194–227 (2010)
3. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on facebook. Science **348**(6239), 1130–1132 (2015)
4. Barabási, A.-L., Bonabeau, E.: Scale-free networks. Sci. Am. **288**(5), 60–69 (2003)
5. Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R.: Tweeting from left to right: Is online political communication more than an echo chamber? Psychol. Sci. **26**(10), 1531–1542 (2015)
6. Barkun, M.: A Culture of Conspiracy: Apocalyptic Visions in Contemporary America, vol. 15. Univ of California Press (2013)
7. Bauman, Z.: Liquid Modernity. Wiley (2013)
8. Bavelas, A.: Communication patterns in task-oriented groups. J. Acoust. Soc. Am. **22**(6), 725–730 (1950)
9. Bayes, T.: An essay towards solving a problem in the doctrine of chances, 1763. MD Comput. Comput. Med. Pract. **8**(3), 157 (1991)
10. Breton, A., Galeotti, G., Salmon, P., Wintrobe, R.: Political Extremism and Rationality. Cambridge University Press (2002)
11. Castellano, C., Munoz, M.A., Pastor-Satorras, R.: Nonlinear q-voter model. Phys. Rev. E **80**(4), (2009)
12. Clifford, P., Sudbury, A.: A model for spatial conflict. Biometrika **60**(3), 581–588 (1973)
13. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**(345), 118–121 (1974)
14. Fronczak, A., Fronczak, P., Hołyst, J.A.: Average path length in random networks. Phys. Rev. E **70**(5), (2004)
15. Galam, S.: Minority opinion spreading in random geometry. Eur. Phys. J. B-Condensed Matter Complex Syst. **25**(4), 403–406 (2002)
16. Garibay, I., Mantzaris, A.V., Rajabi, A., Taylor, C.E.: Polarization in social media assists influencers to become more influential: analysis and two inoculation strategies. Sci. Rep. **9**(1), 1–9 (2019)

17. Grimes, D.R.: On the viability of conspiratorial beliefs. PloS one **11**(1), (2016)
18. Hegselmann, R., Krause, U., et al.: Opinion dynamics and bounded confidence models, analysis, and simulation. J. Artif. Soc. Soc. Simul. **5**, 3 (2002)
19. Latané, B.: The psychology of social impact. Am. Psychol. **36**(4), 343 (1981)
20. Lewandowsky, S., Gignac, G.E., Oberauer, K.: The role of conspiracist ideation and worldviews in predicting rejection of science. PloS one **8**(10), (2013)
21. Newman, M.E.: The mathematics of networks. New Palgrave Encycl. Econ. **2**, 1–12 (2008)
22. Sunstein, C.R., Vermeule, A.: Conspiracy theories: causes and cures. J. Polit. Philos. **17**(2), 202–227 (2009)
23. Wang, X.F., Chen, G.: Complex networks: small-world, scale-free and beyond. IEEE Circ. Syst. Mag. **3**(1), 6–20 (2003)
24. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world'networks. Nature **393**(6684), 440 (1998)

Assessing the Potential of Crowd-Shipping for Food Rescue Logistics Using Agent-Based Modeling

Anuj Mittal, Nilufer Oran Gibson, and Caroline C. Krejci

Abstract Food insecurity in the U.S. is a national concern. More than 30% of available food supply at the retail and consumer level in the U.S. is wasted each year and sent to landfills. This wholesome food could help address food insecurity problem by feeding the hungry families. Restaurants are a major source of food waste in the U.S. However, less than 5% of the more than 1 million restaurants in the U.S. currently donate food to the hungry. One of the biggest barriers to food donation from restaurants is transportation. Crowd-sourced transportation, known as "crowd-shipping," is a potential solution, in which individual volunteers use their personal vehicles to collect donated surplus food from restaurants and deliver it to food-insecure recipients. However, viability and effectiveness of such a program require that the number of participating restaurants and crowd-shippers are appropriately balanced and grow over time. This paper describes a conceptual agent-based model that was developed to examine the impact of initial restaurants and crowd-shipper participation levels on the number of meals delivered to food-insecure people over time. Preliminary experimental results demonstrate that increasing the initial participation levels does not necessarily lead to a uniformly better system performance over time—maintaining the right ratio of crowd-shippers to restaurants is critical to success.

A. Mittal
Department of Industrial Engineering Technology, School of Engineering, Dunwoody College of Technology, Minneapolis, MN, USA
e-mail: amittal@dunwoody.edu

N. Oran Gibson · C. C. Krejci (✉)
Department of Industrial, Manufacturing, & Systems Engineering, The University of Texas at Arlington, Arlington, TX, USA
e-mail: caroline.krejci@uta.edu

N. Oran Gibson
e-mail: nilufer.oran@mavs.uta.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_4

1 Introduction

Food insecurity in the U.S. is a serious humanitarian concern, with 15.6 million American families (12.3% of the U.S. population) lacking consistent access to sufficient nutritious food [1]. This number is particularly concerning, given that approximately 30–40% of the U.S. food supply is wasted [2]. To address the problem of food insecurity, the U.S. Department of Agriculture supports multiple initiatives, including food distribution programs, child nutrition programs, the Supplemental Nutrition Assistance Program (SNAP), and a special SNAP for women, infants, and children (WIC). However, 27% of individuals who are food-insecure may not qualify for federal assistance because their gross monthly income is higher than the maximum allowed income for eligibility to participate in these programs [3].

Another approach to reducing food insecurity is by rescuing food, in which surplus food that is still edible is collected and delivered to food-insecure people. Food rescue activities in the U.S. are typically performed by extra-governmental, community-based charitable programs, such as food banks and pantries [4], which rescue donated surplus food from farms, manufacturers, and retailers [5]. Restaurants are a major source of food waste in the U.S., generating 11.4 million tons each year, of which 390,000 tons could be recovered to yield 643 million meals [6]. However, less than 5% of the more than 1 million restaurants in the U.S. currently donate surplus food [7]. For restaurants, one of the biggest barriers to donating surplus food is logistics. Because the restaurant sector consists of many independent locations with relatively small volumes of rescuable food per location, efficient collection and distribution of restaurant food surplus is particularly challenging [8]. While the donors receive tax benefits for their donations, it is the non-profit food rescue organizations that are typically responsible for managing and covering the cost of transporting surplus food to food-insecure recipients [8].

Crowd-shipping offers a potential solution to the challenge of food rescue logistics. *Crowd-shipping* is defined as "an information connectivity enabled marketplace concept that matches supply and demand for logistics services with an undefined and external crowd that has free capacity with regards to time and/or space, participates on a voluntary basis, and is compensated accordingly" [9]. Examples of commercial crowd-shipping services include Uber Eats and DoorDash, in which customers use an online platform to order food, and a willing driver from a pool of available drivers (i.e., the crowd-shippers) delivers the order from restaurant (typically using his/her personal vehicle) for a predetermined price. The appeal of crowd-shipping lies in its ability to provide a low-cost delivery service with greater flexibility and shorter lead times than conventional transportation service providers.

Using crowd-shipping to rescue surplus food from restaurants and deliver it to food-insecure individuals is a relatively new idea. The non-profit organization Food Rescue US uses an app to recruit volunteer drivers ("Food Rescuers") to pick up surplus food from participating local donors and transport it to receiving agencies, such as soup kitchens and shelters. The service is currently operating in 17 U.S.

locations [10]. Goodr is a for-profit start-up company that uses third-party commercial crowd-shipping services to distribute surplus restaurant food throughout a large network of non-profit recipient organizations in Atlanta. Postmates, a commercial crowd-shipping company, has piloted an initiative in which it uses its own crowd-shippers to transport surplus food from participating restaurants in Los Angeles to local shelters [11].

While still in early stages of development, these programs suggest that food rescue operations can be enhanced through the use of crowd-shipping, providing a potential new avenue for addressing food insecurity. However, as food rescue organizations consider crowd-shipping as a logistics solution, decisions about system design and the best strategy for launching the program will be critical. In particular, the success of any crowd-shipping initiative requires acquiring a critical mass of customer and crowd-shipper participation. If there are too few participants, customers will be dissatisfied by unfilled service requests, crowd-shippers will have insufficient opportunities, and the initiative may never get off the ground [12]. Therefore, it is critical for a nascent crowd-shipping organization to build up its network as quickly as possible, which requires an understanding of the factors that influence potential customers' and crowd-shippers' willingness to participate. Modeling can help program designers to gain a better understanding of these factors, and then use this knowledge to increase the likelihood of system success, in terms of increasing the number of meals delivered to food-insecure people and number of restaurants that continue to participate in the program.

2 Models of Crowd-Shipping Systems

The objective of many existing models of crowd-shipping systems is to predict potential crowd-shippers' willingness to participate in a commercial crowd-shipping system. These are typically statistical models based on survey data and/or data collected from a crowd-shipping platform. A survey was conducted with potential crowd-shippers to develop a statistical model that predicts the likelihood of a crowd-shipper accepting a delivery assignment, given crowd-shipper demographic attributes, the time required to complete the delivery, and the amount of compensation [13]. A similar statistical analysis was done using survey data in another study [14]. Service request records from a crowd-shipping company were statistically analyzed to determine how to increase the odds of successfully recruiting a crowd-shipper to fulfill a given service request [15]. Survey data on potential crowd-shippers' preferences and social network characteristics was used to develop a TRANSIMS model that evaluates the potential of using customers' social network contacts for last-mile delivery [16].

Agent-based modeling (ABM) has also been used to evaluate crowd-shipping systems. ABM is well-suited for this application, allowing potential crowd-shippers to be realistically modeled as autonomous and heterogeneous individuals. ABM was used to explore the relationship between crowd-shippers' characteristics, properties

of crowd-shipping tasks, and system performance metrics [17]. Survey data from Amazon Mechanical Turk's service was used to validate the model. ABM was also used to investigate the use of crowd-shipping for last-mile parcel delivery for a case study in central London [18]. The purpose of the model was to understand how crowd-shippers' actions affect road usage. Likewise, ABM was used to simulate a crowd-sourced last-mile delivery service [19]. The model was used for capacity planning, observing the effects of varying the ratio of the number of crowd-shippers to the number of requested deliveries, as well as the maximum allowable crowd-shipper detour time, on system performance. In another study, survey data was used to inform the development of an ABM that examined the growth of a crowd-shipping system over time [20]. Experimental results show that the number of packages delivered is proportional to crowd-shipper flexibility, the monetary reward they receive for deliveries, and the initial number of crowd-shippers at the start of the simulation run.

This paper describes a conceptual ABM that was developed to provide a better understanding of how to design and launch a successful volunteer-based crowd-shipping system for food rescue. The model can help predict emergent properties of a volunteer-based crowd-shipping system (e.g., number of participants, number of meals delivered) that arise over time as a result of autonomous behaviors and interactions of crowd-shippers and restaurants. This conceptual model was developed from an earlier version of the model [21] and provides a basis for the future development of an agent-based decision-support tool that can assist non-profit and government organizations in initiating food rescue programs that leverage crowd-sourced transportation (part of model description included in this paper is a © [2019] by IEEE and reprinted with permission from [21]). The following sections provide a detailed description of the model, a set of preliminary experiments to demonstrate the model's performance, a discussion of the experimental results, and conclusion and plans for future model development.

3 Agent-Based Model

The ABM was developed using NetLogo 6.0.4. The purpose of the model is to evaluate design parameters of a volunteer-based crowd-shipping system for rescuing food from restaurants. Texas has more food-insecure households than the average across all states in the U.S. (approximately 1.4 million) [22]. Therefore, the model was designed to explore the potential implementation of such a program in the City of Arlington, which is located in North Texas. The City of Arlington is located in a major metropolitan area with more than 1000 restaurants and no existing program to rescue surplus food from these restaurants. The City of Arlington is divided into 84 census tracts and 259 census block groups. A census block group is the smallest entity for which the U.S. Census Bureau collects and publishes demographic data of the residing population [23]. The preliminary model described in this paper focuses on one of the 84 census tracts in Arlington (1224), which contains 5 census block groups and 18 restaurants. Four shelters (one in census tract 1222 and the remaining three

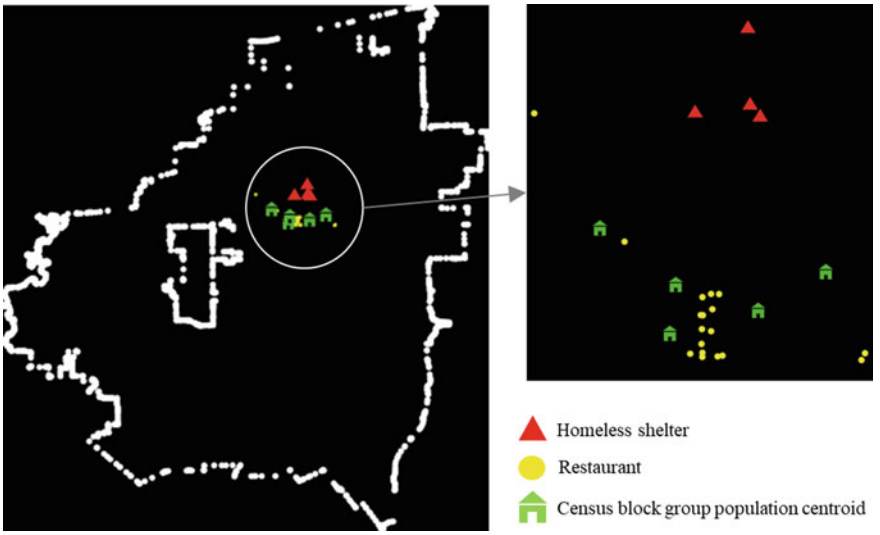


Fig. 1 NetLogo representation of the City of Arlington, showing population centroids of 5 census block groups, 18 restaurants, and 4 homeless shelters (part of the figure on the right is a © [2019] by IEEE. Reprinted, with permission, from [21])

in census tract 1223) are considered as potential recipients of surplus food from the restaurants. U.S. Census Bureau geocoding services were used to obtain the census tracts and block groups corresponding to each restaurant and shelter, based on their street addresses [24]. The locations of population centroids of the 5 census block groups, 18 restaurants, and 4 shelters on the map of the City of Arlington are shown in Fig. 1.

3.1 Model Overview

The ABM contains two types of agents: restaurant agents and crowd-shipper agents. The crowd-shipper agents represent the residents of the five block groups in census tract 1224, all of which (if above age 18) are considered to be potential transportation providers. In each time-step (where a time-step corresponds to one day), the restaurant agents decide whether or not to donate surplus food, and the crowd-shipper agents decide whether or not they will participate in the food rescue program by picking up donations from participating restaurants and delivering them to the assigned shelters. The ABM contains three sub-models: Restaurant Agent Decision-Making, Shelter Assignment, and Crowd-shipper Agent Decision-Making. All three sub-models are executed sequentially in each time-step.

3.2 Sub-Model 1: Restaurant Agent Decision-Making

Each of the 18 restaurant agents is assigned a unique restaurant identification number r. It is assumed that each restaurant agent has surplus food available for donation thrice a week. Each agent's weekly donation schedule is represented by an array of seven binary availability index values ($V_{r,t}$). If restaurant agent r has food available to donate at time-step t, then $V_{r,t}$ will take a value of one, or zero otherwise. Each agent's $V_{r,t}$ values are assigned randomly at the start of the simulation run and are assumed to remain constant for the duration of the run.

A restaurant's decision to donate its surplus food to a food rescue program depends on multiple factors. First, the restaurant must be aware that such a program exists. Once a restaurant learns of the program, its decision to participate may be motivated by sustainability goals (e.g., a desire to prevent food from being sent to landfills) [25] and financial considerations (e.g., tax deductions for charitable donations and reduced waste management fees) [26]. However, many restaurants are discouraged from donating by food safety and liability concerns, being unaware of the Bill Emerson Good Samaritan Act, in which the donor is protected from liability when donating to a non-profit organization [25, 27]. In addition, transportation constraints may prevent restaurants from donating [7]. For example, one restaurant stopped donating its surplus food to a food rescue program after the program's volunteers repeatedly failed to pick up donations at the agreed-upon time [10].

All of these relevant factors were incorporated into the restaurant agents' decision-making logic. In each daily time-step t, if a restaurant agent is aware of the existence of the food rescue program (i.e., its binary awareness variable $A_r = 1$) and it has food available to donate ($V_{r,t} = 1$), it will evaluate its willingness to donate ($W_{r,t}$) based on its total utility ($U_{r,t}$). Total utility is evaluated as the weighted sum of four components and is defined on a scale of -1 to 1: utility due to sustainability goals ($U_{r,s(t)}$), concerns ($U_{r,c(t)}$), past experiences ($U_{r,e(t)}$), and financial benefits ($U_{r,f(t)}$), as given by (1). Larger values of total utility ($U_{r,t}$) correspond to greater donation likelihood.

$$U_{r,t} = \beta_{r,s} U_{r,s(t)} + \beta_{r,c} U_{r,c(t)} + \beta_{r,e} U_{r,e(t)} + \beta_{r,f} U_{r,f(t)} \tag{1}$$

For each restaurant agent in each daily time-step, a random number is generated between 0 and 1. If the number is less than the agent's total utility value ($U_{r,t}$) at time t, the agent is willing to donate food ($W_{r,t} = 1$, or 0 otherwise) and will seek out a crowd-shipper agent for a pick-up. It is assumed that if a restaurant agent successfully finds a crowd-shipper to pick up its donation, it will remain willing to donate food ($W_{r,t} = 1$) in future time-steps until an attempt to find a crowd-shipper fails. In the event of a failure, the restaurant agent will re-evaluate its decision to participate, based on its current total utility ($U_{r,t}$). Also, if a restaurant agent has surplus food available ($V_{r,t} = 1$) and is willing to donate food ($W_{r,t} = 1$) but does not find a crowd-shipper for pick-up three times consecutively, it will stop participating in the food rescue program, with no possibility of rejoining in future time-steps.

Restaurant agents' utility due to sustainability goals ($U_{r,s(t)}$), concerns ($U_{r,c(t)}$), and financial benefits ($U_{r,f(t)}$) are defined on a scale of 0 to 1. Each restaurant agent's utility due to sustainability ($U_{r,s(t)}$) is initially assigned a random value between 0 and 0.5. In each subsequent time-step, $U_{r,s(t)}$ may increase based on interactions with other restaurant agents, in which awareness of the positive social and environmental impacts of food rescue programs is enhanced. These interactions occur via the restaurant agents' social network, which is assumed as an Erdős-Rényi random network [28] with an average degree of connection equal to four. In a given week, the probability of interaction between two socially connected restaurant agents is assumed to be 5%. Upon interaction between two restaurant agents, if one agent is aware of the food donation program, the other agent also becomes aware. Furthermore, the agent with the lower $U_{r,s(t)}$ value will increase this value by 10% of the other agent's $U_{r,s(t)}$ value. The utility due to concern ($U_{r,c(t)}$) for a restaurant agent is given by (2), where $c_{r,t}$ is the agent's r concern level at time-step t. Each agent's $c_{r,t}$ value is initialized as a random value between 0.5 and 1. When two restaurant agents interact via their social network, the concern level of the agent with higher concern decreases by 10% of the concern level of the other agent. Utility gained due to financial benefits from food donation ($U_{r,f(t)}$) has been assigned a value of 0.5 for each restaurant agent, and it remains constant over the duration of the simulation run.

$$U_{r,c(t)} = \frac{1}{e^{2c_{r,t}}} \tag{2}$$

A restaurant agent's utility due to past experiences ($U_{r,e(t)}$) is defined on a scale of -1 and 1 and is a combination of the agent's personal experiences with the food rescue program and the number of interactions (i_r) with other restaurant agents who have stopped participating in the program due to inability to source deliveries from the crowd-shippers. The agent's personal experiences are evaluated using the ratio of the number of days ($N_{r,d}$) in which the agent sought and successfully found a crowd-shipper agent to pick up its donation, to the total number of days (d_r) in which the agent was aware of the food rescue program ($A_r = 1$) and had food available to donate ($V_{r,t} = 1$). $U_{r,e(t)}$ for a restaurant agent is evaluated using (3).

$$U_{r,e(t)} = \frac{N_{r,d}}{d_r} - \frac{i_r}{10} \tag{3}$$

The weights on utility due to sustainability goals ($\beta_{r,s}$), concerns ($\beta_{r,c}$), past experiences ($\beta_{r,e}$), and financial benefits ($\beta_{r,f}$) for the restaurant agents are assumed as 0.1, 0.2, 0.5, and 0.2, respectively. Higher weight is assigned to utility due to past experiences ($U_{r,e(t)}$) based on the assumption that restaurants will be more likely to participate in the food rescue program if they have previously experienced more successful deliveries and have received less negative feedback from restaurants that have stopped participating in the program.

3.3 Sub-Model 2: Shelter Assignment

In each time-step, if a particular restaurant agent r is willing to donate food ($W_{r,t} = 1$), the donation is randomly assigned to one of the four homeless shelters. It is assumed that shelters are able to receive food on any day of the week and have no capacity constraints.

3.4 Sub-Model 3: Crowd-Shipper Agent Decision-Making

There are a total of 4579 crowd-shipper agents in the model, representing residents of census tract 1224. Each crowd-shipper agent belongs to one of the five census block groups in this tract and is assigned a unique identification number, c. Population centroids (latitude and longitude coordinates) of these five block groups were obtained from the U.S. Census Bureau. It is assumed that each crowd-shipper agent's residence is located at the population centroid of its respective block group. Using the centroid of a census block group as a point of origin is a common assumption when calculating travel distances for the population residing within the block group to a particular destination [29].

Crowd-shipper agents are classified using five demographic factors, as per the classification of food rescue program volunteers by [30]: age (18–25, 26–45, or 46–69), gender (male or female), ethnicity (Non-Hispanic White, African American, or Hispanic), education attainment (high school, partial college, college/university, or graduate school), and annual income (<$17,500, $17,500–$47,000, $48,000–$66,000, or $67,000–$80,000). Each agent's demographics are assigned based on 2017 U.S. Census Bureau statistics that correspond to the agent's block group [31]. An agent's demographics are assumed to remain constant throughout each simulation run. It is assumed that every crowd-shipper agent owns a vehicle and is capable of participating in the food rescue program.

Motivations for individuals to participate in food rescue programs include service requirements of a social organization, career improvement, and altruism [30]. Typically, food rescue volunteers are not financially motivated to participate. However, food rescue via crowd-shipping is a relatively new concept—traditionally, volunteer food rescue activities occur at food bank/pantry warehouses. Therefore, encouraging sufficient participation might require some financial incentives. For example, donor restaurants' tax deductions are used to fund one of the largest fresh food donation programs in the North America [32]. A similar scheme could be employed to incentivize food rescue crowd-shippers. Finally, the motivation to serve as a volunteer crowd-shipper may be impacted by previous experiences. For example, a lack of consistent opportunities to participate in the food rescue program could decrease a volunteer's motivation, as continuous participation and enthusiasm to volunteer are interrelated [33].

In this model, it is assumed that each crowd-shipper agent will not volunteer more than once a week (i.e., once in every seven time-steps) to rescue food from a restaurant. Similar to the restaurant agents, each crowd-shipper agent c has a binary awareness variable (A_c) which takes a value of one if the agent is aware of the food rescue program, or zero otherwise. It is also assumed that crowd-shipper agents are available to volunteer three times a week. Each agent's availability schedule is represented by a set of seven binary availability index values ($V_{c,t}$). If a crowd-shipper agent c is available to volunteer at time-step t, then $V_{c,t}$ will take a value of one, or zero otherwise. Each agent's $V_{c,t}$ values are assigned randomly at the start of the simulation run and are assumed to remain constant for the duration of the run. If A_c and $V_{c,t}$ are equal to 1, and if the agent has not volunteered previously for the food rescue program in the current week, the agent evaluates its willingness to volunteer ($W_{c,t}$) at time-step t based on its current total utility ($U_{c,t}$). $U_{c,t}$ for each crowd-shipper agent is defined on a scale of -1 to 1 and is a weighted sum of three components: utility due to motivation ($U_{c,m(t)}$), financial benefits ($U_{c,f(t)}$), and past experiences ($U_{c,e(t)}$), as given by (4). In each time-step, a random number is generated between 0 and 1 for each aware crowd-shipping agent, and if the number is less than the agent's total utility value ($U_{c,t}$), the agent is assumed to be willing to rescue food from a restaurant agent ($W_{c,t} = 1$, or 0 otherwise).

$$U_{c,t} = \beta_{c,m} U_{c,m(t)} + \beta_{c,f} U_{c,f(t)} + \beta_{c,e} U_{c,e(t)} \tag{4}$$

In reality, even if a potential crowd-shipper is willing ($W_{c,t} = 1$) to participate, other obligations and time constraints may prevent him/her from doing so. To allow for these factors, a willing crowd-shipper agent's final decision to choose a particular food donation delivery d is based on its availability at the time of the pick-up of the delivery, given by an agent's pick-up availability index (P_c) and the total time required (T_d) to complete the delivery. The total required time (T_d) includes four components: travel time from the population centroid of the crowd-shipper's census block to the restaurant location, travel time from the restaurant to the assigned homeless shelter, travel time from the homeless shelter back to the census block centroid, and the total time spent in waiting, loading, and unloading food at the restaurant and homeless shelters. Travel times between census blocks, restaurants, and homeless shelters are estimated using the Google Maps API. The total waiting time involved at restaurants and homeless shelters is assumed as 10 min in the model. The pick-up availability index of a crowd-shipper (P_c) is assigned based on its age level, where a higher index value corresponds to a greater probability that the agent will participate. The pick-up availability index (P_c) is assigned a value of 0.5 for crowd-shipper agents that have an age level of 18–25 or 26–45, and a value of 0.75 is assigned for agents with an age level of 45–69. This logic is based on the assumption that senior crowd-shippers (i.e., retired persons) have more availability for volunteer activities.

A willing crowd-shipper agent will look at available deliveries in the list and make its final decision to make a delivery d based on its availability at the pick-up time (P_c) and convenience utility ($C_{c,d}$), which is given by (5). Convenience utility ($C_{c,d}$) for a delivery d is a function of total time involved in completing the delivery (T_d),

with greater the time, lesser the utility value.

$$C_{c,d} = \frac{1}{e^{2T_d}} \tag{5}$$

Two random numbers are generated between 0 and 1 and if each of the numbers are less than P_c and $C_{c,d}$, respectively, the crowd-shipper agent will participate in the food rescue program, and the particular delivery will be removed from the list of potential deliveries for the other crowd-shipper agents. This randomness is introduced to represent heterogeneity in crowd-shipper agent behaviors that is not explicitly represented by the state variables in the model. Also, if a crowd-shipper has been willing to volunteer three times consecutively but was unable to make any delivery due to time constraints (lack of availability at the time of pick-up determined by an agent's P_c and/or lack of convenience determined by $C_{c,d}$), it will stop participating in the program in future time-steps.

Crowd-shipper agents' utilities due to motivation ($U_{c,m(t)}$) and financial benefits ($U_{c,f(t)}$) are defined on a scale of 0 to 1. The initial value of $U_{c,m(t)}$ for each crowd-shipper agent is derived from the motivation scale defined by [30], which is based on survey data collected from volunteers who participate in food rescue programs. The motivation score means, standard deviation, and range for each demographic factor level of the volunteers surveyed is shown in Table 1. These statistics were used to define probability distributions (as shown in Table 1), from which initial $U_{c,m(t)}$ values are drawn for each crowd-shipper agent. The five motivation scores from each demographic factor are then averaged, normalized to a value between zero and one, and assigned to the agents.

Crowd-shipper motivation is assumed to be influenced by social interactions. Results from a national survey indicate that, on average, a person knows approximately 13 people in his/her neighborhood [34]. Thus, an Erdős-Rényi random social network with an average degree of 13 is used to connect the crowd-shipper agents residing within the same census tract. The probability of an interaction between any two connected crowd-shipper agents in a given week is assumed to be 0.5%. If a crowd-shipper agent is aware of the food rescue program ($A_c = 1$) and interacts with an agent in its social network, the other agent also becomes aware. Upon interaction, the crowd-shipper whose $U_{c,m(t)}$ value is lower will increase this value by 1% of the $U_{c,m(t)}$ value of the other crowd-shipper.

Utility due to past experiences ($U_{c,e(t)}$) is defined on a scale of -1 to 1, and is based on the regularity of a crowd-shipper's participation in food rescue program and its interactions with other crowd-shippers who have stopped participating in the food rescue program. $U_{c,e(t)}$ is evaluated using (6), where $N_{c,w}$ is the total number of weeks a crowd-shipper has participated in the food rescue program, w_c is the total number of weeks that the crowd-shipper has been aware of the program (when $A_c = 1$), and i_c is the number of interactions a crowd-shipper has had with other agents who have stopped participating in the food rescue program.

Table 1 Summary statistics and probability distributions used to determine crowd-shipper agent's initial motivation utility values (average of motivation score from each demographic factor was normalized between 0 and 1 to assign to each crowd-shipper agent) (© [2019] by IEEE. Reprinted, with permission, from [21])

Demographic factor: level	M	SD	Range	Assumed distribution
Age: 18–25	8.97	2.97	[1, 14]	Truncated normal (1,14)
Age: 26–45	7.94	4.31	[1, 14]	Truncated normal (1,14)
Age: 46–69	10.93	0.87	[10, 13]	Truncated normal (10,13)
Gender: Men	7.96	2.96	[1, 14]	Truncated normal (1,14)
Gender: Women	9.78	2.94	[1, 14]	Truncated normal (1,14)
Ethnicity: Non-Hispanic White	9.27	2.97	[1, 14]	Truncated normal (1,14)
Ethnicity: African American	8.26	21.97	[3, 14]	Uniform (3,14)
Ethnicity: Hispanic	9.65	2.13	[6, 12]	Truncated normal (6,12)
Education: High school	6.05	6.36	[1, 12]	Uniform (1,12)
Education: Partial college	8.67	1.30	[7, 10]	Truncated normal (7,10)
Education: College/university	9.36	2.84	[1, 14]	Truncated normal (1,14)
Education: Graduate school	3.63	10.24	[1, 8]	Uniform (1,8)
Annual income: <17,500	8.79	2.83	[1, 14]	Truncated normal (1,14)
Annual income: 17,500–47,000	7.53	4.64	[1, 14]	Truncated normal (1,14)
Annual income: 48,000–66,000	11.24	1.48	[10, 14]	Truncated normal (10,14)
Annual income: 67,000–80,000	10.67	1.59	[8, 12]	Truncated normal (8,12)

$$U_{c,e(t)} = \frac{N_{c,w}}{w_c \ (\forall A_c = 1)} - \frac{i_c}{10} \tag{6}$$

It is assumed in this model that the crowd-shipper agents are not financially incentivized to participate in the food rescue program. Therefore, utility gained due to financial benefits ($U_{c,f(t)}$) is assigned a value of 0 for all crowd-shipper agents and it remains constant over the duration of the simulation run. The weights on utility due to motivation ($\beta_{c,m}$), financial benefits ($\beta_{c,f}$), and past experiences ($\beta_{c,e}$) are assumed as 0.25, 0.25, and 0.5, respectively. It was assumed that the crowd-shippers' utility due to past experiences ($U_{c,e(t)}$) was the most influential of the three elements, based on research that suggests that volunteer participation tends to increase with accumulating experience [35] and fewer negative interactions with former volunteers.

3.5 Initialization

The initial number of restaurant and crowd-shipper agents who are aware of food rescue program (i.e., restaurants with $A_r = 1$ and crowd-shippers with $A_c = 1$) is

varied experimentally to identify the effect of initial starting population on the system metrics over the simulation runtime. Figure 2 shows the flowchart representing the three sub-models executed at each time-step in the ABM.

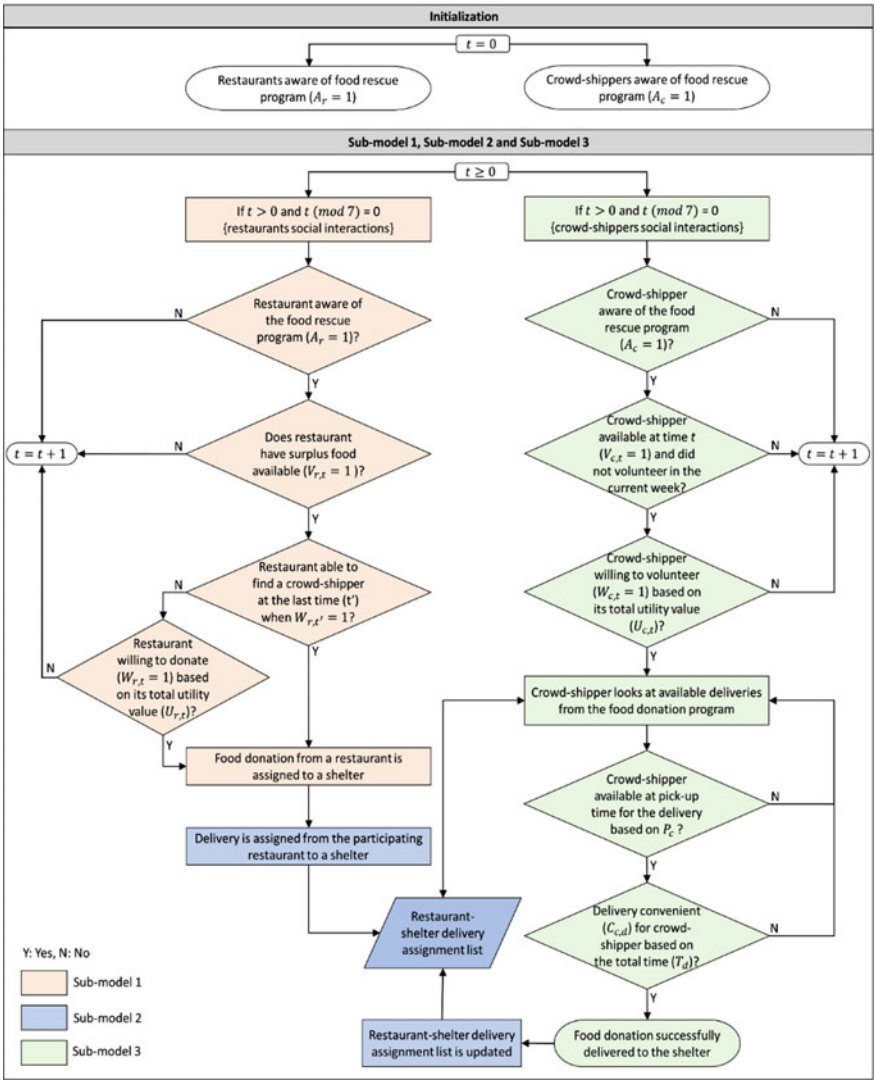


Fig. 2 Flowchart of the three sub-models executed at each time-step in the ABM

4 Experimentation and Results

The ABM was used to investigate factors affecting the viability of a crowd-based volunteer food rescue program. System viability is achieved when the amount of food donated by restaurants and the number of participating crowd-shippers are sufficient to fulfill food donation requests, and this balance is successfully maintained over time. Users who request service via a crowd logistics platform (i.e., restaurants) will only find the platform useful if there are sufficient service providers (i.e., volunteer crowd-shippers), and vice versa. Therefore, it is important to ensure that there is an appropriate balance between the number of service requesters and providers when the program is initially launched, to avoid immediate program failure.

To gain a greater understanding of how to determine the right initial balance, the initial percentages of crowd-shippers (C_0) and restaurants (R_0) who are aware of the food rescue program (i.e., restaurants with $A_r = 1$ and crowd-shippers with $A_c = 1$) were experimentally varied. In each experimental scenario, three key output metrics are captured in each daily time-step: the number of meals rescued and the number of restaurants (R_c) and crowd-shippers (C_c) who are aware of the food rescue program and continue to evaluate participation. On average, each restaurant in the U.S. generates approximately 50,000 lb of surplus food per year and this value was used to determine the potential food donation (three times every week for the 52-week period) for each restaurant in the model [36]. Also, it has been assumed that each pound of food being rescued corresponds to 0.83 meals [6]. For each experimental scenario, 100 replications of 364 daily time-steps were run.

First, the initial percentage of restaurants aware of the program (R_0) was assigned a value of 5%, and the initial percentage of aware crowd-shippers (C_0) was assigned values of 0.5%, 1%, 1.5%, and 2%. Figure 3 shows the total number of meals rescued at the end of one year for the four different values of C_0.

The results indicate that more meals are rescued as the initial value of C_0 is increased, which suggests that increasing C_0 has a positive effect on total restaurant participation. This observation is supported by the data in Fig. 4, which shows the

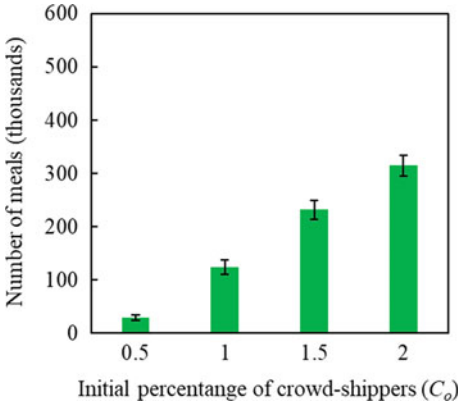


Fig. 3 Number of meals rescued at the end of one year when 5% of restaurants (R_0) were initially assumed to be aware of the food rescue program, with C_0 varied from 0.5% to 2%

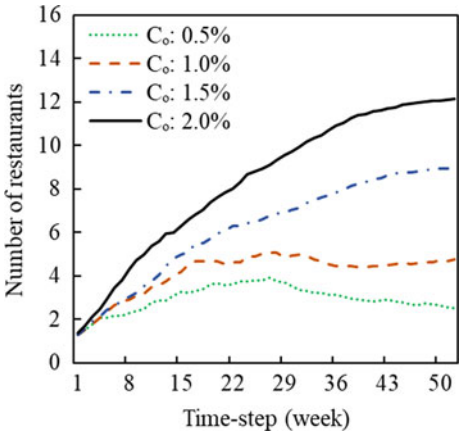


Fig. 4 Number of restaurants that continue to evaluate participation in the food rescue program (R_c), when 5 percent of restaurants (R_o) were initially assumed to be aware of the program with C_0 varied from 0.5% to 2%

number of restaurants every week that continued to evaluate participation (R_c) when C_0 was varied. The number of crowd-shippers available to make deliveries (C_c) followed a similar pattern (Fig. 5), although the value of C_c drops early on in all four cases. The reason for this drop in participation is likely related to the low value of R_c in the initial time-steps—with few delivery requests, crowd-shippers find few opportunities to participate. However, C_c increases in the later time-steps, indicating the mutually positive effect of the increasing number of participating restaurants and the increasing number of crowd-shippers on each other.

The higher initial percentage of aware crowd-shippers (C_o) increases the rate of information diffusion among other potential crowd-shippers, improving food rescue operations by reducing the number of restaurants that stop participating due to repeated failed pick-ups. This suggests that number of crowd-shippers available in the beginning of the program (C_0) is an important aspect of the program design,

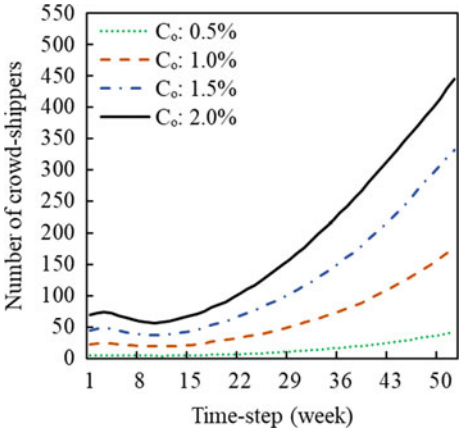


Fig. 5 Number of crowd-shippers (C_c) that continue to evaluate participation in the food rescue program, when 5 percent of restaurants (R_o) were initially assumed to be aware of the program with C_0 varied from 0.5% to 2%

and a higher C_0 value leads to higher number of restaurants who continue to evaluate participation (R_C) in the program.

Given that greater values of C_o tend to yield better system performance, the next set of experiments was performed with C_0 set to 1.5% or 2%, while the initial number of restaurants aware of the program (R_0) was varied. Figure 6 shows the number of meals rescued at the end of one year when R_o was varied and was set to 5%, 25%, 50%, and 75% for these two different values of C_0. In all cases, increasing R_o has a positive effect on the number of meals rescued.

However, Fig. 7a shows that for C_0 =1.5%, increasing R_0 resulted in fewer restaurants that continued to evaluate participation (R_c) at the end of one year. By

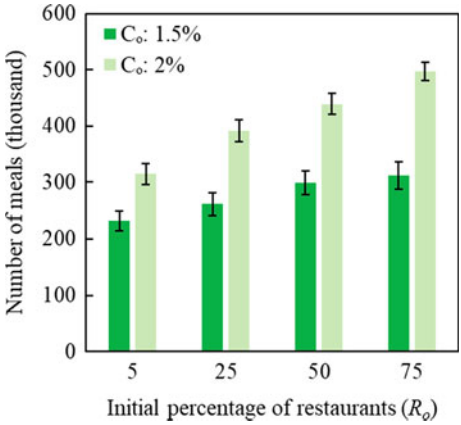


Fig. 6 Number of meals rescued at the end of one year when initial number of restaurants aware of the program (R_o) is varied from 5 to 75%, with C_0 set to 1.5% and 2%

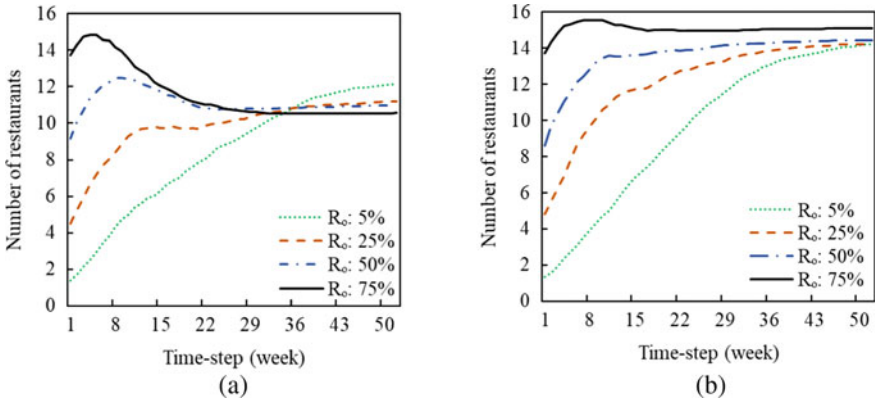


Fig. 7 Number of restaurants that continue to evaluate participation in the food rescue program (R_c) when C_0 is **a** 1.5%, **b** 2.0%

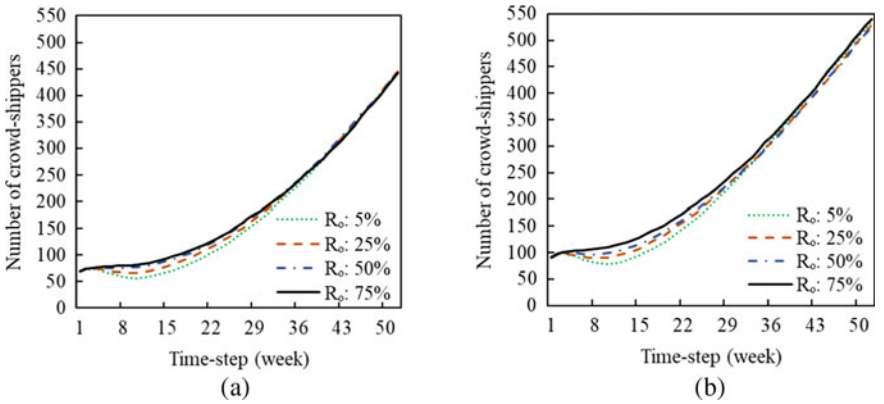


Fig. 8 Number of crowd-shippers that continue to evaluate participation in the food rescue program (C_c) when C_0 is **a** 1.5%, **b** 2.0%

contrast, Fig. 7b shows that when $C_0 = 2.0\%$, the number of participating restaurants (R_c) was always higher when R_0 was higher. In both cases, the number of crowd-shippers available to participate (C_c) in the initial time-steps are greater when the initial percentage of restaurants aware of the food rescue program (R_0) is greater (Fig. 8a, b). Increased availability of restaurants provided greater opportunities for crowd-shippers to deliver food donations, thereby reducing the number who stopped participating in the program due to a lack of available deliveries in the initial time-steps.

5 Conclusion

This paper describes a conceptual ABM that was designed to study the viability of a volunteer-based crowd-shipping program for food rescue. Preliminary experimental results from the model demonstrate the importance of achieving the right balance between the initial number of restaurant and crowd-shipper participants on the program's effectiveness, in terms of the number of meals rescued and number of restaurants who continue to participate in the program. The conceptual model described in this paper will serve as a starting point for future research. Empirical data on crowd-shipper and restaurant behavior will be collected to gain a greater understanding of crowd-shippers' behaviors and preferences, as well as insights into restaurants' decision-making processes. Using this data, the existing model will be enhanced, such that it will be capable of supporting design decisions for new food rescue programs. For example, the model can help to identify the degree to which providing monetary incentives to crowd-shippers supports program effectiveness and long-term sustainability. Developing viable crowd-sourced transportation programs to rescue surplus food will help alleviate food insecurity as well as reduce the food

waste and its accompanying environmental impacts: greenhouse gas emissions from landfills.

Acknowledgements The authors would like to acknowledge Nicole Straight from Food Rescue US and Zebedee McLaurin from Goodr for sharing valuable insights on their respective food rescue operations.

References

1. USDA: Key Statistics & Graphics. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics/#foodsecure (2018). Cited 18 July 2019
2. USDA: OCE | U.S. Food Waste Challenge | FAQ's. https://www.usda.gov/oce/foodwaste/faqs.htm (2019). Cited 18 July 2019
3. Feeding America: Fighting Food Waste with Food Rescue. http://www.feedingamerica.org/our-work/our-approach/reduce-food-waste.html (2018a). Cited 18 July 2019
4. Tarasuk, V., Eakin, J.M.: Food assistance through surplus food: insights from an ethnographic study of food bank work. Agric. Hum. Values **22**(2), 177–186 (2005)
5. Feeding America: Our approach to food waste and rescue. https://www.feedingamerica.org/our-work/our-approach/reduce-food-waste (2018b). Cited 3 May 2019
6. ReFED: Restaurant food waste action guide. https://www.refed.com/downloads/Restaurant_Guide_Web.pdf (2018). Cited 3 May 2019
7. Berkenkamp, J., Phillips, C.: Modeling the potential to increase food rescue: Denver, New York City, and Nashville. National Resources Defense Council. https://www.nrdc.org/sites/default/files/modeling-potential-increase-food-rescue-report.pdf (2017). Cited 3 May 2019
8. Gunders, D., Bloom, J.: Wasted: how America is losing up to 40 percent of its food from farm to fork to landfill. 2nd Edition of NRDC's Original 2012 Report, National Resources Defense Council, New York City, New York (2017)
9. Rai, H.B., Verlinde, S., Merckx, J., Macharis, C.: Crowd logistics: an opportunity for more sustainable urban freight transport? Eur. Transp. Res. Rev. **9**(39), 1–13 (2017)
10. Krejci, C.C., Oran Gibson, N.: Interview with nicole straight, Fairfield county site director of food rescue US (2019). Cited 29 March 2019
11. Chatlani, S.: Delivery companies finding ways to help restaurants donate excess food. https://www.marketplace.org/2019/02/19/wealth-poverty/food-delivery-companies-enable-easy-donation-restaurants-excess (2019). Cited 3 May 2019
12. Frehe, V., Mehmann, J., Teuteberg, F.: Understanding and assessing crowd logistics business models-using everyday people for last mile delivery. J. Business Indus. Market. **32**(1), 75–97 (2017)
13. Miller, J., Nie, Y., Stathopoulos, A.: Crowdsourced urban package delivery: modelling traveler willingness to work as crowdshippers. Transp. Res. Record J. Transp. Res. Board **2610**(1), 67–75 (2017)
14. Le, T.V., Ukkusuri, S.V.: Modeling the willingness to work as crowd-shippers and travel time tolerance in emerging logistics services. Travel Behav. Soc. **15**, 123–132 (2019)
15. Ermagun, A., Stathopoulos, A.: To bid or not to bid: an empirical study of the supply determinants of crowd-shipping. Transp. Res. Part A **116**, 468–483 (2018)
16. Devari, A., Nikolaev, A.G., He, Q.: Crowdsourcing the last mile delivery of online orders by exploiting the social networks of retail store customers. Transp. Res. Part E **105**, 105–122 (2017)
17. Zou, G., Gil, A., Tharayil, M.: An agent-based model for crowdsourcing systems. In: Proceedings of the 2014 Winter Simulation Conference, pp. 407–418 (2014)

18. Wise, S., Cheliotis, K., Bates, O., Friday, A., Allen, J., McLeod, F., Cherrett, T.: Using an agent-based model to explore alternative modes of last-mile parcel delivery in urban contexts. In: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Simulation, pp. 1–4 (2018)
19. Chen, P., Chankov, S.M.: Crowdsourced delivery for last-mile distribution: an agent-based modelling and simulation approach. Int. Conf. Indus. Eng. Eng. Manage. **2017**, 1271–1275 (2017)
20. Van de Westelaken, M., Zhang, Y.: An agent-based model for feasibility and diffusion of crowd shipping. In: 29th Benelux Conference on Artificial Intelligence, p. 419 (2017)
21. Mittal, A., Gibson, N.O., Krejci, C.C.: An agent-based model of surplus food rescue using crowd-shipping. In: 2019 Winter Simulation Conference. National Harbor, Maryland, U.S., pp. 854–865 (2019)
22. Feeding Texas: What is Food Insecurity? https://feedingtexas.org/learn/what-is-food-insecurity/ (2019). Cited 18 July 2019
23. ProximityOne: Census Block Groups and Block Group Codes. http://proximityone.com/geo_blockgroups.htm (2019). Cited 4 April 2019
24. U.S. Census Bureau: https://geocoding.geo.census.gov/geocoder/geographies/addressbatch?form (2019a). Cited 4 April 2019
25. Brahm, J., Helen, C., Samonte, T., Yang, Y.: Increasing restaurant food donations: a strategy for food waste diversion. Sanford School of Public Policy, Duke University, Durham, North Carolina (2014)
26. Feeding America: Changes to Food Donation Tax Incentives. https://www.foodbankcny.org/assets/Documents/feeding-america-2016-food-donation-tax-law-changes.pdf (2016). Cited 15 April 2019
27. Feeding America: Protecting Our Food Partners. https://www.feedingamerica.org/about-us/partners/become-a-product-partner/food-partners (2018c). Cited 2 May 2019
28. Newman, M.E.J.: Random Graphs as Models of Network. Cornell University, Ithaca, New York (2002)
29. Biba, S., Curtin, K.M., Manca, G.: A new method for determining the population with walking access to transit. Int. J. Geogr. Inf. Sci. **24**(3), 347–364 (2010)
30. Mousa, T.Y., Freeland-Graves, J.H.: Motivations for volunteers in food rescue nutrition. Public Health **149**, 113–119 (2017)
31. U.S. Census Bureau: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml (2019b). Cited 4 April 2019
32. Food Donation Connection: Who We Are. https://www.foodtodonate.com/about (2018). Cited 6 April 2019
33. Schanes, K., Stagl, S.: Food waste fighters: what motivates people to engage in food sharing? J. Clean Prod. **211**, 1491–1501 (2019)
34. McCarty, C., Killworth, P.D., Bernard, H.R., Johnsen, E.C., Shelley, G.A.: Comparing two methods for estimating network size. Hum. Organ. **60**(1), 28–39 (2001)
35. Harrison, D.A.: Volunteer motivation and attendance decisions: competitive theory testing in multiple samples from a homeless shelter. J. Appl. Psychol. **80**(3), 371 (1995)
36. Green Restaurant Association: Zeroing in on Restaurant Food Waste. https://www.dinegreen.com/single-post/2017/11/20/Zeroing-in-on-Restaurant-Food-Waste (2017). Cited 31 July 2019

Exogenous Shocks Lead to Increased Responsiveness and Shifts in Sentimental Resilience in Online Discussions

Chathika Gunaratne, Subash K. Ray, Caroline Lourenço Alves, and Maria Waldl

Abstract The effects of real-world events on the dynamics and sentiment expressed through online conversations are not entirely understood. In particular, the dynamics of highly polarized communities, deeply invested in the outcome of a particular event naturally tend to have a contrasting emotional sensitivity to exogenous events. In this study, we analyze Twitter conversations during the Mexico versus Germany group match of the 2018 FIFA World Cup and investigate the effect of the live game outcomes on the conversation dynamics and sentiment. We find that the exogenous events influence conversation volume and virality less, while having high influence on user responsiveness. Interestingly, we observe a shift in the influence that exogenous events have on fans immediately following the only goal scored in the game. The emotional resilience of fans of the advantaged team increased following this exogenous shock. In contrast, the sentiment of the disadvantaged team was left more susceptible to further exogenous events following the exogenous shock. These results support the fact that user engagement and emotional resilience of an online population holding highly polarized stances can be manipulated through an exogenous event of high importance.

C. Gunaratne (✉)
University of Central Florida, Orlando, FL, USA
e-mail: chathika@knights.ucf.edu

S. K. Ray
New Jersey Institute of Technology & Rutgers University - Newark, Newark, NJ, USA
e-mail: sr523@njit.edu

C. Lourenço Alves
University of Sao Paulo, Sao Paulo, Brazil
e-mail: caroline.lourenco.alves@usp.br

M. Waldl
University of Vienna, Vienna, Austria
e-mail: maria@tbi.univie.ac.at

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_5

1 Introduction

Online social networks offer a medium for the discussion of real-world events at global scales, transcending the challenges experienced by other forms of human communication. Individuals with completely different experiences and cultural backgrounds and opposing opinions on a topic may now be spontaneously exposed to one another when discussing a common topic of interest. This results in complex conversation dynamics arising between the social media users due to the many factors driving the expression of ideas and/or opinions, compounded by the constant intrusion of opposing information or opinions.

Online social media discussions are typically initiated by an event exogenous to the social media platform. If a sufficient number and rate of responses to an initial post are received, then the original post (or root) is able to accumulate responses over time, endogenously growing into and online conversation, also referred to as an information cascade. Often, new information is generated in response to an exogenous event in the form of replies to a tweet about the event. This influx of new information contributes in the continuation of a conversation and the expression of opinions and positions.

Understanding the extent to which exogenous events affect online conversation dynamics in comparison to the endogenous forces of discussion is the focus of this study. We expect the impact of exogenous information on online conversation dynamics to be stronger when considering a rapidly discussed topic, with several unprecedented and high-impact events, being discussed among two groups of a highly invested and highly polarized populations, such as a live sports game or an election. Accordingly, we choose the Mexico versus Germany group match of the 2018 FIFA World Cup as our topic of interest, which ended with an unprecedented win by Mexico scoring 1–0. We analyze the conversations that were exchanged over Twitter right before, during, and after the single goal scored over the course of the game. We selected this game in particular, due to the win for Mexico over the team with a stronger historical record, and for the lingual difference between fans of the two teams, for the ease of analysis. The Twitter discussions during the course of this game provide a unique opportunity to study a polarized population engaged in competitive discussion, in isolation, over a fixed time period, along with the explicit specification of exogenous shocks important to the community in the form of the game commentary transcript.

In this study, we analyze the effect that the exogenous events, in particular the game events reported in the game transcript, have on the online conversation dynamics and the sentiment expressed by fans of both teams. Conversation dynamics are quantified via three measurements: (1) *volume*, the total number of replies or retweets to the root tweet or replies or retweets of the resulting conversation, (2) *virality*, the Weiner index of the conversation tree resulting from the root tweet, and (3) *responsiveness*, the rate at which replies and retweets are accumulated by the conversation resulting from the root tweet. These measures have been used in similar investigations in the literature [2]. Furthermore, we performed sentiment analysis [5] on both the Twitter conversations and game commentary to obtain polarity scores, to quantify sentiment.

Finally, we performed transfer entropy analysis on the resulting time-series data to investigate the influence that the game events had on the conversation dynamics and the influence that the commentary sentiment had on the sentiment expressed in the conversations.

Our results indicate that the popularity of the root user had greater influence than the game events on the volume and virality of the Twitter conversations. However, the game events were found to have stronger influence overall on the responsiveness to the conversations. Further, game events at around 7 s prior to the conversation had the highest influence on the conversation volume, while game events around 22–29 s had highest influence on the virality of conversations overall. Interestingly, game events by Germany had a higher influence on the responsiveness than the game events by Mexico, despite Mexico winning the match. The commentary sentiment is shown to have higher influence coming up to the goal than after the goal overall. Finally, we see that users tweeting in German language have higher sentimental influence after the goal than prior to the goal by Mexico, indicating that there was anticipation following the goal by Mexico, that Germany would score in the game.

2 Background

Sentiment analysis is a growing area of Natural Language Processing (NLP) and refers to the extraction and classification of a text primarily available on the web, like social media posts, news articles, and blog posts, to determine the opinions, emotions, and attitude of people [6, 12]. The purpose of this type of analysis is to subjectify people's sentiments about a news item (for e.g., a political situation, sports, and weather), reviews on books, movies, or consumer product, which are carried out by political analysts, marketers, and companies. A typical approach to sentiment analysis is to start with a lexicon of positive, negative, and neutral words and phrases [12]; these entries are labeled with their polarity. In particular, Twitter has been one of the most popular social networking websites generating enormous amount of data because of its millions of followers that are increasing by the day [10]. Therefore, sentiment analysis on data has been an effective tool in discerning public attitude toward a certain topic [8].

Choi et al. [2] use trees to represent the structure of conversations. Vertices of these trees represent comments by users, linked by edges which represent a response to another comment. The authors then characterize these trees with measures of volume, virality, and responsiveness.

In addition, in order to compute the information transfer from the sentiments of the events happening during the game to the sentiments of the tweets posted by people, we used transfer entropy. Transfer entropy is a model-free, information-theoretic tool that measures the amount of directed predictive information flow between two time-series when one of the time-series is hypothesized to influence the other [9]. In specific, transfer entropy from a time-series X to a time-series Y (or $T_{X \to Y}$) is the measure of reduction in uncertainty (or increase in predictability) in the future

values of Y by using the knowledge of the past values of X given the past values of Y. Transfer entropy $T_{X \to Y}$ can be calculated using Eq. 1.

$$T_{X \to Y} = \sum_{x_{i+1}, x_i^{(k)}, y_i} p(x_{i+1}, x_i^{(k)}, y_i) \log_2 \frac{p(x_{i+1}, y_i | x_i^{(k)})}{p(x_{i+1} | x_i^{(k)}) p(y_i | x_i^{(k)})}. \quad (1)$$

Here, X and Y are time-series, with $x_i^{(k)} = x_{(i-k+1)}, x_{(i-k+2)}, ..., x_i$ represents the k successive past values of time-series X at time i and including the value at the current time (i.e., x_i).

3 Methodology

3.1 Data

We collected 382042 tweets by fans posting on the Mexico versus Germany opening game of FIFA 2018 using Developer API. Tweets were collected 30 min before the game started to 30 min after the game ended, from 2018-06-17 14:30:00 to 2018-06-17 17:00:00. Tweets that mentioned the hashtags: #GERMEX, #FIFA, #Germany, or #Mexico, are were created during this period were collected. The collected tweets were in multiple languages, from which English, Spanish, and German were the most commonly used as shown in Fig. 1. The tweet texts were translated to English using Amazon's AWS Translate API with AWS EC2. The tweet texts were then stemmed and run through the Vader sentiment analysis algorithm to extract polarity scores. In addition to collecting the tweets, the game transcript was obtained to represent the effect of the events of the game on the fans.[1]

3.2 Conversation Dynamics

Conversations were treated as trees of response tweets to an originating root tweet. A conversation could have none to many replies, be shallow or deep/viral, and could call low to high response rates, respectively. These conversation dynamics were measured by three properties similar to [2], volume, virality, and responsiveness. Volume was simply measured as the total number of tweet replies, retweets, or quoted tweets that a root tweet received.

Secondly, structural virality indicated the shape of the conversation by the depth of the conversation tree with respect to its width. A tree with high virality would have a long chain of replies (Fig. 2), compared to replies centered around a single tweet. As in [2], structural virality was calculated via the Wiener Index [11], a graph

[1] http://www.espn.com/soccer/commentary?gameId=498193.

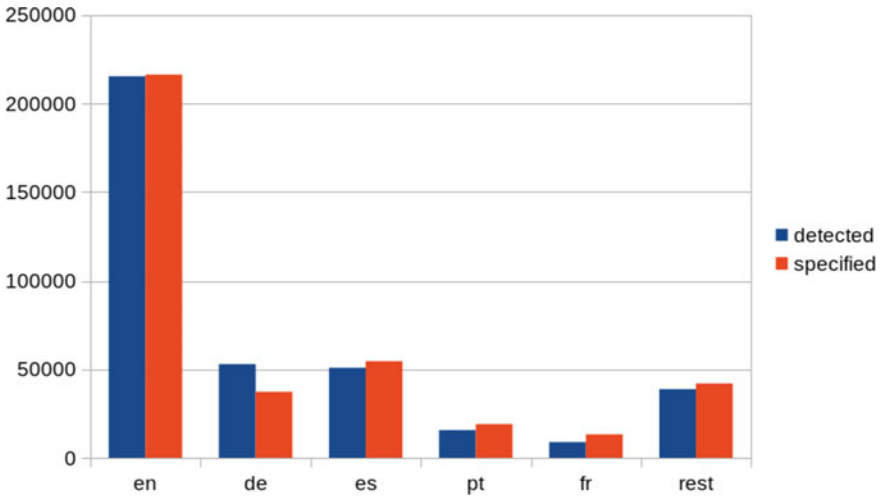


Fig. 1 Histogram of language of collected tweets as specified by Twitter API (red) and detected (blue) through Python-langdetect library. English (en), German (de), and Spanish (es) are the most common, followed by Portuguese (pt) and French (fr)

(a) Large conversation with low virality. (b) Large conversation with high virality.

Fig. 2 Visual examples of two large conversations in the dataset, with 2a low virality and 2b high virality, as determined through the Wiener index. Sentiment polarity scores of each tweet have been depicted as negative (red), neutral (yellow), or positive (green)

property that can be defined as the average distance between all pairs of nodes in a diffusion tree [3], *v(T)*, described in Eq. 2.

$$v(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}, \tag{2}$$

where d_{ij} denotes the length of the shortest path between nodes i and j; and n the number of all nodes.

Thirdly, Responsiveness indicated the rate at which replies would be posted to a parent tweet on a conversation. Responsiveness was calculated as the mean, inverse difference between the time of creation of the reply and the time of creation of the parent tweet, for all replies on a conversation. We used a resolution of seconds for

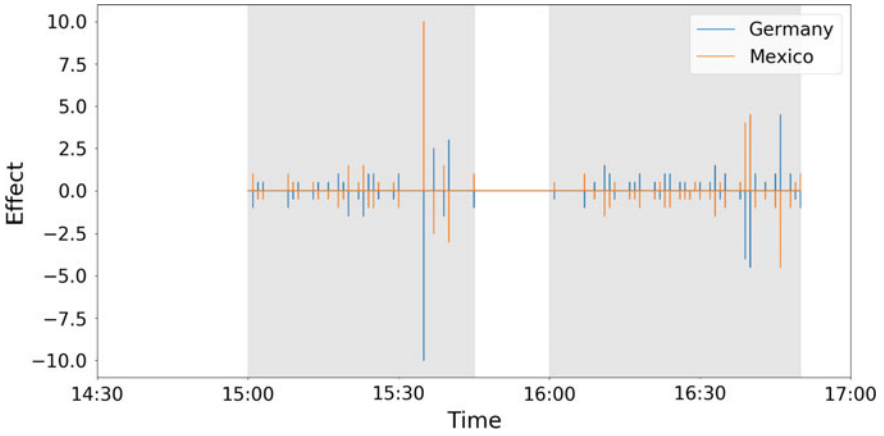


Fig. 3 The game events time-series generated by evaluation of the events reported in the game transcript. Both teams received scores for each event and the scores for Germany are denoted in blue, while those for Mexico are in orange

the creation of timestamps, making the unit of responsiveness used in the study in terms of tweets per second.

Finally, the transcript was used to generate a time-series of game event effects experienced by both the German and the Mexico fans, along with an overall total game events time-series. The game event affects time-series for both the German and Mexican teams were created by applying the following rules, to approximate the intensity of exogenous game events on either team due to the events reported on the transcript: a foul by a team results in −0.5 for that team and +0.5 for the receiving team, a saved goal results in +0.5 for the attacking team and −0.5 for the defending team, a goal results in +10 for the scoring team and −10 for the opposing team, finally, a yellow card results in −3 for the committing team and +3 for the opposing team. The resulting game events time-series is displayed in Fig. 3.

3.3 Sentiment Transfer

3.3.1 Data Pre-processing

The sentiment analysis computation is performed using the Vader Sentiment analyzer for Python[2] [4]. The sentiment analysis provided scores indicating negative, neutral, positive sentiment of text, in addition to a combined compound sentiment score. We used the compound sentiment score for our analysis. The compound score provided by Vader is already normalized between −1 and 1.

[2] https://github.com/cjhutto/vaderSentiment.

We obtained the sentiment time-series by calculating the average sentiment at every minute of the game transcript and sentiments of the tweets (see Fig. 10. We calculated the first derivative of both the time-series and obtained two new times series, i.e., $X = x_1, ..., x_n$ and $Y = y_1, ..., y_n$, corresponding to the game transcript and tweet sentiments, respectively. We assigned a value 3 (i.e., x_i and $y_i = 3$) when the derivative was positive, a value 2 (i.e., x_i and $y_i = 2$) when the derivative was equal to 0, and a value 1 (i.e., x_i and $y_i = 1$) when the derivative was negative. The two time-series were used as inputs for the measurement of transfer entropy explained in the Sect. 1.

3.4 Transfer Entropy

As already mentioned in Sect. 1, we used transfer entropy to compute the information transfer from the sentiments of the events happening during the game to the sentiments of the tweets posted by people on.

Using this definition of transfer entropy, we defined a metric named Total Transfer Entropy (or TTE) to calculate the net information transfer between the time-series from the game transcript and tweets, which is given by Eq. 3.

$$TTE = T_{X \to Y} - T_{Y \to X} \tag{3}$$

where X is the sentiment time-series of the game transcript and Y is the sentiment time-series of the tweets. Therefore, TTE gives the aggregate information flow from the sentiments of the events happening during the game to the sentiments of the tweets posted by people on. This method has been used previously to infer leader-follower relationships in Zebrafish schools and direction of information transfer in Physarum polycephalum's membrane during food choice [1, 7]

We used TTE to measure the aggregate information flow from the game transcript to the sentiments for the duration of the game; and as well as during durations before and after the goal was scored (i.e., TTE between the sentiments before and after the 33rd minute of the game time). Additionally, TTE was measured from the game transcript to the sentiments of all tweets posted by people about the game on, and of people posting in languages English, Spanish, and German.

4 Results

4.1 Conversation Dynamics

Shown in Figs. 4, 5, and 6, are the time-series of volume, virality, and responsiveness of root tweets over the analyzed time period. Shaded in gray are the first and second

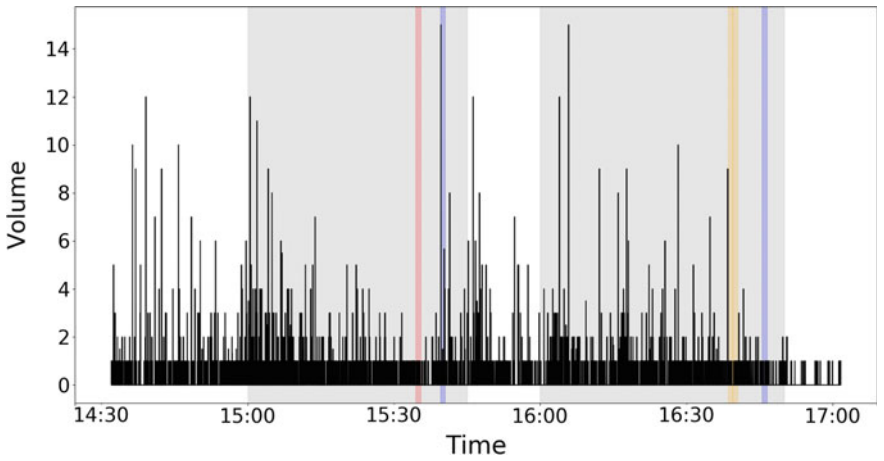


Fig. 4 The eventual volume of conversations originated by tweets over time. The gray areas depict the first and second half of the game. The red bar indicates when Mexico scored the only goal of the game, while the purple and yellow bars indicate when Germany and Mexico players received yellow cards, respectively

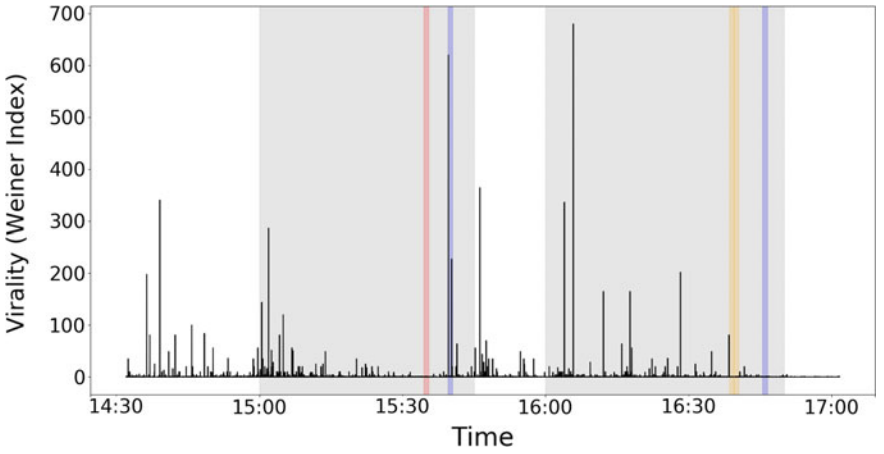


Fig. 5 The eventual virality (Weiner Index) of conversations originated by tweets over time. The gray areas depict the first and second half of the game. The red bar indicates when Mexico scored the only goal of the game, while the purple and yellow bars indicate when Germany and Mexico players received yellow cards, respectively

halves of the game. Additionally, the time of occurrence of the goal (red), two yellow cards to Germany (purple) and two consecutive yellow cards to Mexico (yellow) have been marked. As shown in Fig. 5, we observed that viral conversations initiated at the start of the first and second halves, and the end of the first half. In contrast, no such pattern to the game events is seen with responsiveness.

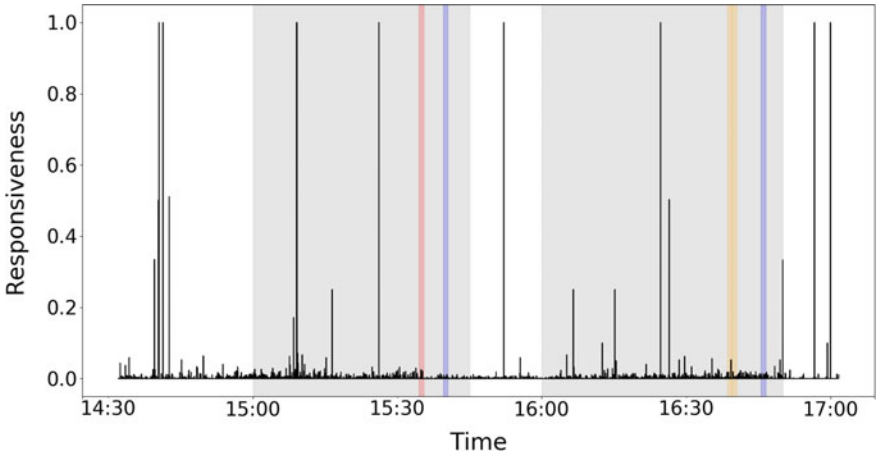


Fig. 6 The eventual responsiveness of conversations originated by tweets over time. The gray areas depict the first and second half of the game. The red bar indicates when Mexico scored the only goal of the game, while the purple and yellow bars indicate when Germany and Mexico players received yellow cards, respectively

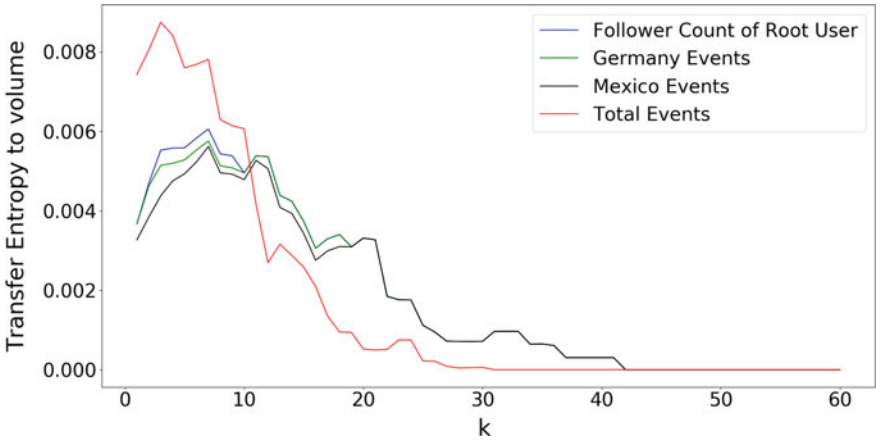


Fig. 7 Comparison of transfer entropy from external effects (Germany (blue), Mexico (green), and combined (black) game events) to the conversation volume time-series against the transfer entropy from the internal effect (the time-series of the number of followers of the conversation root user) over varying k

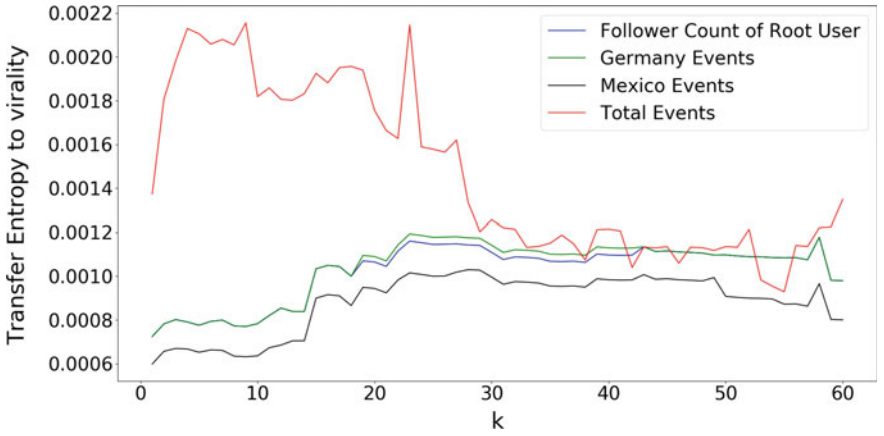


Fig. 8 Comparison of transfer entropy from external effects (Germany (blue), Mexico (green), and combined (black) game events) to the conversation virality time-series against the transfer entropy from the internal effect (the time-series of the number of followers of the conversation root user) over varying k

Figures 7, 8, and 9 plot the change in transfer entropy with the history length parameter, k, for the volume, virality, and responsiveness of root tweets, respectively, due to the performance of the two teams (Germany in blue, Mexico in green, and total in black). Additionally, the transfer entropy change with k for the three measures by the endogenous factor, number of followers of the user posting the root tweet, is also plotted (red) for reference. It can be seen that there is a generally higher transfer entropy from the follower count of the root creators, than the team performance for volume and virality. However, when considering responsiveness the transfer entropy results from the game events are generally higher than from the follower count series of the conversation originators.

Finally, we consider the history lengths at which the maximum transfer entropy values from the three-game event time-series to the three conversation dynamics time-series were reported. The transfer entropy from all three-game event series to conversation volume was maxmimal at a history length of $k = 7$ s; i.e., the best estimate of information transfer from the game events to the conversation volume occurred in the past 7 s. The transfer entropy from all three-game event series to conversation virality was maximal at a history length of approximately $22s < k < 29$ s, i.e., the best estimate of information transfer from the game events to the conversation virality occurred within the past 22–29 s. However, the maximal transfer entropy from the game events of the two teams to conversation responsiveness was different. Transfer entropy from the Mexican team events to conversation responsiveness was maximal at $21 < k < 27$, while the transfer entropy from the German team events to conversation responsiveness was maximal at $31 < k < 33$; i.e., the best estimate of information transfer from the game events by the Mexican team to the conversation

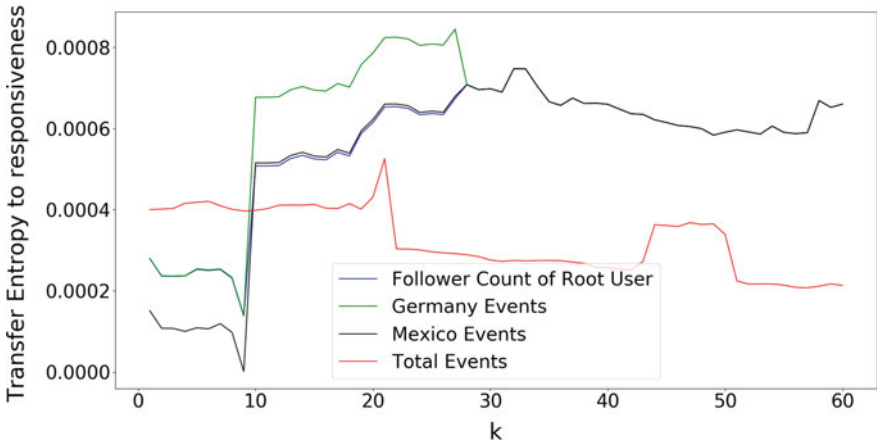


Fig. 9 Comparison of transfer entropy from external effects (Germany (blue), Mexico (green), and combined (black) game events) to the conversation responsiveness time-series against the transfer entropy from the internal effect (the time-series of the number of followers of the conversation root user) over varying k

dynamics occurred in the past 21–27 s, and that by German team occurred in the past 31–33 s.

4.2 Sentiment Transfer

The sentiment time-series of the game transcript and tweets are shown in Fig. 10. The sentiment of the tweets generally stayed in the positive region whereas the sentiments were equally spread across positive and negative regions of the transcript sentiment time-series of the transcript.

4.3 Total Transfer Entropy

Several relationships were found between the sentiment of the commentary and the endogenous sentiment of the Twitter discussions over time, which are summarized in Fig. 11 and discussed below.

4.3.1 Entire Population of People Posting Tweets About the Game

The TTE from the sentiment time-series of the transcript to the sentiment time-series of all tweets is shown in the first column (All Languages) of Fig. 11. The overall

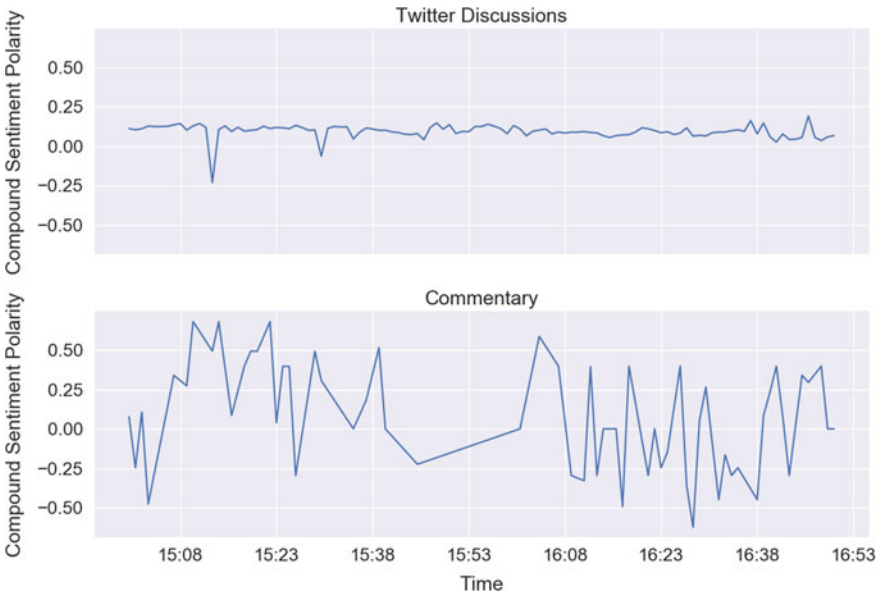


Fig. 10 Sentiment time-series of **a** transcript and **b** tweets posted when the game was in progress

TTE was maximum at $k = 6$ with $TTE = 0.1386$ bits, whereas TTE is maximum at $k = 6$ ($TTE = 0.1379$ bits) before and $k = 4$ ($TTE = 0.0884$ bits) after the goal was scored, respectively. This provides evidence that rate of information flow from the exogenous events of the game into the online discussions was quickened after the goal was scored.

4.3.2 Tweets Posted in Different Languages

Column 2–4 of Fig. 11 presents the TTE observed when the sentiment time-series were categorized by the languages English, Spanish, and German. Overall, *TTE* was maximum at $k = 5$, $k = 3$, and $k = 7$ for English, Spanish, and German speaking fans, respectively. The maximum information flow from the exogenous events into each of these groups was $TTE = 0.0851$, $TTE = 0.0829$, and $TTE = 0.1426$ bits, respectively.

Interestingly, we observe a clear shift in the rate of information flow after the goal is scored for each group. The k for which English and Spanish speaking groups have maximum *TTE* increases after the goal, from $k = 2$ to $k = 7$ for English and $k = 3$ to $k = 5$, indicating that the information flow from commentary sentiment to conversation sentiment was slowed. In contrast, k for maximum *TTE* dropped from $k = 6$ to $k = 5$, with lower values of k generally giving higher *TTE* values after the goal was scored. This indicated that the sentiment of German speaking fans was more sensitive to the sentiment of the commentary after the goal was scored. For German

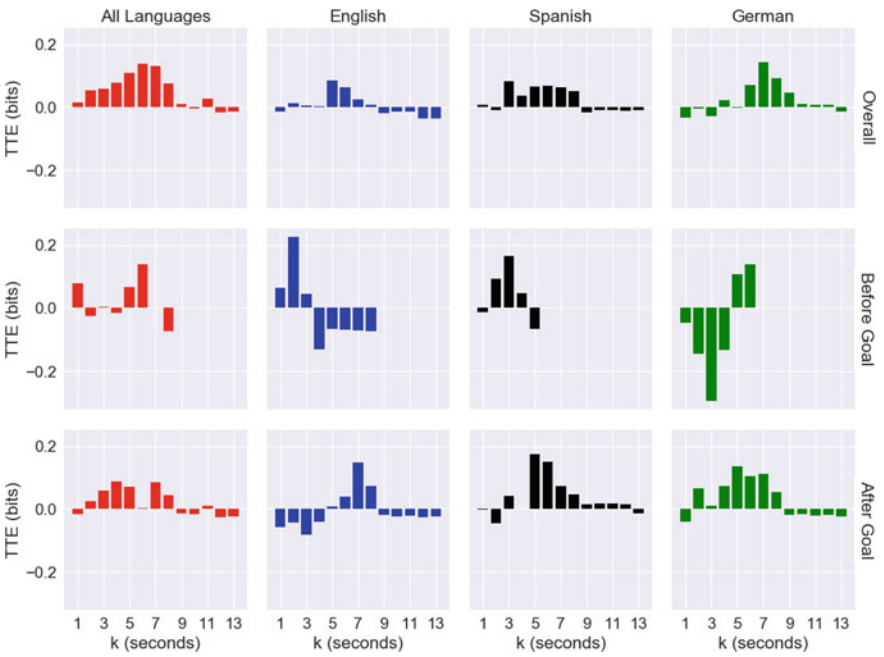


Fig. 11 *TTE* versus k categorized by language (columns), before and after the goal was scored by Mexico (rows). The goal clearly caused a shift in the value of k for which maximum *TTE* was observed for all three languages. There is an increased sentimental resilience shown by English and Spanish speakers and a reduced sentimental resilience shown by German speakers following the goal by Mexico. This is indicated by the increase in k for maximal *TTE* for English and Spanish speaking groups and a decrease for the German speaking group, respectively

speaking fans, before the goal was scored, for $k <= 4$, *TTE* was generally less than 0. This could indicate that there was an increase in uncertainty when considering this history length. Further analysis is required to establish this.

Overall, assuming Spanish speaking fans supported the Mexico team and German speaking fans support the German team, these results indicate that following the exogenous shock, the sentiment of the group that was advantaged was more resilient and less susceptible to exogenous events, while the sentiment group that was disadvantaged was more sensitive and susceptible to exogenous events.

5 Conclusion

This paper explores a case where the discussions within and between a polarized population of two competing groups are affected by exogenous events. We consider the case of Twitter conversations and commentary on the Mexico versus Germany FIFA 2018 qualifier game, in which a single goal was scored by the Mexican team.

This context presents us a unique situation where two polarized competing groups can be studied in isolation. We investigate the effect that game events have on the conversation dynamics, namely, the volume, virality, and the responsiveness of the conversations related to the game. Further, we investigate the emotional influence that the commentary had on the users tweeting about the game.

We discover that while volume and virality were less influenced by the game events and more by the popularity of the root users, responsiveness to conversations was more influenced by the game events.

Interestingly, we find that following the exogenous shock, the advantaged group showed increased emotional resilience to further exogenous events. In contrast, the disadvantaged group was left overall more emotionally susceptible to further exogenous events. These results imply that endogenously grown competitive discussions between two polarized groups that are emotionally invested in a particular topic may be used to manipulate and, possibly, alter the stability of one of the groups involved, while increasing the resilience of the other. If taken to its extreme, these results could serve as a warning to policymakers against the potential harm that could be caused to a population through targeted exogenous events focusing on a topic on which both groups are highly emotionally invested.

Acknowledgements The authors would like to thank the Complex Systems Summer School 2018 conducted by the Santa Fe Institute, at which this project was conducted and without which this collaboration would not have been possible.

References

1. Butail, S., Mwaffo, V., Porfiri, M.: Model-free information-theoretic approach to infer leadership in pairs of zebrafish. Phys. Rev. E **93**(4) (2016)
2. Choi, D., Han, J., Chung, T., Ahn, Y.-Y., Chun, B.-G., Kwon, T.T.: Characterizing conversation patterns in reddit: From the perspectives of content properties and user participation behaviors. In: Proceedings of the 2015 ACM on Conference on Online Social Networks, pp. 233–243 (2015)
3. Goel, S., Anderson, A., Hofman, J., Watts, D.J.: The structural virality of online diffusion. Manag. Sci. **62**(1), 180–196 (2015)
4. Hutto, C. J., Gilbert, E.: Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14) (2014). http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf
5. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: the good the bad and the omg. In: The International AAAI Conference on Weblogs and Social (2011)
6. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
7. Ray, S.K., Valentini, G., Shah, P., Haque, A., Reid, C.R., Weber, G.F., Garnier, S.: Information transfer during food choice in the slime mold physarum polycephalum. Front. Ecol. Evol. **7**, 67 (2019)
8. Saif, H., He, Y., Alani, H.: Semantic sentiment analysis of twitter. In: International Semantic Web Conference, pp. 508–524 (2012)
9. Schreiber, T.: Measuring information transfer. Phys. Rev. Lett. **85**(2), 461 (2000)

10. Sehgal, D., Agarwal, A.K.: Real-time sentiment analysis of big data applications using twitter data with hadoop framework. In: Soft Computing: Theories and Applications, pp. 765–772. Springer (2018)
11. Wiener, H.: Structural determination of paraffin boiling points. J. Am. Chem. Soc. **69**(1), 17–20 (1947)
12. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354 (2005)

The Cat and Mouse of Getting Around the Law

Daria Roithmayr, Justin Chin, Fei Fang, and Bruce Levin

Abstract Actors in certain fields of regulation—tax shelters, payday lending, cybersecurity—play a costly game of cat-and-mouse. When a new regulation is enacted, those who are regulated quickly innovate to avoid complying, and in turn, regulators must enact a new regulation. Payday lenders, for example, have responded to regulation by affiliating with Indian tribes to claim immunity from regulation. Drawing from work on drug resistance, we develop a non-equilibrium population dynamics model of regulatory resistance, with an eye towards advising regulators about how to optimize the speed and strength of regulation responding to innovators. Counterintuitively, we find that in certain conditions, a regulator's best option is to weaken or slow the regulatory response to an emerging workaround. If coupled with a fast response, weakening regulation can reduce the incentive for actors to innovate. Likewise, if coupled with a strong response, slowing the regulatory response can optimally extend the duration of each regulation and still limit the evolution of resistance. These findings demonstrate that math modeling can help legal policymakers and regulators to better understand and take into account the downstream dynamic consequences of their choices and to optimize accordingly.

D. Roithmayr (✉) · J. Chin
USC Gould School of Law, Los Angeles, CA, USA
e-mail: droithmayr@law.usc.edu

J. Chin
e-mail: jchin@law.usc.edu

F. Fang
Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: fang@cmu.edu

B. Levin
Emory University, Atlanta, GA, USA
e-mail: blevin@emory.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_6

1 Introduction

In many fields of law, actors are constantly hunting for new ways to escape regulation. Apple directs as much of its innovative energy towards tax avoidance as it does towards new products. Auto manufacturers import cars with backseats and then strip them out once they are in the country to gain a regulatory advantage. Uber hackers do constant battle with Uber in an attempt to game the digital platform and defraud the company.

Regulatory resistance is costly. Both regulators and regulated lenders spend millions of dollars trying to outcompete each other, neither side obtaining a lasting advantage. Payday lenders have been especially creative in coming up with loopholes, largely because they can innovate around regulation with the stroke of a pen to create new loopholes.

Most recently, in response to a usury limit on loans, Maryland payday lenders have partnered with Indian tribes to claim immunity from all regulations. State regulators have now had to re-regulate to address this latest innovation.

There is limited legal literature on the arms race of innovation between regulators and those they regulate. One study on tax regulation [1] models the regulatory arms race as a static model but does discuss endogenous innovation.

Practical remedies to address costly regulatory resistance are also few and far between. Some legal scholars have recommended the adoption of general anti-abuse rules meant to penalize attempts to innovate around legal regulation. These rules penalize innovative workarounds if they are undertaken in bad faith for the sole purpose of avoiding regulation [2].

Practitioners and scholars alike find these remedies incomplete, in part because innovators can engage in prohibited behavior even during the relatively short window of time that it takes for an agency or a court to declare that the workaround was undertaken with bad intent [3]. In addition, general standards against loopholes are vulnerable to arguments that they are too vague and/or that they stifle desirable innovation [2, 4].

In this paper, we analyze the cat-and-mouse dynamics of getting around the law, with an eye towards advising regulators on how best to manage innovators. Insight comes from an unexpected place: recent population dynamics models of drug resistance. As with regulatory cat-and-mouse, bacteria can mutate to get around constraining effects of antibiotic "regulation." Thanks to population dynamics models, the public health community now understands that overuse of antibiotic treatment can actually speed up the emergence of drug resistance and the loss of drug efficacy.

Borrowing from these models, we develop a simple coupled differential equation model of the cat-and-mouse game. We tailor the model for legal regulation, with the goal of helping a regulator optimize payday lending regulatory strategy. In our model, regulators choose how fast and how strong to make regulation, and lenders innovate with some probability of discovering a strategy that gets around the regulation. Using synthetic data, we optimize to find strategies that will achieve optimal results, which

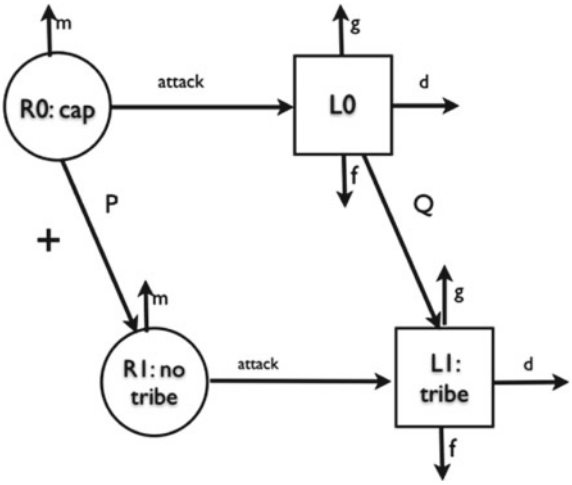


Fig. 1 Regulatory resistance model

we define as minimizing both the number of payday lenders and the frequency of innovation.

As detailed at more length in the Online Supplement, we first develop a toy population dynamics model to describe the dynamics of regulatory resistance, shown here below in Fig. 1.

Unique to legal regulation, this model maps the relationship between a single regulator and a population of actors, here lenders, who develop resistance to the regulation by innovating around the regulation. As will be specified mathematically below, over time, both the regulator and the lenders change over time. The single regulator's strategy becomes more mature (as court cases clarify the regulation's power and scope) and therefore more effective at regulating. The growth rate of the population of lenders also changes, as more lenders enter and exit.

At the outset, the regulator adopts a strategy (here, an interest rate cap) to "attack" the population of lenders. In response, some lenders "die" or exit the system, but some mutate to adopt a new strategy (here, affiliating with an Indian tribe). In turn, as the tribal lenders grow, the regulator must now adopt a new strategy (here, a regulation prohibiting tribal lenders), which in turn "attacks" the mutant lenders.

Developed from this simple predator–prey model, Eqs. 1 and 2 are the coupled differential equations that describe the dynamic patterns of resistance between the Regulator and Lenders.

In (1), lenders using the strategy of generation i grow logistically at rate g to carrying capacity K and fail to survive at rate f. When the regulator enacts a new regulation, lenders using that strategy die off at rate d under regulatory pressure; the variable d is a function of the regulator's choices about how strong to make the regulation. In response to such pressure, lenders using strategy replace it with some probability Q to use mutant strategies that escape regulation.

$$\frac{dL_{\mathrm{i}}}{dt} = \left[L_{\mathrm{i}} g\left(1 - \frac{L_T}{K}\right)\right] - (L_{\mathrm{i}} f) - (L_{\mathrm{i}} R_{\mathrm{i}} d) + \left[L_{\mathrm{i}-1}\left(Q_{\mathrm{i}-1,\mathrm{i}}\right) - L_{\mathrm{i}}\left(Q_{\mathrm{i},\mathrm{i}+1}\right)\right] \quad (1)$$

In (2), a single regulator uses strategy (which regulates lender strategy); this regulation matures logistically from 0 to full power 1, at rate m. In response to lender innovation, strategy is replaced with some probability P by a mutant strategy; the variable P specifies a function of the regulator's choices about how fast to respond to lender innovation.

$$\frac{dR_i}{dt} = P_{i-1,i} + R_i m_i \left(1 - \frac{R_i}{\mathrm{K_R}}\right) \quad (2)$$

We now simulate the model with a full range of synthetic data for key variables to see how varying the speed and strength of the regulatory response affects regulatory performance. To assess performance, we score regulatory choices by how well they (i) minimize the average number of payday lenders and (ii) minimize the average frequency of re-regulation (regulating to address a new innovation) in a given period of time. We optimize the regulator's strategic choices by scoring regulatory choices on the objective function and choosing those choices that score the highest.

The objective value function seeks to minimize two performance criteria: the average frequency of re-regulation in a given time period and the average number of payday lenders measured in the same time period:

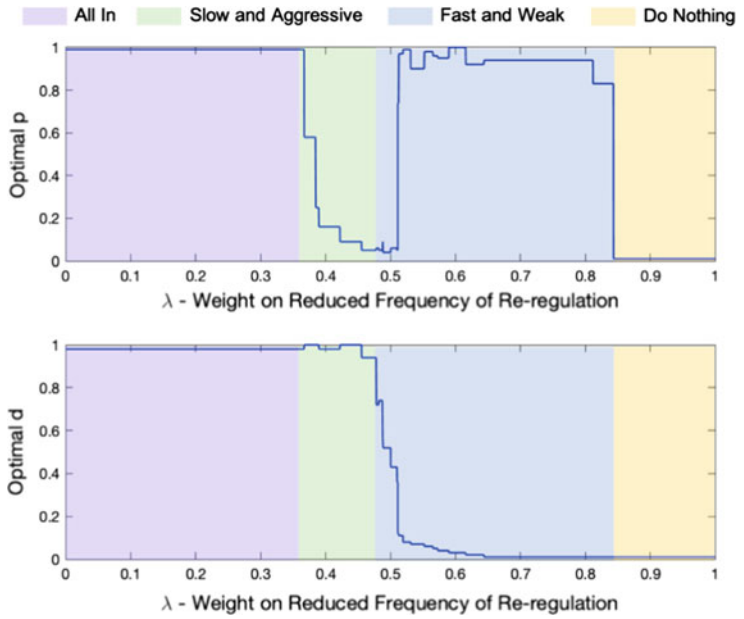
$$\mathrm{J}(\lambda) = \lambda(\text{Avg. Reregulation Frequency}) + (1 - \lambda)(\text{ Avg. Lenders}) \quad (3)$$

$$\mathrm{J}(\lambda) = \lambda(\text{avgIR}) + (1 - \lambda)(\text{avgLT})$$

In (3), we include a term that describes the "weight" of the regulator's relative priority for each of the two performance criteria in the function. While λ represents the weight of reducing the frequency of re-regulation, $1 - \lambda$, represents the weight of reducing the average number of payday lenders. For a regulator who cares only about reducing the frequency of re-regulation, $\lambda = 1$. For a regulator who cares only about reducing the number of payday lenders, $\lambda = 0$. Optimizing will depend very much on regulator priority, as we observe below.

2 Results

Surprisingly, optimal regulator choices about the speed and strength of re-regulation vary widely and in complex ways, depending on regulator priority. Figure 2 displays the regulator's optimal choices on speed (p) and strength of re-regulation (d) as a function of the relative weight a regulator puts on minimizing the frequency of re-regulating.



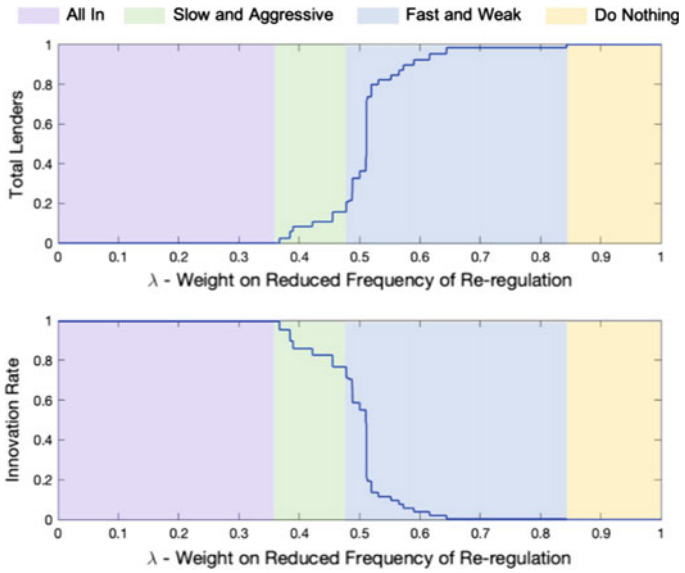
dt= .02, L1 = 10, K=800, g =2, f =0.4, s=0.1, cc= 0.1

Fig. 2 Optimal p and d across regulator priorities

Figure 2 helps us to understand why optimal strategies vary with regulator priorities. For example, as the regulator begins to care more about reducing re-regulation frequency, moving from left to right on the x-axis, the optimal speed for re-regulation (P) drops. As we reach $\lambda = 0.5$, the point at which the regulator cares equally about reducing lenders and reducing re- regulation, the optimal speed for re-regulation suddenly rises, quite dramatically.

Figure 2 identifies four key optimal strategies that regulators should pursue, depending on their priorities. These four strategies include Fast and Strong ("All In"), Slow and Strong ("Slow and Aggressive"), Fast and Weak, and Slow and Weak ("Do Nothing"). Each of these strategies is optimal depending on the way in which the regulator weights the importance of reducing payday lenders relative to the importance of reducing the number of times the regulator must go back to the drawing board to re-regulate.

Figure 3 depicts the performance "scores" of these four optimal strategies. Here, we measure the performance of the regulatory strategies on the two regulatory objectives: reducing payday lender numbers and reducing the frequency of innovative re-regulation (on the y axis). Again, we observe the very steep slope of change in both panels, indicating dramatic differences in lender numbers and re-regulation frequency over a small difference in regulator priorities. Below, we discuss the reason for these dramatic differences, and more generally the key strategies that regulators can pursue to optimize regulation.



dt= .02, L1 = 10, K=800, g =2, f =0.4, s=0.1, cc= 0.1

Fig. 3 Lender number and innovation rates across regulator priorities

3 Optimizing Strategies

Drawing from Figs. 2 and 3, we can now identify four optimal strategies that work for a range of regulator priorities. In consumer regulation, for example, different regulators have different priorities—tax regulators might want to prioritize eliminating the prohibited behavior while other regulators with fewer resources might want to balance limiting prohibited behavior with limiting the expense of re-regulation.

Optimizing depends wholly on regulatory priority. A regulator who prioritizes reducing payday lenders will optimize by re-regulating with maximum speed and with maximally strong regulation. In contrast, a regulator who cares equally about eliminating prohibited behavior and minimizing costly re-regulation will re-regulate quickly but with relatively weak regulation. We explore the way in which optimal strategies change as the regulator becomes increasingly focused on limiting costly re-regulation.

3.1 Fast and Strong ("All in")

For the regulator who prioritizes eliminating prohibited behavior, the optimal strategy is a fast and strong (or "all in") re-regulation strategy. This strategy combines strong regulation and a quick response to a new lender loophole. This strategy is optimal

for the regulator who has prioritized reducing lender numbers ($\lambda = 0 - 0.38$) and who is not willing to reduce that priority in order to lower the cost of re-regulating. For example, cybersecurity regulators who are tasked with protecting voting systems or the country's infrastructure might focus exclusively on reducing the number of cyber-attackers and attach no weight to minimizing the cost of revising regulation. The Fast and Strong strategy is also optimal for the regulator for whom re-regulation is costless or very low cost.

For these regulators, the optimal strategy is to adopt the strongest regulation possible (maximum d) and to re-regulate as quickly as possible to a new regulation (maximum p). As Fig. 3 shows, the all in strategy reduces lending to a minimum, while leaving the frequency of re-regulation at a maximum. Though strong regulation induces the evolution of resistance, the regulator takes care of resistant innovators by adapting quickly and paying the cost of re-regulation.

3.2 Slow and Strong ("Slow and Aggressive")

Slow and Strong generates a small increase in lenders and a small decrease in innovation rate relative to the Fast and Strong strategy. For the regulator who prioritizes limiting prohibited behavior but still cares about limiting costly re-regulation, the optimal strategy is slow and strong re-regulation. Slowing the response to regulation gives the regulator more time with each regulation and requires less frequent re-regulation. Though slowing the response tends to promote an increase in the number of innovators, strong regulation helps to keep these innovators in check once the regulator actually responds. In terms of performance, Fig. 3 shows that Slow and Strong generates a small increase in lenders and a small decrease in innovation rate relative to the Fast and Strong strategy.

3.3 Fast and Weak

For the regulator who cares equally about bad behavior and costly re-regulation, the optimal strategy is a fast and weak re-regulation. The Fast and Weak strategy combines very weak regulation with the maximum speed of regulatory response. This strategy works best for a regulator who prioritizes both minimizing lender numbers and reducing the costly re-regulation frequency, but who cares slightly more about reducing the number of times the regulator must go back to the drawing board ($\lambda = 0.48 - 0.6$). This strategy might be appropriate for a small state with big resource constraints.

Fast and Weak works by minimizing the strength of regulation, which reduces the incentive for lenders to innovate. Reducing the strength of regulation enables lenders to continue to pursue their standard business model, and accordingly, lenders will be less invested in innovating around the law. As noted earlier, a payday lender subject

only to an APR disclosure requirement will be less motivated to find a loophole than the lender facing an interest rate cap. To keep the occasional lender innovation in check, a fast re-regulation response (high p) enables the regulator to "catch" innovative lenders before others copy them.

In terms of performance, Fig. 3 shows that Fast and Weak generates a significant increase in payday lending (rising steeply to near maximum) and a dramatic fall in re-regulation frequency (falling steeply to near minimum).

3.4 Slow and Weak ("Do Nothing")

For the regulator who prioritizes limiting costly re-regulation, the optimal strategy is re-regulating slowly and with weak regulation. Slow and Weak can be thought of as a do-nothing strategy. This strategy minimizes both the strength of regulation and the speed of the regulatory response. Slow and Weak is optimal for regulators who strongly prioritize reducing the frequency of costly re-regulation ($\lambda = 0.58 - 1$) relative to the priority of reducing lender numbers. This strategy is appropriate in theory for the state in which the cost of maintaining re-regulation outweighs the benefits of maintaining re-regulation. Alternatively, this strategy is appropriate for a state that prioritizes providing access to funds for credit-constrained borrowers and is thus un-interested in lowering lender numbers.

This strategy works by offering little in the way of regulatory pressure, letting market forces regulate the number of lenders, to limit the population to an upper bound. In terms of performance, Fig. 3 shows that a Slow and Weak strategy generates a maximum number of payday lenders in a given period of time and minimizes the frequency with which regulators must re-regulate.

Figure 2 and 3 displays abrupt switches between strategies, which appear to have a "bang-bang" like property associated with linear optimal control. We observe that the best solutions for the regulator are mostly on the boundaries, maximizing either speed or strength, rather than combining half-measures of both speed and strength. The switchover points in Fig. 2 reflect (i) the point at which reducing speed produces no more improvement in reducing costly re-regulation, and (ii) the point at which minimizing both speed and strength are needed to further limit re-regulation.

These simulated results demonstrate the potential power of a dynamic approach to regulatory optimization. Our toy model enables regulators and policymakers to see downstream consequences of regulatory choices that are otherwise hard to predict. Regulators can use such insight to better manage evolving loophole innovations. In future work, we plan to offer advice to lenders as well.

4 Discussion

Biology offers models that can shed light on these dynamics: models resembling ours have described the co-evolutionary relationship between predator and prey (often parasite and host), the antigenic arms race between the immune system and pathogens, and the evolution of drug resistance in cancer and more generally [5–8].

Such frameworks, which endogenize the role of the regulator and allow the modeling of non-equilibrium behavior, are potentially of immense importance to legal regulators. Importantly, this model sensitizes regulators to the potential negative consequences of responding quickly and aggressively to actors who are "hacking" legal regulation.

Some evidence from the real world indicates the potential impact of the four key strategies we identify. One or two regulators have begun to experiment with strategies that vaguely resemble the Fast and Weak strategy. For example, the Consumer Finance Protection Board (CFPB) has considered offering lenders the option of an eighteen-month window free of regulation in order to experiment with lending strategies that meet certain guidelines outlined by the CFPB. In exchange, lenders would give the CFPB historical data on lending levels and borrower profiles [9].

Likewise, scholars have argued for sunset provisions or regular review as a means to delay or systematize regulatory responses when waiting would be better than acting [10]. The value of these strategies becomes more apparent when we set them against the framework of a dynamic understanding of their downstream consequences.

5 Conclusion

This paper investigates the cat-and-mouse game of regulatory avoidance. Borrowing from models of pathogen resistance, we tailored a simple population dynamics model to advise regulators on optimal strategies. Regulators often mistakenly assume that the best re-regulation is fast and strong [9, 10]. This model suggests that for some regulators, fast and strong strategies are counterproductive and contribute to the emergence of "resistance."

Data are key in developing effective strategies for regulating behavior in the face of ongoing regulatory resistance. Modeling may help regulators to be far more data-driven in choosing optimal strategies. Biologists have for some decades used data in conjunction with mathematical models to design effective strategies that minimize the emergence of pathogenic resistance. This approach may prove equally effective in the world of legal regulation, particularly as data collection improves.

More generally, this project underscores the importance of evolutionary dynamics in law. Dynamic modeling enables policymakers to investigate the unanticipated consequences of their interventions. Understanding human activity as in transition from one equilibrium to the next, and often not in equilibrium at all, helps scholars to ask different questions about the nature of legal and social change. Analyzing the

role that innovation and novelty play in shaping the evolution of law and the social world helps us to begin charting a path to a new and better world.

(Online Supplement at http://www.dariaroithmayr.com/pdfs/Online-Supplement.pdf).

References

1. Hines, J.: On the Timeliness of Tax Reform. J Publ. Econ. **88**, 1043 (2003)
2. Burk, D.L.: Perverse innovation. Wm Mary L. Rev. **58**, 1 (2016)
3. Kleinbard, E.: Interview with Former Chief of Staff of U.S. Congress' Joint Committee on Taxation (2017)
4. Katz, L.: A theory of loopholes. J. Legal Stud. **39**(1), (2010)
5. Abel zur Wiesch, P., Kouyos, R., Abel, S., Viechtbauer, W., Bonhoeffer, S.: Cycling empirical antibiotic therapy in hospitals: meta-analysis and models. PLoS Pathol. **10**(6), e1004225. https://doi.org/10.1371/journal.ppat.1004225
6. Beardmore, R.E., Pena-Miller, R.: Rotating antibiotics selects optimally against resistance in theory, 7(3) Math. Biosci. Eng. **527**, (2010)
7. Levin, S.A.: Ecological issues related to the release of genetically modified organisms in the Environment in Introduction of Genetically Modified Organisms into the Environment. Mooney, H. A., and Bernardi, G. (eds.) (1990)
8. Read, A.F., Lynch, P.A., Thomas, M.B.: How to make evolution-proof insecticides. PLoS Biol. **7**, e10000058 (2009)
9. Ficklin, P.: Interview with general counsel, Consumer Financial Protection Bureau (2016)
10. Romano, R.: Regulating in the dark. J. Fin. Persp. **1**, 1 (2013)

The Degree-Dependent Threshold Model: Towards a Better Understanding of Opinion Dynamics on Online Social Networks

Ece Çiğdem Mutlu and Ivan Garibay

Abstract With the rapid growth of online social media, people become increasingly overwhelmed by the volume and the content of the information present in the environment. The fact that people express their opinions and feelings through social media channels, influence other people, and get influenced by them has led researchers from various disciplines to focus on understanding the mechanism of information and emotion contagion. The threshold model is currently one of the most common methods to capture the effect of people on others' opinions and emotions. Although many studies employ and try to improve upon the threshold model, the search for an appropriate threshold function for defining human behavior is an essential and yet an unattained quest. The definition of heterogeneity in thresholds of individuals is oftentimes poorly defined, which leads to the rather simplistic use of uniform and binary functions, albeit they are far from representing reality. In this study, we use Twitter data of size 30,704,025 tweets to mimic the adoption of a new opinion. Our results show that the threshold is not only correlated with the out-degree of nodes, which contradicts other studies, but also correlated with nodes' in-degree. Therefore, we simulated two cases in which thresholds are out-degree and in-degree dependent, separately. We concluded that the system is more likely to reach a consensus when thresholds are in-degree dependent; however, the time elapsed until all nodes fix their opinions is significantly higher in this case. Additionally, we did not observe a notable effect of mean-degree on either the average opinion or the fixation time of opinions for both cases, and increasing seed size has a negative effect on reaching a consensus. Although threshold heterogeneity has a slight influence on the average opinion, the positive effect of heterogeneity on reaching a consensus is more pronounced when thresholds are in-degree dependent.

E. Ç. Mutlu · I. Garibay (✉)
Complex Adaptive System Laboratory, University of Central Florida, Orlando, Florida, USA
e-mail: ivan.garibay@ucf.edu

E. Ç. Mutlu
e-mail: ece.mutlu@ucf.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_7

1 Introduction

While studying networks is not new to humanity, its focus has evolved from physical proximity and socio-economic-based to social media-based networks. This change is arguably the product of the fast-paced information flow that is engendered by the technological advances of the twenty-first century and the resulting impact on people's needs and lifestyles. The need to address the newly emerged phenomenon that people create, receive, and disseminate information on online social networks has amplified the interest in the field of network science. Indeed, network science applications have extended to the field of marketing [2, 6], sociology [9], political science [15], physics [14], economics [12], and biology [10], in attempting to reveal the interdependency between units of interest. For instance, the shift from traditional advertising to digital marketing applications, or the political campaigns being organized on social media channels, has allowed people's opinions to be voiced freely and with far-reaching consequences. This reciprocity in information flow, the increase in the volume of information received and sent, and the ease of relaying information have made it imperative that researchers and practitioners understand the dynamics of information and opinion formation, propagation, and exchange [1, 18, 19, 21].

In the mid-twentieth century, the field of sociology pioneered the development of information and opinion diffusion as a subject of study, which has remained relevant and popular to this day. One of the early studies is the Markovian linear threshold model introduced by Granovetter [5], which investigates the opinion dynamics of people. According to the threshold model, individuals adopt a new opinion only if a critical fraction of their neighbors have already adopted the new opinion. Granovetter suggests that the threshold of individuals can be different and are influenced by demographic and psychographic factors such as socio-economic status, education, age, and personality type. However, this heterogeneity among researchers is poorly defined, which leads to an extensive use of homogenous (uniform) [11, 16, 17] and binary [20] thresholds in many studies. Arguably, this assumption of homogenous or binary thresholds is an oversimplification of reality and may produce misleading results. To remedy this oversimplification and thereby provide a more holistic and accurate model, more complex threshold models such as tent-like function [22], truncated normal distribution function [7], or sigmoid function [3] are also used in the literature. Our Twitter data mining results show that the threshold of an individual for adopting a new opinion (retweeting a tweet) is affected either by his out-degree (number of followers) or his in-degree (number of following/followee). Some studies have already employed degree-dependent threshold models in explaining the dynamics of information diffusion [4, 8]; however, the degree dependency of an individual's threshold is associated only with his out-degree. Additionally, these studies have implemented threshold heterogeneity by using custom threshold functions, which renders the results less robust and less reliable. Therefore, we want to analyze the sensitivity of information diffusion dynamics to in-degree and out-degree dependencies of thresholds. Another purpose of this study is to under-

stand how threshold heterogeneity and network properties (seed size, mean-degree) affect information diffusion dynamics when thresholds are in-degree and out-degree dependent, separately.

The remaining part of this paper is structured as follows: First, we provide an overview of the Twitter data set and its subsequent analysis. Then, we describe the methodology that we used to generate networks, assign thresholds, and run the simulations in the "Method" section. We give the results of simulations in the "Simulation Results" section. Finally, we discuss the results and explain the contributions of this study in the "Conclusion" section.

2 Methods

2.1 Data Set and Twitter Analysis Results

The Twitter data set used for this study contains 30,704,025 tweets from the cybersecurity-related events from March 2016 to August 2017, of which 16,884,353 are retweets. We first collected follower and following counts of each user to relate the retweeting probability of users with the two aforementioned counts. We generated a matrix of which rows represent follower count clusters and columns represent following count clusters of all users in our data set (Fig. 1a). Then, we filtered users who have retweets only and generated the same matrix (Fig. 1b). Preliminary results show that majority of the users are clustered around the areas where follower and following counts are not extreme, and the matrix of retweeted users also shows a similar pattern unsurprisingly. Since retweeting probabilities of users in each cluster, i.e. the threshold for adopting a new opinion, are not clear from these matrices only, we calculated the element-wise division of these two matrices to figure out the ratio of the number of retweeters to the number of all users in each cluster. Results show that the retweeting probability of users who have relatively lower following count is higher, i.e. threshold of a node seems to be positively correlated with his out-degree. On the other hand, the effect of varying follower count on the retweeting probability is not obvious since the left-bottom of the matrix is empty (Fig. 1c). Therefore, we extracted the three most retweeted tweets (RT1, RT2, RT3) of retweet sizes 138,969, 58,546, and 57,280, respectively. We divided users into 8 clusters with respect to their follower (Fig. 2a–c) and following counts (Fig. 2d–f), independently rather than jointly clustering. For each cluster, we calculated the ratio of the number of users who retweeted RT1, RT2, or RT3 to the number of all users, respectively, as in Fig. 2c. The only difference is that instead of all retweeters, we just focused on retweeters of RT1, RT2, and RT3. Thus, we could prevent the masking effect of non-active users in the whole data set. The results show that both follower and following count have a negative effect on the retweeting probability of users. Furthermore, we applied a one-sided Chi-square test ($\alpha = 0.5$) to understand whether this decreasing pattern is statistically significant. We included

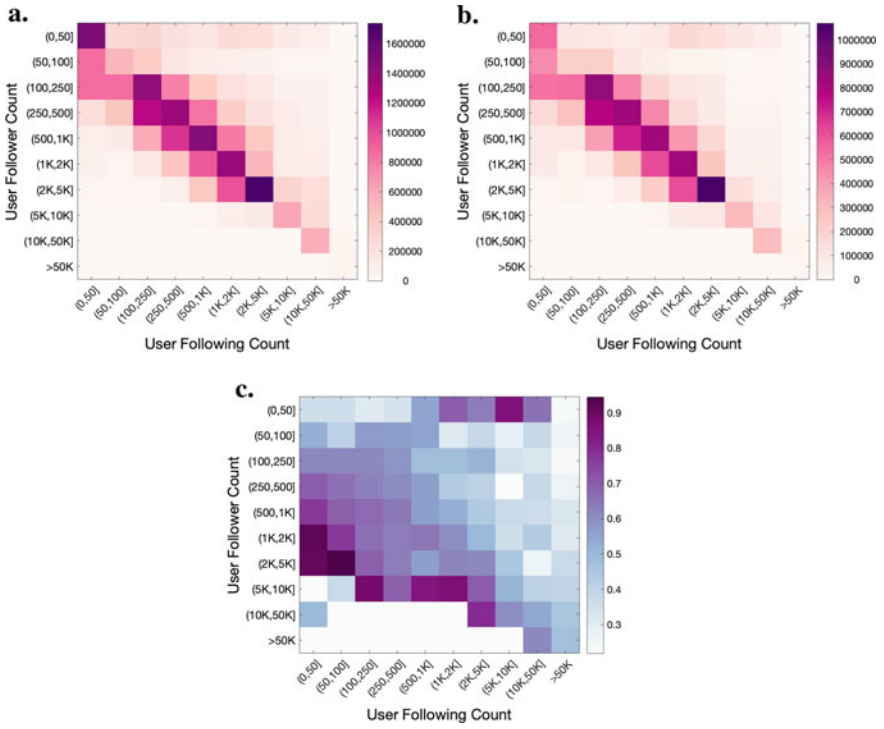


Fig. 1 **a** Number of users **b** Number of retweeters **c** The retweeting probability of users in each cluster-the element-wise ratio of number of retweeters in b to the number of users in **a**

relative χ^2 values if the retweeting ratio in the cluster is significantly higher than that of the next cluster (p-value is lower than 0.005). We observed that the retweeting probability decreases when follower count increases and this decreasing pattern is significant for almost all consecutive clusters. Nevertheless, the decrease between the consecutive clusters defined by following counts was significant only when the following counts are not high. This is probably because clustering users according to their follower and following counts with the same limits affect the results of test statistics notably, since distributions of follower counts and following counts of users are not similar in the data set, i.e. the 0.8 and 0.9 quantiles and the maximum of the user following counts are 1916, 3860, and 3,136,215; while those are 2332, 5639, and 94,833,565 for user follower counts. When we decrease the number of clusters from 8 to 3 ($(0, 1K]$, $(1K, 10K]$, $(>10K]$), we observed that the retweeting probability decreases when following count increases and this decreasing pattern is significant for all consecutive clusters ($\chi^2 = \{12602.8, 18087, 272.3\}$ for RT1, $\{2762.3, 1807.2, 52.0\}$ for RT2, and $\{1299.3, 987.1, 87.1\}$ for RT3). Thus, our data analysis shows that thresholds of individuals to accept a new opinion are positively correlated with their in-degree and out-degree.

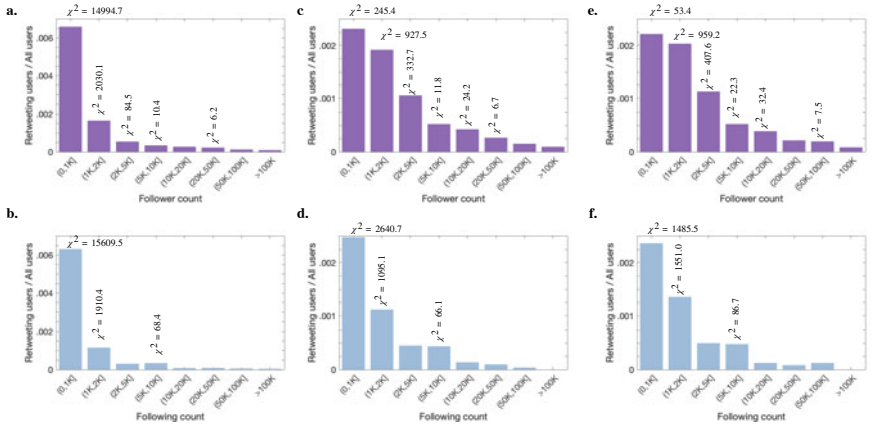


Fig. 2 The ratio of retweeters of RT1, RT2 and RT3 to all users when we cluster users according to the their follower counts and following counts independenty. Bar plots in **a** and **b** shows the results for RT1 users, **c** and **d** shows the results for RT2 users and **e** and **f** for RT3 users

2.2 Generating Networks

The main aim of this paper, as mentioned, is to investigate the effect of threshold heterogeneity on opinion spreading dynamics when thresholds are correlated with the degree distribution of the nodes in a network. For this purpose, we generated power-law distributed random numbers (x_i) to further assign them to the desired degree distribution of the network. To understand the effect of out-degree-dependent threshold and in-degree-dependent threshold on the dynamics of opinion spreading separately, we created two independent networks as:

(i) Out-degrees of the nodes (k_{out}) are power-law distributed and has the form $\sqrt{N}x^{\gamma}$ and in-degrees are kept constant (M_{in}).
(ii) In-degrees of the nodes (k_{in}) are power-law distributed and has the form $\sqrt{N}x^{\gamma}$ and out-degrees are kept constant (M_{out}).

Here, N denotes the number of nodes in the network (seed size) and $\gamma = 3$ for both cases for a fair comparison. Then, we added directed links between randomly selected node pairs (i, j) by employing configuration model [13] if $i \neq j$ and $k_{out} < x_i$ for i., $k_{in} < x_i$ for ii. This wiring process continued until all possible links are formed. In this network structure, self-edges are not allowed while multiple edges between the same node pairs are possible. Since total in-degree in the network should be equal to the total out-degree in the network, one can easily realize that the mean-degree of the network is equal to:

(i) Fixed in-degree (M_{in}) when out-degrees of the nodes are power-law distributed.
(ii) Fixed out-degree (M_{out}) when in-degrees of the nodes are power-law distributed.

2.3 Assigning Thresholds

After generating networks, we employed the degree-dependent threshold model by assigning the threshold of node i to accept a new opinion (ϕ_i) as correlated with:

(i) its out-degree when out-degrees are power-law distributed and in-degrees are constant in the network.
(ii) its in-degree when in-degrees are power-law distributed and out-degrees are constant in the network.

Since threshold heterogeneity is one of our main concerns in this study, we divided nodes into N_{th} groups by their ranks which can be obtained by sorting their

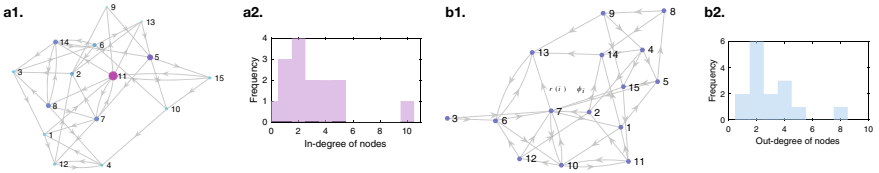
(i) out-degrees when out-degrees are power-law distributed and in-degrees are constant in the network.
(ii) in-degrees when in-degrees are power-law distributed and out-degrees are constant in the network.

Then, we assigned thresholds as evenly spaced N_{th} points between 0.5 and 1 to prevent the confounding effect of the mean-threshold, i.e. the average threshold is always constant as 0.75. Thus, increasing N_{th} yields more heterogeneity among thresholds of individuals.

$$\phi_i = \begin{cases} 0.5 & \text{if } r(i) \leq \frac{N}{N_{th}} \\ 0.5 + \frac{0.5}{N_{th}-1} & \text{if } \frac{N}{N_{th}} < r(i) \leq \frac{2N}{N_{th}} \\ \ldots & \ldots \\ 0.5 + \frac{0.5(N_{th}-2)}{N_{th}-1} & \text{if } \frac{(N_{th}-2)N}{N_{th}} < r(i) \leq \frac{(N_{th}-1)N}{N_{th}} \\ 1 & \text{if } \frac{(N_{th}-1)N}{N_{th}} < r(i) \leq N \end{cases}$$

where $r(i)$ represents the rank of the node when they are sorted according to their i. out-degree and ii. in-degree.

An example of network generation for two cases (i. and ii.), out-degrees and in-degrees, and relative threshold values of the nodes are shown in Fig. 3.



a3.

	out-degree	in-degree	$r(i)$	ϕ_i
1	2	3	8	0.75
2	3	3	6	0.75
3	2	3	9	0.75
4	2	3	10	0.75
5	5	3	2	1
6	3	3	7	0.75
7	4	3	3	1
8	4	3	4	1
9	1	3	14	0.5
10	2	3	11	0.5
11	8	3	1	1
12	2	3	12	0.5
13	1	3	15	0.5
14	4	3	5	1
15	2	3	13	0.5

b3.

	out-degree	in-degree	$r(i)$	ϕ_i
1	3	2	11	0.5
2	3	3	6	0.75
3	3	0	15	0.5
4	3	2	10	0.75
5	3	2	9	0.75
6	3	5	3	1
7	3	10	1	1
8	3	1	13	0.5
9	3	1	14	0.5
10	3	4	4	1
11	3	2	8	0.75
12	3	1	12	0.5
13	3	5	2	1
14	3	4	5	1
15	3	3	7	0.75

Fig. 3 The representation of network when $N = 15$ and **a1** out-degrees are power-law distributed and in-degrees are kept constant as $M_{in} = 3$, **b1** in-degrees are power-law distributed and out-degrees are kept constant as $M_{out} = 3$. Histogram plots of **a2** out-degrees in **a1**, **b2** in-degrees in **b1**. In addition to out-degree and in-degree of nodes, their ranks $r(i)$ and thresholds ϕ_i are also given in the table for **a3** the network in a1, **b3** the network in **b1**

2.4 Running Simulations

We initialized the opinions of individuals as a Bernoulli distributed random variable with an initial probability (p), i.e. the opinion of the node i (s_i) might equal to 1 with a probability p and equal to 0 with a probability $1 - p$. We assumed that the opinion change process is reversible; thus, individuals may change their opinions continuously rather than changing one time.

After generating the network, assigning thresholds, and initializing the opinions, we run the opinion change simulations. The process of updating their opinions is as follows:

1. Picking a node i randomly.
2. Calculating the weighted average of the opinions of its in-neighbors ($\bar{o}_i$). Here, weights are the multiple edges formed between node i and its neighbors.
3. Updating the opinion of node i (s_i) according to the criteria as follows:

   a. **if** $s_i = 0$ and $\bar{o}_i - s_i > \phi_i$,
   then $s_i = 1$ in the next step.
   b. **if** $s_i = 1$ and $\bar{o}_i - s_i < -\phi_i$,
   then $s_i = 0$ in the next step.

This Markovian chain is repeated until all possible opinion changes are made and individuals fix their opinion. We carried out all the simulations on MATLAB and repeated these simulations 10,000 times.

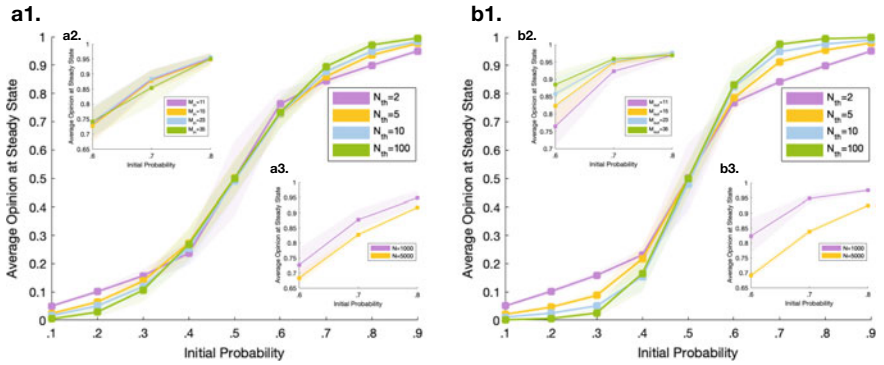


Fig. 4 Simulation result of average opinion at steady state as a function of initial probability (p) **a1** with varying threshold heterogeneity (N_{th}) when $N = 1000$ and $M_{in} = 15$, **a2** with varying in-degree (M_{in}) when $N = 1000$ and $N_{th} = 10$ and **a3** with varying seed size (N) when $M_{in} = 15$ and $N_{th} = 10$ if thresholds are out-degree dependent and out-degrees are power-law distributed. Additionally, simulation result of average opinion at steady state as a function of initial probability (p) **b1** with varying threshold heterogeneity (N_{th}) when $N = 1000$ and $M_{in} = 15$, **b2** with varying out-degree (M_{out}) when $N = 1000$ and $N_{th} = 10$, and **b3** as a function of initial probability (p) when $M_{out} = 15$ and $N_{th} = 10$ if thresholds are in-degree dependent and in-degrees are power-law distributed

3 Simulation Results

In the current study, we first aim to analyze the effect of in-degree and out-degree dependence of thresholds on the average opinion at steady state ($\bar{s}$). Therefore, after all the individuals fix their opinions in a network, we averaged the opinions of them by using the equation below:

$$\bar{s} = \frac{1}{N} \sum_{i}^{N} s_i \tag{1}$$

where s_i is the opinion of node i at a steady state. We conducted our simulations to measure $\bar{s}$ as a function of initial probability (p) with varying mean-degree ($M_{\text{in}}/M_{\text{out}}$), seed size ($N$), and threshold heterogeneity N_{th}. Line plots in Fig. 4 represent the expected value of 10,000 Monte Carlo simulations, and shaded areas with the same colors denote the relative one standard deviation from the expected value of these simulations. Here, Fig. 4a1–a3 shows the simulation results when thresholds are out-degree dependent and out-degrees are power-law distributed while in-degrees are kept constant. Figure 4b1–b3, on the other hand, indicates the simulation results when thresholds are in-degree dependent and in-degrees are power-law distributed while out-degrees are kept constant.

Figure 5, on the other hand, shows the time elapsed until all individuals fix their opinion (t_f) as a function of p with varying $M_{\text{in}}/M_{\text{out}}$ and varying N_{th}. We did not

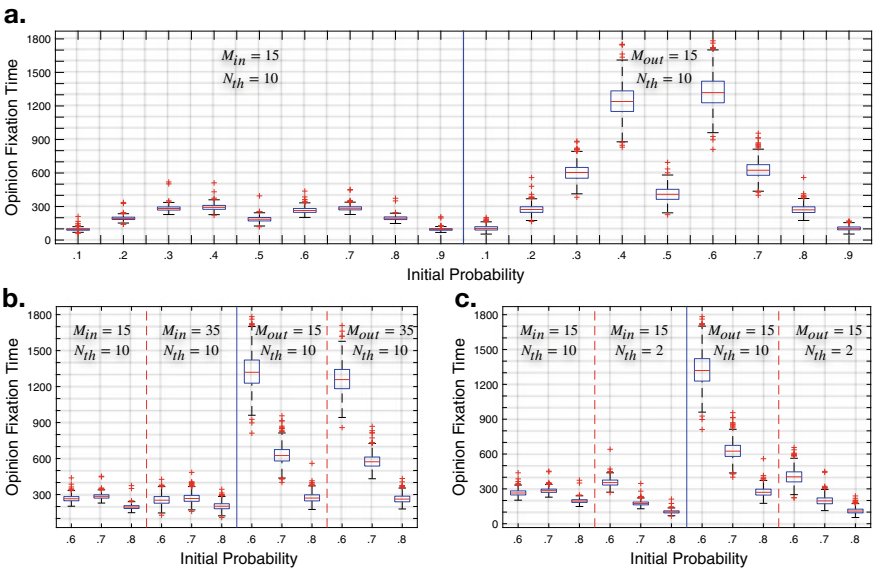


Fig. 5 The comparison of fixation time of individuals as a function of initial probability **a** when thresholds are out-degree dependent (left) and in-degree dependent (right) **b** with varying mean-degree (M_{in}/M_{out}) when thresholds are out-degree dependent (left) and in-degree dependent (right) **c** with varying threshold heterogeneity (N_{th}) when thresholds are out-degree dependent (left) and in-degree dependent (right)

simulate the effect of varying N on t_f since it is obvious that increasing the seed size causes more deviation in the opinions and increases t_f.

Figure 4a1 and b1 shows $\bar{s}$ as a function of p at various N_{th} values. Since the standard deviation of the simulations are highest in the range $0.35 \lesssim p \lesssim 0.65$, we especially focus on the results when $p \lesssim 0.35$ and $p \gtrsim 0.65$. In general, the system is more likely to reach a consensus when thresholds are in-degree dependent, and there is a clear asymmetry before and after $p = 0.5$ in both cases. Therefore, we just focused on the region $0.6 \leq p \leq 0.8$ for further analyses. Although threshold heterogeneity of nodes in the system has a slight effect on the resulting average opinion when thresholds are out-degree dependent, we can say that the probability that the system reaches a consensus increases as threshold heterogeneity increases, and this increase is more pronounced when thresholds are in-degree dependent. E.g. $\bar{s} = 0.8412$ when $N_{\text{th}} = 2$, while $\bar{s} = 0.9742$ when $N_{\text{th}} = 100$ at $p = 0.7$ (Fig. 4b1). This can be explained as follows: When $N_{\text{th}} = 2$, thresholds are distributed as 0.5 or 1 and a node which has opinion 0 can change its opinion from 0 to 1 when 8 neighbors or all of his neighbors have opinion 1 if $N_{\text{th}} = 2$, respectively. On the other hand, thresholds may take values of 0.500, 0.625, 0.750, 0.875, or 1.000 when $N_{\text{th}} = 5$, and a node can change its opinion when 8, 10, 12, 14, or all of his neighbors have opinion 1 if $N_{\text{th}} = 5$, respectively. When the initial probability is higher than 0.7, one may expect that a node has more than 10.5 ($M_{\text{in}}xp = 15 \times 0.7$) neighbors

who have opinion 1 initially, and exceeding thresholds are easier when thresholds are not equal to 1. Therefore, the number of nodes which have opinion 1 is higher at the steady state when threshold heterogeneity is higher. When it comes to the effect of heterogeneity on the opinion fixation time, t_f increases with increasing N_{th} when thresholds are in-degree dependent. When thresholds are out-degree dependent, on the other hand, the effect of N_{th} on f_t is very minimal and the relation between N_{th} and t_f depends on p, e.g. increasing N_{th} causes the people to fix their opinions more lately when $p \gtrsim 0.7$, while the effect is opposite when $0.6 \gtrsim p \gtrsim 0.7$.

Figure 4a2 and b2 shows $\bar{s}$ as a function of p at various M_{in} and M_{out} values. Results show that the change in the mean-degree has no prominent effect on the average opinion at the steady state when thresholds are out-degree dependent; however, increasing mean-degree seems to facilitate reaching a consensus when thresholds are in-degree dependent if $p \lesssim 0.7$. If $p \gtrsim 0.7$ in the same case, $\bar{s}$ values are very close to each other again. Since standard deviations of the results are high, we can conclude that mean-degree does not affect average opinion at steady state either when thresholds are in-degree dependent or out-degree dependent. This is not surprising when we redefine the threshold model. The threshold model basically takes the ratio of the node's threshold to the average opinion of his neighbors and the node changes his opinion if the ratio is higher than 1. Since the ratio does not change with changing mean-degree when the initial, $\bar{s}$ is not affected from M_{in} and M_{out}. In fact, we would expect t_f to increase because the number of links between nodes increased and this increase will cause more changes in ideas, and the results show that the change in mean-degree has no effect on t_f when thresholds are in-degree or out-degree dependent.

Increasing node size in the network decreases $\bar{s}$ significantly when thresholds are in-degree dependent, whereas it has very little effect when thresholds are out-degree dependent. Since we analyze the effect of seed size on $\bar{s}$ when $0.6 \leq p \leq 0.8$, it means that there is more diversity in the opinions when seed size is higher, e.g. $\bar{s} = 0.9480$ when $N = 1000$, while $\bar{s} = 0.7155$ when $N = 5000$ at $p = 0.7$. It means that almost %70 of the population has opinion 1 and %30 has opinion 0 when $N = 5000$. Low standard deviation in the Monte Carlo simulations also demonstrates the consistency of simulation results in every trial. As we mentioned before, we did not simulate a case in which $N = 5000$ for the analysis of opinion fixation time since we expect the result is not novel and obvious.

4 Conclusion

People make decisions in their daily lives on shopping, career, politics, and so on. Although it looks like we make these decisions by ourselves, other people have a great influence on us since we are not isolated from each other. While face-to-face interactions used to be a main communication tool in the past, today's communica-

tion happens mostly on social media. Therefore, social network analysis has become very important to understand the dynamics of opinion formation, change, and propagation. One of the most common methods used to understand these dynamics is the threshold model, in which individuals adopt a new opinion only if a critical fraction of their neighbors have already adopted the new opinion. First studies of social contagion have used homogeneous binary threshold model due to its simplicity; however, people show different attitudes to adopt a new opinion, which renders the use of heterogeneous thresholds a must. Even though more complex thresholds are used in social contagion analysis nowadays, none of them validates their model with real data analysis. The main novelty of this study is that the degree-dependency of thresholds is inferred by using real-world Twitter data. Social data analysis shows that the threshold of a node does not only depend on his out-degree but also depend on his in-degree. Although the examples of out-degree-dependent threshold models can be found in some studies, we also examined the results of opinion change simulations either for the in-degree-dependent threshold model and out-degree-dependent threshold model. Another contribution of this study is to investigate the effect of heterogeneity in thresholds on reaching a consensus for the first time. Our simulations demonstrated that the system is more likely to reach a consensus when thresholds are in-degree dependent, rather than being out-degree dependent; however, people change their opinion more and fix their opinion more later in this case. More heterogeneity in the thresholds is more likely to result in consensus but reaching a consensus takes more time, which is more significant when thresholds are correlated with the in-degree of nodes. Additionally, increasing seed size in the network makes the formation of consensus more difficult regardless of the dependence of threshold to the in-degree or out-degree. Another important point is that, as mean-degree increases, diversity in opinions of individuals decreases when thresholds are in-degree dependent while it has no effect when thresholds are out-degree dependent. For future works, one may cluster Twitter users by using transfer entropy analysis to understand the dependence of users' threshold to their in-degree and out-degree. Thus, degree-dependence in threshold might be modeled more reasonably.

Acknowledgements This work was partially supported by grant FA8650-18-C-7823 from the Defense Advanced Research Projects Agency (DARPA).

References

1. Bovet, A., Makse, H.A.: Influence of fake news in twitter during the 2016 us presidential election. Nat. Commun. **10**(1), 7 (2019)
2. Culotta, A., Cutler, J.: Mining brand perceptions from twitter social networks. Mark. Sci. **35**(3), 343–362 (2016)
3. Fink, C., Schmidt, A.C., Barash, V., Kelly, J., Cameron, C., Macy, M.: Investigating the observability of complex contagion in empirical social networks. In: Tenth International AAAI Conference on Web and Social Media (2016)
4. Gleeson, J.P.: Binary-state dynamics on complex networks: pair approximation and beyond. Phys. Rev. X **3**(2), 021004 (2013)

5. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociol. **83**(6), 1420–1443 (1978)
6. Iyengar, R., Van den Bulte, C., Valente, T.W.: Opinion leadership and social contagion in new product diffusion. Mark. Sci. **30**(2), 195–212 (2011)
7. Karampourniotis, P.D., Sreenivasan, S., Szymanski, B.K., Korniss, G.: The impact of heterogeneous thresholds on social contagion with multiple initiators. PloS one **10**(11), e0143020 (2015)
8. Lee, Eun, Holme, Petter: Social contagion with degree-dependent thresholds. Phys. Rev. E **96**(1) (2017)
9. Lewis, K., Kaufman, J.: The conversion of cultural tastes into social network ties. Am. J. Sociol. **123**(6), 1684–1742 (2018)
10. Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Stærfeldt, H.H., et al. A scored human protein–protein interaction network to catalyze genomic interpretation. Nat. Methods **14**(1), 61 (2017)
11. Liu, Q.H., Lü, F.M., Zhang, Q., Tang, M. and Zhou, T.: Impacts of opinion leaders on social contagions. Chaos: Interdiscip. J. Nonlinear Sci. **28**(5), 053103 (2018)
12. Mishra, S., Singh, S.P., Johansen, J., Cheng, Y., Farooq, S.: Evaluating indicators for international manufacturing network under circular economy. Manag. Decis. **57**(4), 811–839 (2019)
13. Newman, M.: Networks: An Introduction. Oxford University Press, Oxford (2010)
14. Ruan, Z., Iniguez, G., Karsai, M., Kertész, J.: Kinetics of social contagion. Phys. Rev. Lett. **115**(21) (2015)
15. Shi, F., Teplitskiy, M., Duede, E., Evans, J.A.: The wisdom of polarized crowds. Nat. Hum. Behav. 1 (2019)
16. Singh, P., Sreenivasan, S., Szymanski, B.K., Korniss, G.: Threshold-limited spreading in social networks with multiple initiators. Sci. Rep. **3**, 2330 (2013)
17. Sprague, D.A., House, T.: Evidence for complex contagion models of social contagion from observational data. PloS One **12**(7), e0180802 (2017)
18. Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.: Structural diversity in social contagion. Proc. Natl. Acad. Sci. **109**(16), 5962–5966 (2012)
19. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
20. Wang, W., Tang, M., Shu, P., Wang, Z.: Dynamics of social contagions with heterogeneous adoption thresholds: crossover phenomena in phase transition. New J. Phys. **18**(1), (2016)
21. Watts, D.J., Dodds, P.S.: Influentials, networks, and public opinion formation. J. Consum. Res. **34**(4), 441–458 (2007)
22. Zhu, X., Wang, W., Cai, S., Stanley, H.E.: Dynamics of social contagions with local trend imitation. Sci. Rep. **8**(1), 7335 (2018)

Modeling Genocide: An Agent-Based Model of Bystander Motivations and Societal Restraints

Elizabeth von Briesen, Amy Canevello, Samira Shaikh, John Cox, and Mirsad Hadžikadić

Abstract Genocide does not occur within an isolated range of the societal spectrum; rather, it is the outcome of destructive processes that reach from elite governmental offices to the civilian bystander–and beyond. This research applies an agent-based computational model to the problem of identity-based conflict, exploring the dynamics of bystander resistance and its impact on outcomes. Macro-level conditions can lead to difficulties that affect the micro-level, psychological states of people living within a society. This model establishes a connection between these macro- and micro-states, seeking to better understand, explain, and quantify how the motivations and choices of bystanders impact the likelihood of genocide. Early results show that the model reproduces expected behavioral patterns, and also reveals the sensitivity of a genocidal outcome to in-group bystander willingness to intervene in behalf of out-group members.

Keywords Agent-based model · Genocide · Bystanders · Factors of restraint · Ecosystem · Egosystem · Motivational orientation

E. von Briesen (✉) · A. Canevello · S. Shaikh · J. Cox · M. Hadžikadić
UNC Charlotte, 9201 University City Blvd, Charlotte, NC, USA
e-mail: evonbrie@uncc.edu

A. Canevello
e-mail: acanevel@uncc.edu

S. Shaikh
e-mail: samirashaikh@uncc.edu

J. Cox
e-mail: jcox73@uncc.edu

M. Hadžikadić
e-mail: mirsad@uncc.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_8

1 Introduction

Genocide is the result of human interaction, and can be examined using both micro- and macro-level theories. An agent-based model (ABM) is a suitable choice for simulating this problem, as it can capture individual interactions within these complex social systems, providing an artificial world in which to explore emergent outcomes. The social science theories selected to inform the model are thus "candidate explanations" for outcomes [15].

This work combines social psychology and political science theories in order to determine the most appropriate set of agent attributes, system conditions, and interaction rules. The Ecosystem and Egosystem Theory of Motivational Orientations provides a micro-level framework for understanding how people are motivated to support one another, dividing the individual's motivations into two independent systems that are "scaffolded onto" the "evolved motivations of our species" for self- and species-preservation [14]. Narrowing our focus to identity-based conflict, Social Identity Theory provides a well-defined framework from which to describe and model people as they interact both with those sharing their identity, and with those whom they consider "the other." Ervin Staub's framework for modeling the dynamics of genocide is particularly useful for understanding the role of bystanders in this process, and it informs individual agent decision-making in the model [25]. Finally, this research relies heavily on the theories of political scientist Scott Straus in order to simulate a society's macro-level "factors of restraint" against out-group persecution [28].

Prior ABMs of conflict tend to fall into two categories: those that are highly generalized and theoretical, with limited ability have their output validated using real-world data [6, 9, 16, 20, 21, 23], and those that are scenario specific, producing output that is strongly validated on historic event data [7, 8, 34]. This work seeks to combine elements of the above theories to produce an ABM of identity-based conflict that is theoretical, in that, it can model basic social patterns understood to exist across social systems, yet is sufficiently customizable in order to obtain output that sufficiently corresponds to data from different historic scenarios. Generality can be obtained by using an ABM to model personality traits and behaviors at the individual level, which, in this case are human motivations for self- and species-preservation. Specificity is possible through the use of a system-level function that is an abstract representation of factors of restraint as understood by Straus [28]. Early model results are encouraging, in that, they show how factors of restraint can significantly affect levels of violence against out-group members. While the model is not yet validated on historic events, it does yield expected patterns of behavior in a general sense, and is well situated for incorporation of historic data and attempts to achieve output validation for different scenarios.

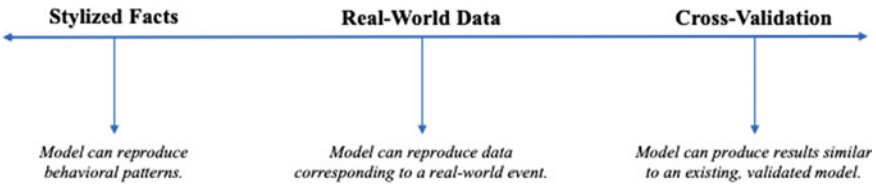


Fig. 1 Diagram of ABM validation as a spectrum defined by the data the models produce, ranging from reproduction of behavioral patterns, real-world event data, or data produced by a different model

1.1 Review of Prior Work

1.1.1 Understanding the Spectrum of Output Validation

This review of prior work focuses on ABMs specific to conflict. The research and model presented in this paper seek to address two specific gaps within the prior work: the inability of generalized models to reproduce a variety of historic scenarios and the lack of flexibility in models that can simulate specific historic scenarios. Validation can be understood as a spectrum, and the level of validation required is dependent on the purpose and goals of the model. Figure 1, derived from the work of Rand and Rust, visualizes this spectrum [22].

Models validated on stylized facts are useful, in that, they can reproduce behavioral patterns that are important to researchers. Section 1.1.2 reviews a number of models in this category, and outlines the significance of being able to reproduce and study general behavioral patterns in ABMs of conflict. Section 1.1.3 reviews models that use real-world data to model specific scenarios. These also reproduce known behavioral patterns; however, do so within specific contexts. The latter category of models are valuable to those interested in using the validated model to explore outcomes resulting changes to the environment through theoretical policy initiatives and interventions.

1.1.2 Validation on Stylized Facts

The first major gap in prior work is that, theoretical models of behavior typically lack event validation through the use of real-world data; rather, they are validated according to subject matter experts' knowledge of the domain. These models reside in the left-most part of the validation spectrum shown in Fig. 1. This section reviews models in this category.

Epstein questioned if simplicity in an ABM could effectively "...generate recognizable macroscopic revolutionary dynamics of fundamental interest." His seminal model of civil violence includes attributes primarily limited to Hardship, Legitimacy, and Grievance, and this simple model reveals complex, emergent phenomena such as "deceptive behavior" and development of "safe havens" [16]. Bhavnani developed

an ABM that explored the 1994 Rwandan genocide through the lens of ethnic norms. His model unveils critical connections between punishment, heterogeneity of a society, and influential people, as each interacts to contribute to the spread of violence [6]. Bhavnani and Miodownik's ABM exploring the dynamics of ethnic salience is also important. The model results clearly establish that ethnic salience is a critical variable that can moderate the effects of conflict and polarization [9]. In Kustov's ABM of ethnic structure, as it relates to civil conflict, he finds that, making ethnicity "bidimensional" is essential to restraining the effects of grievances in a society. His results indicate that "crosscuttingness" of ethnicity can reduce the salience of any single dimension, leading to a reduction in the grievances that lead to violence [21]. Ibrahim and Hassan extended prior work to develop a "revolutionary crowd model" in order to account for the unique dynamics of modern communication through social media and other digital outlets. Their results indicate that the influence of "acquaintances" is not sufficient to maintain protest momentum; rather, that the momentum of a movement can only be maintained by "... constant stimulation that affirms and repeats signals" [20]. Finally, Shults et al. developed an ABM of the "mutual escalation of anxiety between religious groups." Their major finding is that, escalating anxiety between groups over extended periods occurs in the presence of two factors: minimal difference in the size of these groups and agents encountering environmental "social and contagion hazards" exceeding their anxiety threshold [23].

The ABMs outlined above have high value because they allow the exploration of social dynamics leading to conflict for which there is often no available data. For example, Bhavnani and Miodownik found the lack of individual-level data about ethnic salience made it difficult, or even impossible, to validate the model on an event [9]. For Ibrahim and Hassan, validation was not robust in a quantitative sense; however, qualitative comparisons suggest that the model is a good representation of real-world scenarios [20]. In other cases, the research is in its early stages, and the authors see the work as having the potential to yield more advanced simulations, including event validation through the use of real-world data [6, 23].

1.1.3 Strong Event Validation Using Real-World Data

Next are the models that lie in the central portion of the validation spectrum shown in Fig. 1. Here, three conflict models stand out with respect to event validation: a model of Afghanistan, one of Palestine, and one of Baghdad [7, 8, 34].

Bhavnani and Choi modeled civil violence in Afghanistan [7]. In this model, civilian agents have attributes of identity, ethnic salience, character, and propensity for risk taking. Political agent attributes include military and logistical capability, distance from their "stronghold," and a vision radius. In the model, political actors attempt to control areas of the environment, and are supported by civilian agents who "denounce" their neighbors in order to "collaborate" with the dominant political actor, which is either the Taliban or coalition forces [7]. The distribution of ethnicity across Afghanistan is informed using data from the ACLED dataset [1]. The model accurately predicted areas in which violence actually occurred as reported in the

ACLED data. Given the validation, Bhavnani and Choi were then able to explore the model and perform "counterfactual analysis." They used this model to explore the effects of heterogeneity and "territorial control," finding that civilians living in more heterogeneous areas were at higher risk for violence if they encountered a rival, or if a dominant "political actor" did not have full control of the region.

The next model in this category is of violence and segregation in Palestine. Here, Bhavnani, et al. developed a model that is an accurate geographic and ethnic representation of Jerusalem. In addition to other attributes, agents have a "perception of discrimination" against their group, and the model's neighborhoods have levels of past violence and policing. Groups have a fixed "mobility" that is determined based on empirical data for their ethnicity. In validating their model, the researchers found that the social distance and segregation parameters yielding the "best-fit" were realistic in representing ethnic tensions in Jerusalem. Confident that their model was reliable, the researchers then used it to explore theoretical scenarios based on proposed solutions for reducing tensions and violence. In comparing policy scenarios, they found that a return to the 1967 borders yielded the highest reduction in violence [8].

Finally, Weidmann and Salehyan implemented an ABM of violence and segregation in Baghdad. The model environment uses empirical data to generate a geographically accurate representation of Baghdad's ethnic neighborhoods, including levels of violence at those locales from 2006 to 2007. The researchers were able to parameterize the model to produce "empirically plausible runs," and they explored how the ratio of insurgents to civilians impacted migration and violence. One of the main findings was that "...ethnic settlement patterns influence where violent attacks are likely to occur." Additionally, small groups of minorities were at higher risk for violence, and migration in the quest for safety increased segregation, bringing with it a reduction in violence. Weidmann and Salehyan were careful to specify that, while the model could not "re-run history and draw definitive conclusions about what actually happened in Iraq," it did increase general understanding of ethnic conflict and its dynamics [34].

The above models are complex and scenario specific. They are highly useful to policy makers, as they provide a realistic artificial world in which to explore a variety of policy initiatives and probable outcomes within the specified society. It is likely that transferring any of these models to a different domain or example would be difficult, as they are designed to tightly fit these particular regions within specific time frames.

1.2 Addressing the Research Gap

This research seeks to develop a model that can be validated in the area *between* "Stylized Facts" and "Real-World Data" shown in Fig. 1. Theoretical models often lack data that is suitable for validation [9], and models validated on real-world events are highly scenario specific and tightly constrained [7, 34]. The model presented below

currently resides in the left-most portion of the validation spectrum, capitalizing on the ability of ABMs to model general scenarios. However, the selected combination of parameters increases the probability of future validation using real-world data. The global function representing factors of restraint is a mechanism designed with this in mind, as it can later be informed by data that quantifies how free the people of a given society are to oppose the persecution of out-groups (see Sects. 2.2.1 and 4 for more detail).

2 Research Approach

This model simulates the dynamics of in-group, civilian bystanders living in a society, in which, there is perpetrator aggression against out-group civilians. Early results are promising, in that, they clarify the importance of bystander intervention. As such, this model provides a framework from which to determine an appropriae source of data for validation in the near future.

In their research, von Briesen et al. developed a detailed, system-level diagram of the dynamics of genocide, shown below in Fig. 2. This diagram furnished the context for the development of an ABM, in which, agents had the following attributes: Ideology, Influence, Susceptibility, Threshold-to-Act, and Radius of Sight. There were two identity groups in the model, with each agent willing to act against one from the opposing group if its Ideology exceeded its Threshold-to-Act. The authors found that for uniform Threshold-to-Act across all agents, higher thresholds had an exponentially beneficial effect in reducing violence [10].

The current model modifies the above framework in order to focus on the role of in-group bystanders within identity-based conflict. Section 2.1 provides an overview of the selected social science theories informing the new model. Next, Sect. 2.2 presents a description of the model implementation.

2.1 Social Science Theories

This section details theories from political science and social psychology that form the framework and components of the ABM implemented in this research. Each theory has a specific significance, and together, they range from understandings of the highest levels of a society to the fundamental motivations of any individual person.

2.1.1 Political Science

Figure 2 is a "simplified causal loop diagram" developed by von Briesen et al. [10]. It graphically depicts the dynamic relationships between variables and actors in

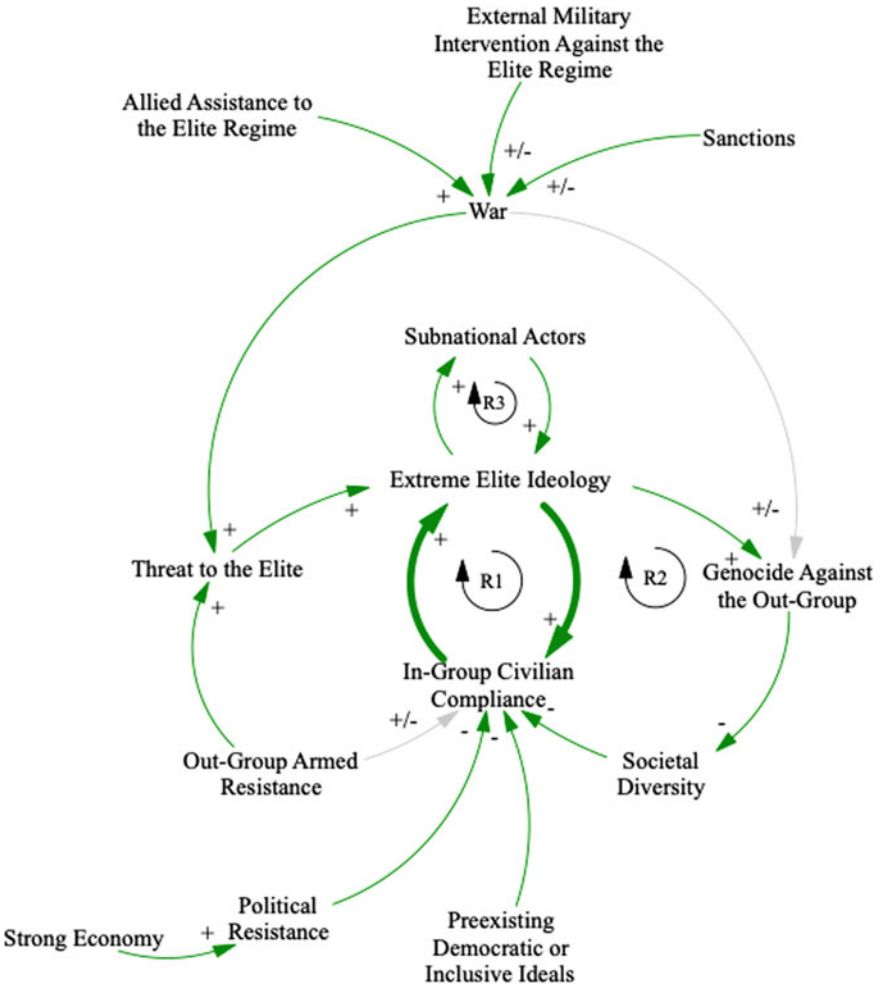


Fig. 2 Causal loop diagram of the system-level dynamics of genocide [10]

societies at risk of genocide. These researchers conducted an extensive review of social science literature in order to create a visualization of the most commonly cited variables and their relationships. This diagram relies most heavily on the work of political scientist Scott Straus [27, 28], with additional contributions from other social scientists and historians [13, 31–33].

As stated above, this causal loop diagram is strongly based on the research findings and models of Scott Straus. In his work, Straus applies the term "factors of restraint" to macro-level conditions that constrain the escalation of extreme ideologies and genocide [27]. Variables that represent restraints on a system moving toward genocide include "Preexisting Democratic or Inclusive Ideals," and "Political Resistance."

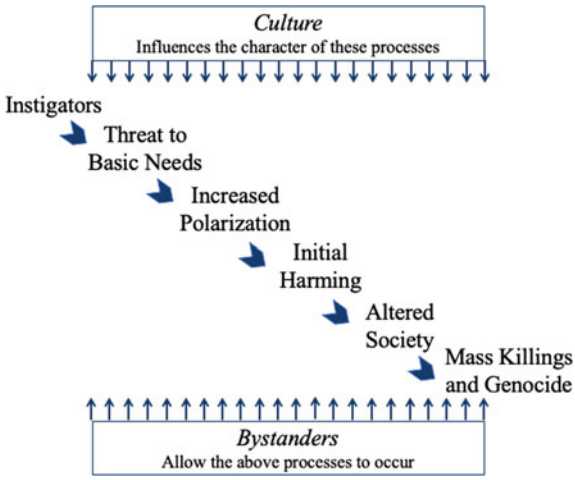


Fig. 3 Diagram of dynamics occurring in a society moving toward genocide, highlighting the influence of culture and bystanders on the process. Adapted from [25]

These are of interest due to their effect on "In-Group Civilian Compliance." In order to explore the effect of these restraints on civilians, we introduce into the model a function, β, that simulates factors of restraint as system-level, aggregate abstraction. A compliant civilian population in the face of extremist ideology is likely to be composed of many bystanders, and the next relevant theory highlights the critical nature of their role.

2.1.2 Social Identity Theory—Bystanders

Figure 3 is adapted from Ervin Staub's diagram of the "Influences and Processes Contributing to Genocide and Mass Killing" [25]. Staub outlines the dynamic processes in a society moving toward mass killing and genocide. Each element in the evolving situation represents a general characteristic, with its specific nature being determined by the culture of the society.

As stated above, we assume that "In-Group Civilian Compliance" implies the existence of many bystanders. Staub states that violence will almost certainly increase in intensity if bystanders do not intervene [25]. Through his years of research, he finds that "[b]ystanders can exert powerful influence. They can define the meaning of events and move others toward empathy or indifference. They can promote values and norms of caring, or by their passivity or participation in the system, they can affirm the perpetrators" [24, p. 87].

In a more recent work, "The Roots of Goodness and Resistance to Evil: Inclusive Caring, Moral Courage, Altruism Born of Suffering, Active Bystandership, and Heroism," Staub provides an in-depth exploration of the nature and role of bystanders. He outlines three bystander categories, detailed below in Table 1. Each category is

Table 1 Descriptions of bystander types as defined by Staub [26]

Bystander Category	Description
Active	Can be found "…speaking out in behalf of their values and in behalf of people who are harmed" (pg. 33).
Passive	Stand in contrast to active; however, do not engage with perpetrators in a way that makes them complicit.
Complicit	Do not intentionally "…support harmdoing but by their actions, make perpetrators believe that at the very least they accept what they do" (p. 14).

included in the model implementation, with Staub's descriptions helping to inform their decision-making process [26, pp. 13–36].

Staub's theory supports the importance of bystanders in identity-based conflict, and also provides a framework for characterizing their motivations and actions. Active bystanders try to help others; passive bystanders remain neutral and do not intervene either way; complicit bystanders support perpetrators at least in an abstract sense. Next, we dig more deeply into the question of the fundamental human motivations toward the "self" and those toward "others."

2.1.3 Motivational Orientation

What motivates people to help others, especially when doing so may be of no personal benefit, or may even be to their own detriment? Reaching beyond the issue of identity in answering this question, we turn to Crocker and Canevello's Ecosystem and Egosystem Theory of Motivational Orientations. Figure 4 is a visualization of this theory, showing "...two types of social motivation scaffolded onto evolved motivations for self-preservation and species-preservation" [14]. The theories of Straus and Staub outlined above suggest that the appearance of specific types of threats can trigger an already polarized society to move closer to violence against out-groups [24, 25, 27]. Ecosystem and Egosystem Theory of Motivational Orientations proposes that the individual's egosystem and ecosystem orientations determine the qualitative nature of his or her micro-level response to external stimuli, and that these systems operate independently of one another [14].

Egosystem. In the egosystem orientation, an individual's goals are oriented toward constructing, maintaining, and defending their desired images of the self; these are termed "self-image goals." Generally, goals in this orientation are "...to be seen by others as having desirable characteristics, and not be seen as having undesirable characteristics." Crocker and Canevello's studies revealed that participants' self-image goals were predictive of feelings of competitiveness, conflict, confusion, and fear. In sum, the egosystem links back to the evolved human motivation for self-

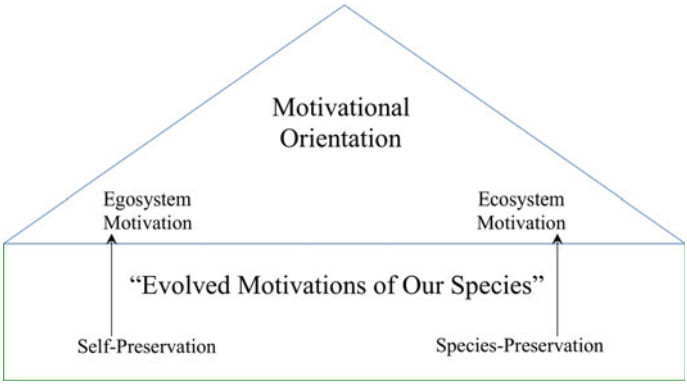


Fig. 4 Visualization of the scaffolding of motivational systems as defined in the ecosystem and egosystem theory of motivational orientations [14]

preservation focusing on personal needs not only for basic survival, but also for maintenance of status (as a proxy for survival) within the group [14].

Ecosystem. In the ecosystem orientation, an individual's goals are formed with the intention to support things like a person, institution, or idea; these are termed "compassionate goals." Having compassionate goals "...promotes caring for the well-being of others and energizes behaviors to protect them and support them to thrive" [14]. Note that this is very similar to Staub's characterization of active bystanders as being altruistic and prosocial [26, pp. 35–36]. People with ecosystem motivations experience higher relationship cohesiveness (less avoidance of partners), and have less anxiety than those with egosystem motivations. This research found that compassionate goals predicted participants' feelings of cooperativeness, peace, love, and clarity. Ecosystem motivation links to the evolved motivation of species-preservation because it strengthens the network of social connections within a group through having an orientation toward others that is non-competitive, caring, and giving without concern for reciprocation [14].

Table 2 brings together the Ecosystem and Egosystem Theory of Motivational Orientations with the bystander types discussed in Sect. 2.1.2. The table shows how this research elects to link motivations, individual goals, and associated affective states as described in [14], with increased probabilities of an individual choosing to be a complicit or active bystander as understood by Staub [26].

The above section outlined the theories underpinning the model framework described below in Sect. 2.2. Factors of restraint, as defined by Straus [27, 28] provide a system- level parameter. Staub's explanation of bystander categories provides a set of low level attributes and mechanisms for decision- making [26]. Finally, Crocker and Canevello's theory provides the lowest level of granularity in clarifying the independent motivational attributes of agents for self- and species-preservation [14].

Table 2 Connections between motivational states, goals, affect, and classification of bystanders. These connections provide justification for model parameters and rules related to BystanderType, as outlined in Sects. 2.2.3 and 2.2.5

Motivational Orientation	Goals	Affect	Bystander classification
Egosystem	Self-image	Competitiveness Conflict Confusion Fear	High Egosystem ⊢ (yields) Higher probability of Complicit Bystander
Ecosystem	Compassionate	Cooperativeness Peace Love Clarity	High Ecosystem ⊢ (yields) Higher probability of Active Bystander

2.2 ABM Implementation

This research uses NetLogo 6.1.0 to implement its model [35]. The following section describes the model environment, agents and their attributes, and interaction rules for different scenarios.

2.2.1 Environment

β-System-Level Factors of Restraint. Figure 5 is a simplification of the larger causal loop diagram from von Briesen et al. shown in Fig. 2 [10]. Here, we narrow our focus to the relationship between Elites and In-Group Civilians. The current model examines the impact of societal factors of restraint on this relationship and outcomes by collapsing them into a global function, β.

The purpose β in this model is to provide a dynamic, system-level attribute that acts as an additive "push" to an in-group agent's motivational state, moving it in the direction of the ecosystem. This simulates societal restraints on the persecution of out-groups. These restraints affect bystanders by increasing their probability of becoming active in protecting out-group members. There are many examples of societies or systems that are repressive and of those that are relatively free. For the purposes of this study, we have chosen two of each in order to provide context for the β function.

Low values of β represent societies that have limited tolerance for civilian opposition to government sponsored acts, such as in an oppressive dictatorship or authoritarian government. Among many modern examples of such societies are The Democratic People's Republic of Korea under Kim Jong-un and his predecessors. North Korea represents an extreme example, as all media is government controlled and its

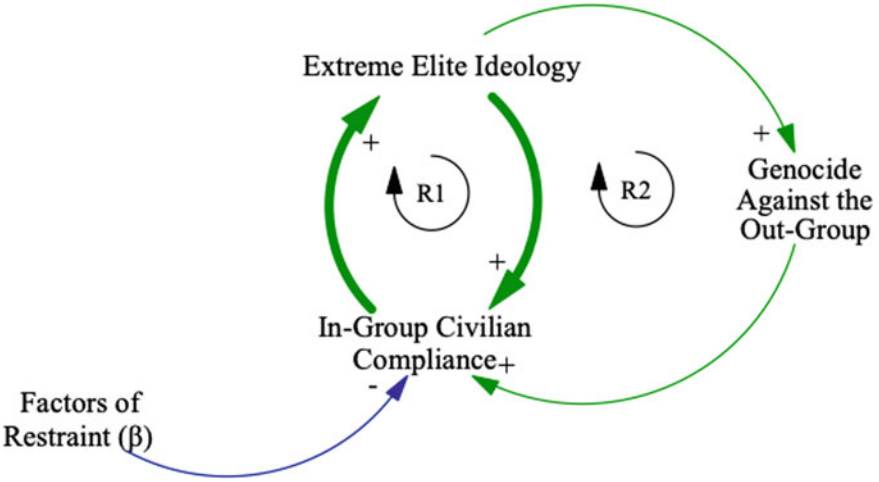


Fig. 5 Causal loop diagram showing elite and in-group civilian dynamics in the presence of generalized factors of restraint as a system-level function β

leadership stands accused of many human rights violations [3, 29]. In some respects, Rwanda under Paul Kagame also fits in this category. Kagame is seen by most experts as an oppressive leader, and his government stands accused of multiple human rights violations including "arbitrary arrest," and hindering freedoms in the electoral process [4, 30].

High values of β represent the opposite—societies that support pluralism and allow citizens to protest governmental actions and policies. The United States stands as one example. Despite evidence of "partisan manipulation of the electoral process," an unfair and discriminatory criminal justice system, and worsening inequality on many fronts, US citizens continue to have relatively high levels of freedom of speech, assembly, and more [19]. Denmark provides a second example. While there are recent issues concerning the treatment of immigrants and refugees, as well as controversy around regulation of public wearing of face coverings, the country has well documented free and fair elections, independent media, full political rights of all citizens, and more [18].

The examples provided above show that there is a great deal of subtlety when classifying a society with respect to how free its citizens are to act as restraints on governmental action. While future work will allow for β to be dynamic in order to model societal changes in time, the current implementation has a constant β value to allow for simplicity.

Location Attribute. Discrete locations in the environment at which violence occurs are coded to appear visually distinct through the use of color. Locations at which violence occurs are colored differently (yellow) from all others (white). In

addition to visualizing violent locations, each locale has a "deathCount" attribute. This attribute records the number of violent acts against out-group agents at that site.

Temporal Scale. A time-step in the model represents one 24-hour day.

2.2.2 Perpetrator Agent Attributes

Perpetrators are an entirely simplistic agent in this model. They are an abstract representation of subnational actors (see Fig. 2) and others directly assisting an Elite regime in out-group persecution and violence [17, 28]. While perpetrators have no dynamic attributes and are only identified on the interface according to their shape, they are coded to be a distinct breed of agent in order to allow for ease of modification in future models. These agents move randomly through the environment, placing an out-group agent is at risk of death when a perpetrator is in its local neighborhood.

2.2.3 Civilian Agent Attributes

Civilian agents in the model represent people who belong one of two distinct identity groups. As noted in the description of the model environment, this research modifies and extends the work of von Briesen, et al. [10]. With respect to agents in the model, the current work continues to implement attributes of Identity, Susceptibility, and Influence, and adds new attributes of Egosystem, Ecosystem, and BystanderType. Table 3[1] details these agent attributes, and is followed by a brief description of each.

Identity. This model will allow for agents to be associated with one of two identity groups. Identity in the model is static and assumed to be salient for all agents.

Susceptibility. This variable is a measure of an agent's susceptibility to change due to micro- or macro-level influences. Micro-level influences are those of other, more influential agents who can cause an agent to update its Egosystem or Ecosystem variables in the influencing agent's direction. The model's macro-level influence is the β function representing factors of restraint.

Egosystem & Ecosystem. The motivation behind the Egosystem and Ecosystem attributes is explained in Sect. 2.1.3. These attributes are dynamic and independent.

Bystander Type. According to their Egosystem and Ecosystem attributes, and accounting for any β function, agents determine if they will be active, passive, or complicit bystanders according to the rules described below in 2.2.5.

[1] A co-author and expert in social psychology provided guidance on the distribution parameters for personality traits: [0.0, 5.0], $\sigma \approx 0.8$. All distributions in the model use a normalized version of these settings: [0.0, 1.0], $\sigma = 0.16$, $mean = 0.5$.

Table 3 Civilian agent attributes as implemented in the model used in experiments outlined in Sect. 3

Attribute name	Type	Range	Initialization
Identity	static, binary	ID $\in$ {A, B}	user determines number of agents per identity group
Susceptibility	static, float	S $\in$ [0.0, 1.0]	normal distribution ($\mu = 0.5$, $\sigma = 0.16$)
Egosystem	dynamic, float	EG $\in$ [0.0, 1.0]	normal distribution ($\mu = 0.5$, $\sigma = 0.16$)
Ecosystem	dynamic, float	EC $\in$ [0.0, 1.0]	normal distribution ($\mu = 0.5$, $\sigma = 0.16$)
Influence	static, float	INF $\in$ [0.0, 1.0]	normal distribution ($\mu = 0.5$, $\sigma = 0.16$)
BystanderType	dynamic, ternary	bType $\in \{-1, 0, 1\}$	1 = active bystander 0 = passive bystander -1 = complicit bystander

2.2.4 Additional Model Settings

The following are additional global settings:

- **Radius of Sight**: uniform, global value determining how far any agent can see from its location.
- **Probability of Mutation**: chance that any agent will reset its attributes randomly (see Table 3 for attribute range and distribution).
- **Probability of Death**: chance that any agent will die. Agents who die automatically produce one offspring that only inherits the identity of its parent. All other offspring attributes are randomly set (see Table 3 for attribute range and distribution).
- **Susceptibility Fraction**: used to slow the rate of change by using a fractional amount of an agent's Susceptibility during updates.

2.2.5 Scenarios and Interaction Rules

This section details the mechanics and logic of the model. In-group agents in the model adapt at each time-step to local *and* global influences. The prosocial and altruistic behaviors of active bystanders, noted by Staub [26, pp. 13–36], can be contagious, leading observers of active bystandership to be more likely to become active bystanders themselves. Perpetrators can commit violence against out-group agents, leading in-group agents to become fearful, and thus have a higher likelihood of becoming complicit bystanders.

Local Adaptation and Accounting for β. As an in-group agent moves randomly through the environment, it updates its Egosystem and Ecosystem attributes according to the following rules:

Algorithm 1 Adaptation According to Agent Influence

let i = *adapting agent*
let j = *random in-group neighbor agent*
if $Influence_j > Influence_i$ **then**
$Egosystem_{i(new)} \leftarrow [Egosystem_{i(old)} + (Egosystem_j - Egosystem_{i(old)})] * \frac{Susceptibility_i}{SusceptibilityFraction}$
$Ecosystem_{i(new)} \leftarrow [Ecosystem_{i(old)} + (Ecosystem_j - Ecosystem_{i(old)})] * \frac{Susceptibility_i}{SusceptibilityFraction}$
end if

Once the agent has updated its Egosystem and Ecosystem attributes, it determines its Bystander Type according to these rules:

Algorithm 2 Effect of β on BystanderType

let $OrientationDifference = Egosystem - Ecosystem$
if $\beta = TRUE$ {There is a β function in the system} **then**
$OrientationDifference \leftarrow OrientationDifference - \beta$
end if
if $OrientationDifference = 0$ **then**
Agent is a **passive** bystander
end if
if $OrientationDifference > 0$ **then**
Agent probability to become a **complicit** bystander is determined according to the magnitude of the OrientationDifference. Higher values yield a higher probability.
end if
if $OrientationDifference < 0$ **then**
Agent probability to become an **active** bystander is determined according to the magnitude of the OrientationDifference. Higher values yield a higher probability.
end if

Note that, the above decision-making process for BystanderType accounts for the global influence of β, meaning there are factors of restraint inherent in the society. The calculation will cause in-group agents to have a higher probability of becoming active bystanders. For example, β = 0.01 will increase this probability by 1% for all in-group agents.

Active Bystander Contagion. Crocker and Canevello find that people with compassionate goals, or ecosystem orientation, "foster trust" in their relationships [14]. The higher a person's compassionate goals, the more they engender trust. Given this finding, the model allows the user the option to include a contagion effect from local, active bystanders. The rules for this contagion effect are as follows:

Algorithm 3 Active Bystander Contagion

let $averageOD$ = *average OrientationDifference of active bystander neighbors*
if $OrientationDifference < 0$ {an agent has a probability to become an active bystander} **then**
$OrientationDifference_{new} \leftarrow OrientationDifference_{old} - averageOD$
end if

This updated value of OrientationDifference gives the agent in question a higher probability of becoming an active bystander when it determines its BystanderState during local adaptation. The magnitude of this probability is higher when its active bystander neighbors have a greater magnitude of difference between their Ecosystem and Ecosystem attributes, on average.

Violence Against Out-Group. If an out-group agent has a perpetrator agent in its local radius, it is at risk of violence. This risk is higher if there are local complicit bystanders and is lower if there are local active bystanders. The probability of a successful attack by a perpetrator is determined as follows:

Algorithm 4 Violence Against Out-Group

let $totalBystCt$ = *total number of all in-group civilians in local radius*
let $activeBystCt$ = *number of in-group active bystanders in local radius*
let $complicitBystCt$ = *number of in-group complicit bystanders in local radius*
if $complicitBystCt > activeBystCt$ **then**
$probabilityofDeath \leftarrow (\frac{complicitBystCt}{totalBystCt} - \frac{activeBystCt}{totalBystCt})/10$
{probability is reduced by a factor of 10 in order to allow for longer model runs}
end if

It is important to note that acts taken against the out-group in this model have an extremely simplified representation. In actual cases of genocide, this persecution can range from basic discriminatory measures, to forced exile, to death. In order to keep the model as simple as possible, persecuted out-group agents simply disappear.

In-Group Fear. In order to model the effect of fear on the psychological state of bystanders, the model can be set to allow an increase in the probability that a bystander will become complicit given their proximity to current or past acts of violence. Recall from Sect. 2.2.1 that a location at which violence occurs stores a total count of all such incidents in its deathCount attribute. The rules for in-group agents are as follows:

Algorithm 5 In-Group Fear

let $totalDeaths$ = *sum of deathCount from all locations within agent's vision*
if $totalDeaths > 0$ **then**
$OrientationDifference_{new} \leftarrow OrientationDifference_{old} + \frac{totalDeaths}{100}$
end if

For example, two violent acts in an agent's locale will cause it to have a 2% higher probability of becoming a complicit bystander when it determines its bystander type during local adaptation.

2.2.6 Model Verification

Egosystem and Ecosystem are initialized as random normal distributions (see Table 3), and the first step in model verification involved tuning the model and confirming that these distributions did not converge over long periods of time, as this would not correctly simulate a typical social system. The setting of ProbabilityofMutation, outlined in Sect. 2.2.4, allows the user to introduce a small amount of random change into the model. In their model of cultural conformity and consistency, Bednar et al. were motivated to accurately capture cultural heterogeneity over time. They found that introducing "small amounts of noise or error" positively impacted the level of heterogeneity in the system, and that these "behavioral trembles" could be a candidate explanation for how a society remains diverse even when its members have a tendency to conform to the group [5]. Model verification was completed with only local adaptation (Algorithm 1) and no β function. This ensured model stability prior to running additional experiments. The process and settings are as follows:

- There is a global **Susceptibility Fraction of 1/10,000**. This was the optimal fractional reduction of Susceptibility during adaptation. Lower values caused the distributions to quickly converge, and higher values prevented the model from producing significant results over anticipated time frames. In the case of this work, this value was adjusted to ensure stability over a minimum of 20,000 d, which represents approximately 55 years. While genocidal violence is unlikely to occur over this long period of time, the model is designed to simulate the evolution of a society towards, or away from, violence against out-groups as visualized in Fig. 3, and requires the ability to span long time frames.
- Every agent has a **Probability of Death of 1 in 25,000** in each time-step. With the death comes reproduction, and offspring inherit only the parent's identity. All other attributes are set randomly.[2]
- Every agent has a **Probability of Mutation of 1 in 10,000** in each time-step. We determined this as the optimal value by running a series of experiments, closing on the value that gave the smallest change in range, mean, and standard deviation of distributions after 20,000 d.

Additional global settings for the experiments presented below are:

- Total number of agents: 500
- In-Group percentage of total: 70

[2] Note that this is not currently an accurate representations of death rates. According to CIA Word Factbook [12] 2018 estimates, the global death rate was 7.7 deaths/1,000 population. The global birth rate was 18.2 births/1,000 population. Future models will carefully incorporate birth and death rates for the region in question, or global rates in generalized scenarios.

- Radius of Sight: 10 patches in NetLogo
- World size: 75 × 75 patches in NetLogo.

This completes the outline of the ABM implementation. Below are the results of two experiments. We explored the effect of active bystander contagion and fear on violence, and then performed sensitivity analysis on system-level factors of restraint (β).

3 Results

3.1 Experiment 1: Effect of Contagion and Fear on Violence

The model runs for this experiment measure the effect of the number of perpetrators on the length of time before an out-group is annihilated. In all scenarios, violence against the out-group is possible (Algorithm 4), and $\beta = 0$ (no system-level factors of restraint). Figure 6 shows the average results over 3 runs, incrementing perpetrators in the environment from 0 to 100, for all combinations of Active Bystander Contagion (Algorithm 3, AB Contagion in chart legend) and InGroup Fear (Algorithm 5). In the legend, 'T' indicates that the simulation included the logic of the algorithm in question, and 'F' indicates that it did not. Future experiments will greatly increase the model runs in order to provide more reliable average values.

The results show that the model performs as expected. It takes longest to reach out-group annihilation when the in-group is not afraid because of violence (*In-Group Fear = F*), and compassionate goals engender trust (*Active Bystander Contagion = T*). The worst-case scenario is the opposite, *In-Group Fear = T* and *Active Bystander Contagion = F*.

Of interest here is that regardless of the model settings, once there were a sufficient number of perpetrators in the environment (a global setting), the out-group could not survive for very long. In fact, with only 20 or more perpetrators, all runs ended before 5000 d. This aligns with Scott Straus' finding that regardless of an individual's personal beliefs and morals, their ability to make a difference in restraining violence diminishes as the drive of elites to persecute out-groups increases [27].

These results inspired the next level of exploration. We were particularly interested in the linear region in Fig. 6. Here the rate of death is generally constant over all scenarios, regardless of the number of perpetrators. This led to the question of how factors of restraint (β) might impact out-group survival.

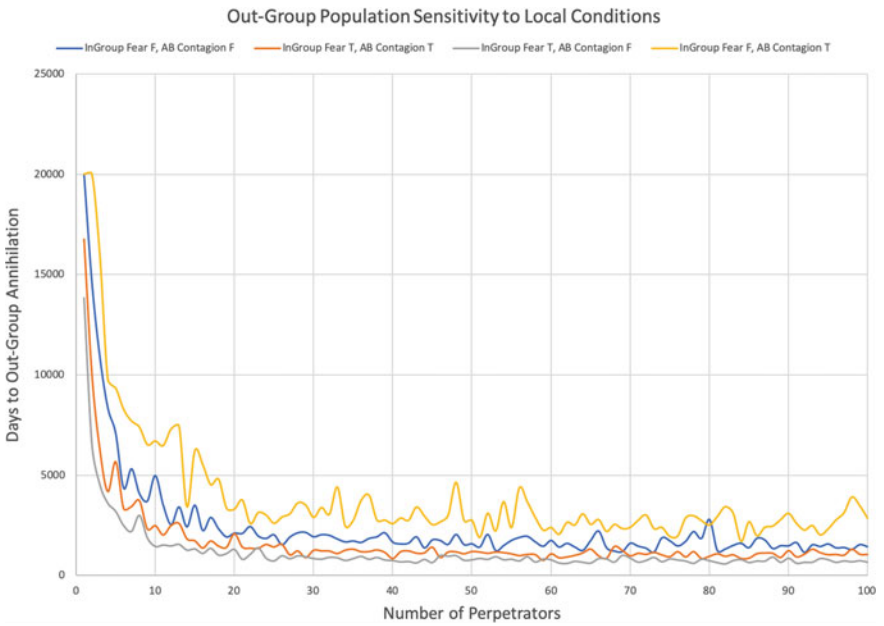


Fig. 6 Out-group population sensitivity to all combinations of use for Algorithms 3 (Active bystander contagion effect) and 5 (In-group fear in the presence of violence)

3.2 Experiment 2: Sensitivity Analysis—System-Level Factors of Restraint (β)

Drawing on insights gained from Experiment 1 (Sect. 3.1), we elected to perform sensitivity analysis of the global function of β, which simulates societal restraints on the persecution of an out-group. In the case of this experiment, β is a constant value through the duration of each model run, ranging from 0.09 to 0.30 in steps of 0.01. The results shown in Fig. 7 are averaged over 3 runs for every value of β, with each run allowed a maximum of 10,000 d.

All model settings remained as presented above, with the following customizations:

- Number of Perpetrators: 50 (constant)
- Active Bystander Contagion: TRUE
- In-Group Fear: TRUE.

The chart above shows that for low values of β ($\beta \in [0.09, 0.14]$), the out-group population drops in a roughly exponential fashion. Toward the middle of the chart ($\beta \in [0.15, 0.23]$), the data shows the rate of population decline becomes more linear. However, the result with the greatest significance is at the top of the chart ($\beta \in [0.24, 0.30]$), where the slopes of these lines begin to approach zero.

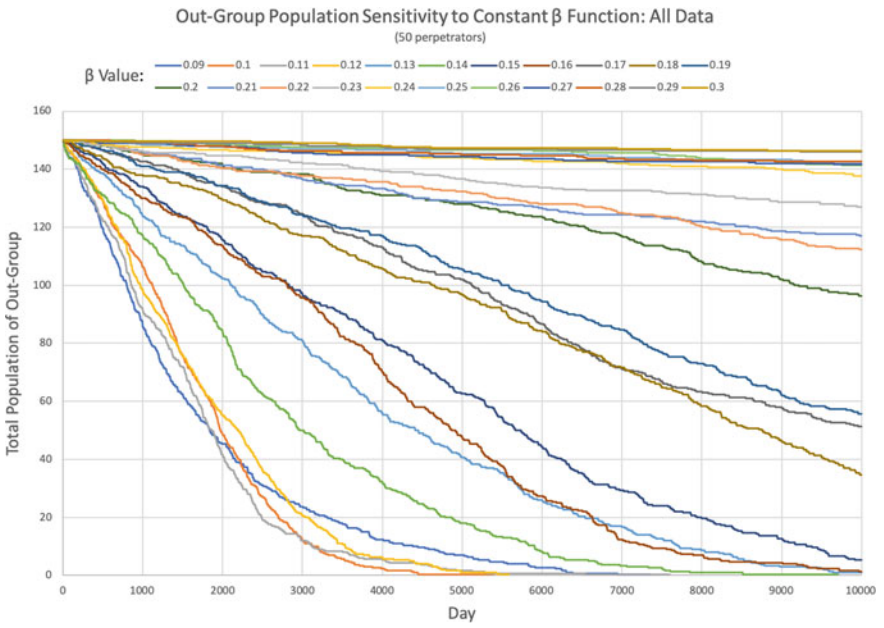


Fig. 7 Out-group population sensitivity to a constant β function providing system-level restraint against out-group violence

Fitting the results shown in Fig. 7 to a linear function for $\beta \in [0.19, 0.30]$ provides a powerful visualization of the dramatic change in death rate within this small range. Figure 8 shows that the change in the slopes of these lines according to increasing β approaches zero at the maximum value of $\beta = 0.30$.

As highlighted in Fig. 8, $\beta = 0.30$ represents a society in which any civilian has a 30% higher probability of becoming an active bystander than in one with $\beta = 0$. In the latter case, there are *no* societal restraints against out-group persecution, such as in a strong authoritarian state. This contrasts with societies in which citizens are more willing to stand in support of persecuted groups due to norms of pluralism, democracy, and so on. These conditions vary in quality, strength, and scope across societies, as discussed in Sect. 2.2.1.

Figure 9 provides a sample visualization of BystanderType counts for four different values of β. The gray line tracks the count of passive bystanders, green tracks active bystanders, black tracks complicit bystanders, and the thin blue line shows the total out-group population over the model run. To provide a full visualization of out-group annihilation, these runs were allowed to continue for up to 20,000 d.

As the results in Fig. 9 show, β has a direct effect on the number of active bystanders from the beginning of the simulation run. This is expected, as is the increase in complicit bystanders with an increase in out-group deaths. Of interest, here are the widening gap between active and complicit bystanders, and the narrowing gap between active and passive bystanders. The number of active bystanders becomes signifi-

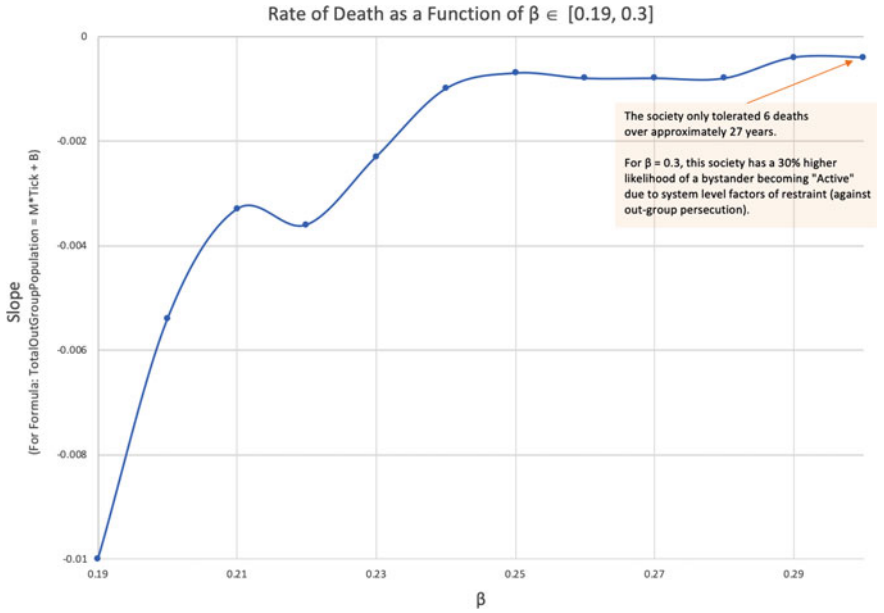


Fig. 8 Effect of β on out-group rate of death according to slopes of lines in Fig. 7 for $\beta \in [0.19, 0.30]$

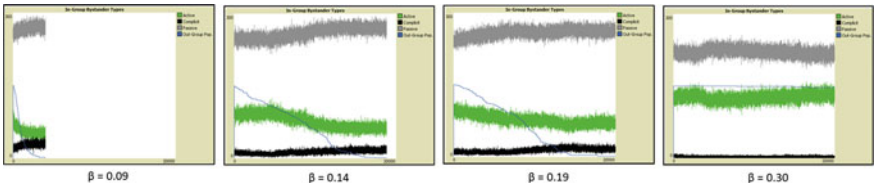


Fig. 9 Visualization of total number of bystanders according to type, over time, for given β values

cantly closer, yet never exceeds, the number of passive bystanders when $\beta = 0.30$. As previously noted, these active bystanders are sufficient to provide protection to the out-group over very long periods of time.

The model described above provides the foundation for continued research examining the evolution of genocide from identity-based conflict. These initial experiments show the model's usefulness for visualizing scenarios that range from peace, to out-group annihilation. We conclude with the anticipated next steps in this research and provide some final thoughts.

4 Conclusions and Future Work

The above research detailed an ABM of identity-based conflict with the potential of genocide. The work relies on theories from social psychology and political science to develop a model framework and set of interaction rules that is generalized, yet has the potential to allow for event validation using real-world data from historic scenarios. The work is currently in an exploratory stage, and does not yet achieve this validation goal. Early results show that in-group bystanders play a critical role, with their ability to restrain violence being highly sensitive to societal levels of restraint against out-group persecution.

There remain several items to address with the current model. One of the first steps will be perform experiments to more finely tune the model. Global settings like the SusceptibilityFraction and ProbabilityofMutation should be more firmly established as "optimal" through an extensive series of model runs and testing. ProbabilityofDeath should be joined by a ProbabilityofReproduction, and both adjusted to accurately reflect death rates found in data [12].

After the above are complete, the following are additional planned experiments using the model in its current form:

- A more granular sensitivity analysis of β, to determine the nature of what appears to be a clustering of results in Fig. 7.
- Sensitivity analysis of the number of perpetrators. This was fixed at 50 during the sensitivity analysis in Experiment 2 (Sect. 3.2).
- Addition of a "buffer" around the zero value in OrientationDifference (Algorithm 2), which will determine a range in which an agent will choose to be a passive bystander. Sensitivity analysis is required here to explore the effects of smaller and larger ranges.
- Sensitivity analysis with respect to population and landscape sizes to determine model stability given variation in population density, as well as to verify that the model will still run correctly at extremes.

Next, it is essential that this early model be tested for its usefulness with respect to validation. Here, the function β is significant, as it allows the use of data to model societal levels of factors of restraint. As explained above, Crocker and Canevello [14] identify specific micro-level emotional states associated with their motivations. If these emotions can be identified in an appropriate source of text, sentiment analysis has the potential to provide an index of emotion that can inform system-level factors of restraint in the model. Alternatively, measures of political conditions for selected validation cases can provide similar measures of the change in factors of restraint in a society. The ICPSR "Nations, Development, and Democracy 1800–2005" database includes data on "Category of democracy," "Centralization of state authority," and "Freedom of demonstration" [2]. The ICPSR database merges variables from a number of existing datasets, including Polity III and IV [11], and development indicators from the World Bank [36]. We will work to use this data to validate the model by

determining the most relevant quantitative measures of factors of restraint, obtaining data for the appropriate location and time-range from the ICPSR database, using this data to inform the model's β function, and then attempting to simulate a historic event such as the 1994 Rwandan genocide.

Finally, the model will be modified to more thoroughly explore this problem, and also capitalize on the advantages of ABM. A more rounded exploration of the problem should include model modifications that introduce: salience of identity, out-group agent adaptation and resistance, and perpetrator adaptation. Two additional modifications will be added in order to capitalize on the advantages of an ABM. First is the introduction of agent memory such that past interactions continue to influence an agent's current state. Second, agents will be linked in social networks in order to explore the effect of small world and star networks on the outcome. These network models will be compared to results using agent geographic intersection to determine the advantages and disadvantages of each.

Studying the problem of genocide at any level, even through a lens as abstract as computer modeling and simulation, causes the researcher to continually be reminded of the darkest and most troubling sides of humanity. The research presented in this paper attempts to take a fresh approach that integrates micro- and macro-level factors in such a way that the resulting model is simple, efficient, and interpretable. It is the hope of all researchers working on this project that this model will contribute to greater understanding of the dynamics of genocide, and that the results will be of benefit to those who seek to prevent the next such tragedy.

References

1. Armed Conflict Location & Event Data Project.: Acled Data—Bringing Clarity to Crisis. https://www.acleddata.com/
2. Wejnert, B.: Nations, development, and democracy, 1800–2005 (ICPSR 20440) (2007). https://doi.org/10.3886/ICPSR20440.v1. https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/20440/datadocumentation
3. BBC.: North Korea country profile—BBC News. https://www.bbc.com/news/world-asia-pacific-15256929
4. BBC.: Rwanda country profile—BBC News. https://www.bbc.com/news/world-africa-14093238
5. Bednar, J., Bramson, A., Jones-Rooy, A., Page, S.: Emergent cultural signatures and persistent diversity: a model of conformity and consistency. Rat. Soc. **22**(4), 407–444 (2010). https://doi.org/10.1177/1043463110374501
6. Bhavnani, R.: Ethnic norms and interethnic violence: accounting for mass participation in the rwandan genocide. J. Peace Res. **43**(6), 651–669 (2006). https://doi.org/10.1177/0022343306069290
7. Bhavnani, R., Choi, H.J.: Modeling civil violence in Afghanistan: ethnic geography, control, and collaboration. Complexity **17**(6), 42–51 (2012)
8. Bhavnani, R., Donnay, K., Miodownik, D., Mor, M., Helbing, D.: Group segregation and urban violence. Am. J. Polit. Sci. **58**(1), 226–245 (2014)
9. Bhavnani, R., Miodownik, D.: Ethnic polarization, ethnic salience, and civil war. J. Confl. Resolut. **53**(1), 30–49 (2009)

10. von Briesen, E.M., Bacaksizlar, N.G., Hadzikadic, M.: Modeling genocide at the system and agent levels. J. Policy Complex Syst. **3**(2), 31–48 (2017). https://doi.org/10.18278/jpcs.3.2.3. http://www.ipsonet.org/publications/open-access/policy-and-complex-systems/journal-of-policy-and-complex-systems-volume-3-number-2-fall-2017
11. Center for Systemic Peace.: The Polity Project. http://www.systemicpeace.org/polityproject.html
12. CIA.: World—The World Factbook—Central Intelligence Agency. https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html
13. Cox, J.M.: To Kill a People: Genocide in the Twentieth Century. Oxford University Press, New York, New York (2017)
14. Crocker, J., Canevello, A.: From egosystem to ecosystem: motivations of the self in a social world. In: Elliot, A. (ed.) Advances in Motivation Science, vol. 5, 1st edn., chap. 2, pp. 41–86. Elsevier Inc., New York (2018). https://doi.org/10.1016/bs.adms.2018.01.003. http://www.sciencedirect.com/science/article/pii/S2215091918300038
15. Epstein, J.M.: Agent-based computational models and generative social science. Complexity **4**(5), 41–60 (1999)
16. Epstein, J.M.: Modeling civil violence: an agent-based computational approach. Proc. Natl. Acad. Sci. U. S. A. **99**(3), 7243–7250 (2002). https://doi.org/10.1073/pnas.092080199
17. Finkel, E.: Macro, meso, and micro research on genocide: gains, shortcomings, and future areas of inquiry. Genocide Stud. Prev. **7**(1) (2012)
18. House, F.: Denmark | Freedom House. https://freedomhouse.org/report/freedom-world/2019/denmark
19. House, F.: Freedom in the World 2019 | United States Country Report. https://freedomhouse.org/report/freedom-world/2019/united-states
20. Ibrahim, Y., Hassan, R.: A revolutionary crowd model: implemented to contrast oscillating to consistent media influence on crowd behavior. Simulation **93**(11), 951–971 (2017)
21. Kustov, A.: How ethnic structure affects civil conflict: a model of endogenous grievance. Confl. Manag. Peace Sci. **34**(6), 660–679 (2017)
22. Rand, W., Rust, R.T.: Agent-based modeling in marketing: guidelines for rigor. Int. J. Res. Mark. **28**(3), 181–193 (2011). https://doi.org/10.1016/j.ijresmar.2011.04.002
23. Shults, F.L., Gore, R., Wildman, W.J., Lynch, C., Lane, J.E., Toft, M.: A generative model of the mutual escalation of anxiety between religious groups. J. Artif. Soc. Soc. Simul. **21**(4), 7 (2018). https://doi.org/10.18564/jasss.3840. http://jasss.soc.surrey.ac.uk/21/4/7.html
24. Staub, E.: The Roots of Evil: The Origins of Genocide and Other Group Violence, 1st edn. Cambridge University Press (1989)
25. Staub, E.: Individual and group identities in genocide and mass killing. In: Social Identity, Intergroup Conflict, and Conflict Reduction, Rutgers Series on Self and Social Identity, vol. 3, chap. 7, pp. 159–184. Oxford University Press, New York (2001). https://ostromworkshop.indiana.edu/library/node/67059
26. Staub, E.: The Roots of Goodness and Resistance to Evil: Inclusive Caring, Moral Courage, Altruism Born of Suffering, Active Bystandership, and Heroism, 1 edn. Oxford University Press, New York, NY (2015). https://global.oup.com/academic/product/the-roots-of-goodness-and-resistance-to-evil-9780195382037
27. Straus, S.: Retreating from the brink: theorizing mass violence and the dynamics of restraint. Perspect. Polit. **10**(02), 343–362 (2012). https://doi.org/10.1017/S1537592712000709
28. Straus, S.: "Destroy them to save us": theories of genocide and the logics of political violence. Terror. Polit. Violence **24**(4), 544–560 (2012). https://doi.org/10.1080/09546553.2012.700611
29. United Nations.: OHCHR | People in North Korea trapped in vicious cycle of deprivation, corruption and repression—UN human rights report. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24651
30. United Nations.: OHCHR | Rwanda—Summary (2019). https://www.ohchr.org/EN/Countries/AfricaRegion/Pages/RWSummary2019.aspx
31. Verdeja, E.: On genocide: five contributing factors. Contemp. Polit. **8**(1), 37–54 (2002). https://doi.org/10.1080/13569770220130112

32. Verdeja, E.: Genocide: clarifying concepts and causes of cruelty. Rev. Polit. **72**(03), 513–526 (2010). https://doi.org/10.1017/S0034670510000343
33. Waller, J.: Perpetrators of genocide : an explanatory model of extraordinary human evil. J. Hate Stud. **52**, 5–22 (2001)
34. Weidmann, N.B., Salehyan, I.: Violence and ethnic segregation: a computational model applied to Baghdad. Int. Stud. Q. **57**(1), 52–64 (2013)
35. Wilensky, U.: NetLogo (1999). http://ccl.northwestern.edu/netlogo/
36. World Bank.: World Development Indicators (WDI) | Data Catalog. https://datacatalog.worldbank.org/dataset/world-development-indicators

Global News Sentiment Analysis

Eric Castellanos, Hang Xie, and Paul Brenner

Abstract The Journalist's Creed, a declaration of the principles, values, and standards of a journalist, states that a journalist should "believe that clear thinking and clear statement, accuracy and fairness are fundamental to good journalism". However, in recent years there has been concern that personal, corporate, and government biases and opinions have had an effect on the fairness and integrity of modern journalism. Many mainstream media outlets are well known to have certain political inclinations and occurrences of "fake news" have become common. Studying mainstream international news sources is important for numerous reasons, such as national security, cultural stability, and the development of social/economic policy. Through the use of modern natural language processing techniques, we performed sentiment analysis on 10,809 articles published by ten globally distributed sources covering eleven globally impactful events, spanning a period of eleven years. We tested three open-source sentiment analysis tools and one commercial tool from Microsoft. We also tested three text processing approaches, i.e., sentence, paragraph, and article-based approaches. We observed that sentiment polarity was generally consistent across news agencies, however, the relative percent differences in sentiment magnitude varied by multiple standard deviations. Further, we found general agreement in sentiment polarity across the four analysis tools strengthening our finding regarding the news source content sentiment trends. The work provides both a reproducible computer-based metric for quantifying differences in global journalism and insights into the effective use of modern sentiment analysis tools.

E. Castellanos · P. Brenner (✉)
Center for Research Computing, University of Notre Dame, Notre Dame, USA
e-mail: paul.r.brenner@nd.edu

H. Xie
Department of Computer Science, Wenzhou-Kean University, Wenzhou, China
e-mail: xieh@kean.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_9

1 Introduction

Sentiment analysis is often used to measure and classify bias in social media posts and customer reviews [1, 2]. In comparison, however, there are a lack of sentiment analysis studies that focus on major/mainstream news outlets, and among those which do, most studies focus solely on U.S. news sources. Our study expands sentiment analysis to global news sources which can affect national security, social and economic policy, and cultural stability. While many believe that mass media does not have the ability to direct government [3, 4], several studies have provided evidence for the claim that news coverage does have the ability to shape public policy [5–7]. Most of these studies focus on how coverage of foreign events in U.S. media affects U.S. public policy. Few published studies focus on coverage of internationally significant events in foreign media and the coverage's subsequent influence on public opinion. In the specific case of U.S. events with international impact, the coverage in foreign media shapes public policy and opinion in those foreign countries. For instance, if a foreign news outlet were to report negatively about the performance of automobiles manufactured in the U.S., this could lead to a decrease in purchases of the automobile in that country and other negative economic effects. As a second example, if a foreign news outlet were to report negatively about U.S. international intervention, it could lead to harboring hostility toward the U.S. and in the worst case, military intervention.

In addition, the loss of ground truth and the emergence of so-called "fake news", also contributes to the shaping of public policy [8]. Negative news reporting of U.S. events and observations of "fake news" can lead to the loss of efficiency in a society and stifle democracy. This problem can be amplified by the development of country-level "echo chambers". For example, Russia plans to introduce legislation that will isolate its internet servers from the rest of the world [9]. If Russia were to then publish untrue articles, Russian citizens would have few internal ways to verify the truth. Further, politicians could manipulate media and information to a point approaching a dictatorship. While it is important to recognize this rise of "fake news", it is tangential to the focus of our study. The rise of "fake news" and political spin has more broadly brought into question the validity of reporting in today's largest and most respected global newsrooms.

We must discuss the fine line between perspective and bias. Perspective can be defined as a viewpoint or frame from which a person sees an event. A journalist covering a particular event often has a deeper understanding of the supporting background than the average citizen. Providing this citizen reader with some perspective is part of the journalist's task. However, it seems that sometimes reporting can go beyond perspective, and venture into bias. Bias implies that truth and results are unfairly prejudiced in favor of one person or group [10]. It is difficult to discern what is perspective and what is bias in journalism, and so our study seeks to enhance the discernment with reproducible algorithmic metrics.

We leverage mainstream sentiment analysis tools to better identify bias and differences in perspective across major global news sources in order to address these

concerns regarding public disinformation. The paper is organized as follows. The first section reviews relevant previous work that has been done for sentiment analysis of media. The second section will cover methodology/experimentation where issues with the collection and querying of data, details of the chosen sentiment analysis libraries, three approaches to process the news articles and design of experiment will be discussed. In the third section, we will present our data analysis results. Fourth, we discuss possible reasons for the output of our chosen sentiment analysis libraries. Finally, we will conclude with summary observations and acknowledgements.

2 Related Work

Multiple works have explored the relationship between computational sentiment analysis studies and bias. For example, Khan and Taimoor identify bias analysis as a subfield of sentiment analysis [11]. Additionally, Zhang, Kawai, Nakajima, and Matsumoto investigate sentiment bias when websites present information [12].

Multiple studies on media bias leverage sentiment analysis. Pang and Lee published a work describing the state of the field of sentiment analysis, which covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems [13]. Most of the works focus on social media rather than traditional news (newspapers, TV stations, etc.), and those that do focus on traditional news, generally restrict the study to U.S.-based sources. Abdul and Diab (a non-US media-focused work) describe their methodology for performing sentiment analysis on Arabic language social media [14]. Balahur describes the difficulty of analyzing sentiment of news articles, and notes the differences between analyzing sentiment of social media and sentiment of news articles [15]. In the following paragraphs, we discuss works similar to ours which focus on news articles.

Bautin et al. focus on machine translation and if sentiment will be affected when translating from English to a foreign language [16]. They explored an approach utilizing state-of-the-art machine translation technology and performs sentiment analysis on the English translation of a foreign language text. The study is relevant to ours because they used news articles from international sources as their data. One of our news sources is based in Argentina and the Spanish articles from this source required automatic translation from Spanish to English prior to the sentiment analysis phase.

Budak et al. describe a machine learning approach to classify news articles and quantify bias [2]. Through a combination of machine learning and crowdsourcing techniques, it investigates the selection and framing of political issues in fifteen major US news outlets. The study utilizes a database of over 800,000 articles. Like ours, the study aims to quantify bias, but does not use sentiment analysis to do so. In addition, the data used in the study is restricted to articles from only US news sources. However, the study serves as a good reference as their methodology/experimentation is similar to ours. Kaya et al. describe the difficulties of performing sentiment analysis on political news articles [17]. They state that sentiment analysis struggles on news

texts due to the lack of large-scale golden standard datasets and the high context-dependency of sentiment-inducing phrases.

3 Methodology and Experimentation

In this section, we break down our methodology into three sections. We first discuss how we gathered our data, initial challenges we faced and how we eventually utilized the LexisNexis data repository. We then discuss how we extracted and cleaned articles, followed by selection of the sentiment analysis algorithms (with basic details on each tool). Then, our reasoning for selection of topics will be discussed. Finally, we introduce three approaches to process news articles and explain why we chose a paragraph-based approach to process articles.

As pointed out by Hamborg et al., finding a data repository of news articles for a computational study is not an easy task [18]. Two datasets available to the public are Integrated Crisis Early Warning System (ICEWS) and Global Database of Events Languages and Tone (GDELT). ICEWS consists of coded interactions between socio-political actors such as cooperative or hostile actions between individuals, groups, sectors, and nation states [19]. GDELT is a large, comprehensive, open database of human society. It is a platform that monitors the world's news media in print, broadcast, and web formats, in over 100 languages, and stretches back to January 1, 1979 through present day [20]. The datasets are vast, date back many years, and contain useful information about events and the articles written about them. However, they do not provide access to full article text, which is essential to our study. We then discovered NewsAPI, a simple and easy-to-use API that returns JSON metadata for headlines and articles [21]. Further, NewsAPI gave us access to title and short, two-sentence summaries of news articles. Ultimately we decided that in order to perform a more complete study, we needed access to full article text. NexisUni, a data repository provided by LexisNexis, features more than 15,000 news, business and legal sources, including U.S. Supreme Court decisions dating back to 1790 [22]. The biggest advantage that NexisUni had to offer was the availability of an API that allows developers to query the NexisUni database on things like article topics and dates, and return the full article content in XML format.

Python was chosen as the language for our analysis. We utilized the lxml library for extraction of the article text and organization of files by article topic and news source. Next, we had to decide which sentiment analysis algorithm and library we were going to use. We researched many different popular natural language processing libraries available in the Python language [23]. While there are many natural language processing libraries and sentiment analysis algorithms available, many of those tools are ill-equipped for dealing with texts like news articles. Most sentiment analysis libraries are intended to be used to analyze small bodies of text, no more than a few sentences. In addition, many of the tools are not very good at dealing with highly contextualized texts, which news articles generally are [24]. As a result, we focused on selecting a sentiment analysis tool that attempted to tackle these chal-

lenges. We initially selected the VADER Sentiment Analysis tool available in the Natural Language Tool Kit 3.4.1 (NLTK) library in Python [25]. VADER's documentation states that the algorithm is equipped to deal with somewhat long bodies of text and can manage "tricky" sentences, and was evaluated by the data from Twitter, movie reviews, technical product reviews, and 500 New York Times opinion editorials [25]. In addition, VADER has been recognized as one of the best tools available for sentiment analysis. VADER combines the use of a lexicon and processing characteristics of sentences to determine the polarity of a sentence. For words in the lexicon, sentiment scores were assigned by human raters in the range of -4 (extremely negative) to 4 (extremely positive), and the overall sentiment score of a sentence is obtained by adding all sentiment scores of words that appear in both VADER's lexicon and sentence. Then, VADER will normalize this final score to a value between -1 and 1 (the normalize process will be discussed further in a later section). VADER also takes into account punctuation, capitalization, and n-gram examination for things like negation and context [26].

To improve our tool evaluation, we also leveraged three other sentiment analysis tools: TextBlob 0.15.3, Afinn 0.1, and Microsoft Azure Text Analytics 0.2.0 (MATA). TextBlob is considered as an easily accessible NLP package and it stands on the shoulder of NLTK and Pattern [27]. It is an easy-to-use library, equipped with impressive functionality, which attracts a wide range of people for their projects' initial prototyping purpose. TextBlob provides two different algorithms to perform sentiment analysis: one is the implementation of the "pattern-en" module [28], which is similar to the lexicon-based approach that VADER adopts, and another one is the Naive Bayes classifier which is trained on movie reviews. In our research, we choose to use the default approach, i.e., lexicon-based approach, as the way to calculate the sentiment. Afinn is also a popular lexical-based sentiment analysis toolkit whose lexicon mainly comes from Twitter [29]. According to Ribeiro, Araujo, Goncalves, Goncalves, and Benevenuto, for the 3-Classes sentiment analyzer, VADER ranked first and Afinn ranked third after comparing 24 sentiment analysis methods on a benchmark of eighteen labeled datasets [26]. Thus, we decided to use this package due to Afinn's excellent performance. Different from VADER, TextBlob, and Afinn which are free to use and can be implemented locally, MATA API is a cloud-based service for natural language processing and has been commercially available for several years [30]. One of the reasons we choose to use this tool was to determine if there is a significant difference between commercial and open-source natural language processing libraries.

For our study, we chose ten different news sources that cover the six populous continents. Choices for selection of news sources were made based on: rough global coverage, popularity of the news sources in their respective regions, availability in English, availability of articles mapping to target events, modestly opposing political spectrum (CNN vs New York Post) and their availability in the NexisUni data repository. The news sources are La Nacion (Argentina), All Africa, Xinhua, The Times of India, Al Jazeera, BBC, RIA Novosti, The Australian, CNN, and The New York Post as shown in Fig. 1. It is important to note that La Nacion is an Argentinian newspaper written in Spanish. Since Matheus et al. [31] suggested that applying machine

Fig. 1 Distribution map for selected news sources

translation to Spanish input text prior to using an English sentiment tool can be a competitive strategy, we utilized Google's translate API in order to prepare articles from La Nacion for sentiment analysis. We created an article count matrix to visualize how many articles were available for each topic, as Fig. 2 shows. As expected, there were significantly more articles returned for the mainstream popular news outlets like CNN, BBC, and Xinhua, while more regional media sources or those with a smaller number of English readers like All Africa and Al Jazeera, respectively, often did not have many articles available for the target events. Therefore, the results for sources such as BBC will have higher statistical relevance than those of All Africa.

The selected events are the execution of Saddam Hussein, the Saffron Protests of Burma/Myanmar, the Election of Barack Obama, the indictment of Omar Al-Bashir (Sudan), Kim Jong Un's succession, the Fukushima Nuclear Disaster, the Benghazi Attacks, the Snowden Incident, Russia's annexation of Crimea, the Legalization of Same-Sex Marriage in the US, and the Election of Donald Trump, as shown in Table 1. Articles were selected over an eleven year time period (2006–2016). One major event was selected for each year, and articles were gathered from the ten sources for each event. In attempt to observe differences in reporting, five selected events were about topics that primarily concerned the U.S., and the other six were about international events.

After querying the NexisUni database, extracting article text from the XML documents, and data cleaning, the articles were ready to be analyzed by the sentiment

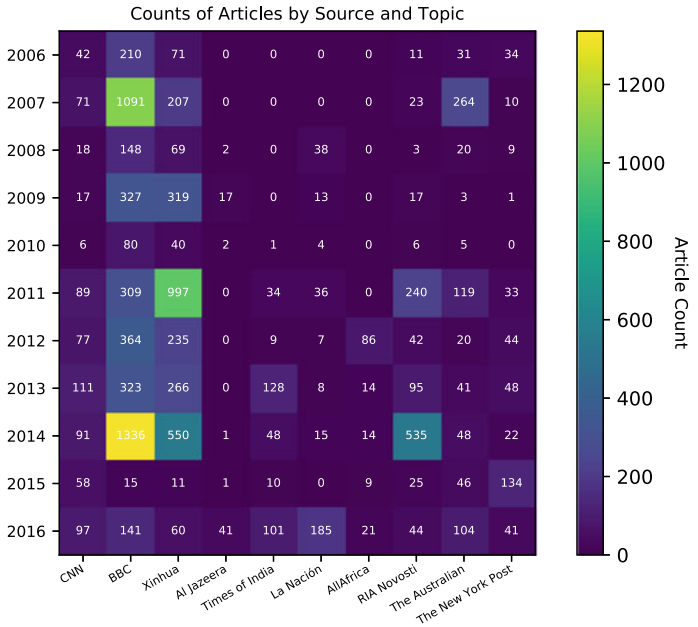


Fig. 2 Article count matrix

Table 1 Selected global events

Year	Event
2006	Saddam Hussein sentenced to death
2007	Saffron protests in Burma/Myanmar
2008	Barack Obama elected as US President
2009	Omar Al-Bashir convicted on humanity crimes
2010	Kim Jong Un announced as Kim Jong Il succesor
2011	Fukushima-Daichi nuclear disaster
2012	US facilities attacked in Benghazi
2013	Edward Snowden leaks sensitive information
2014	Russia annexes Crimea
2015	Same sex marriage legalized in the US
2016	Donald Trump elected as US President

analysis libraries. While the algorithms were reasonably equipped to handle large bodies of text, our initial study utilized paragraph-based approach to process the news articles, where each article was broken down by paragraph, a sentiment score was assigned to each paragraph, and finally each article was assigned the average of all its paragraphs' sentiment scores.

4 Results

4.1 Initial Observation

We first calculated the mean sentiment values of all articles by source and then by event. Not all events were covered by all the news outlets, and thus we see some years missing in the charts. The majority of the sources had some polarity alterations for mean sentiment value for the various events, with some events having positive mean sentiment values and other events having negative or neutral mean sentiment values. Due to the publication space limits, the compound sentiment scores of all articles are available by contacting the authors.

Some events were chosen for their potential to be controversial, while other events had the potential to be either unanimously positive or unanimously negative. For instance, we chose the Fukushima Nuclear Disaster because intuitively, we expected the sentiment of news on this topic to be negative, while we chose Donald Trump's election since we surmised it might be a controversial topic around the world. We analyze the article bias for these two events here in the paper and provide full data (online) for readers to evaluate bias in the other events and sources.

In Figs. 3 and 4, we show the sentiment distribution histogram for Fukushima Nuclear Disaster and Donald Trump's Election from different sources. The X-axis

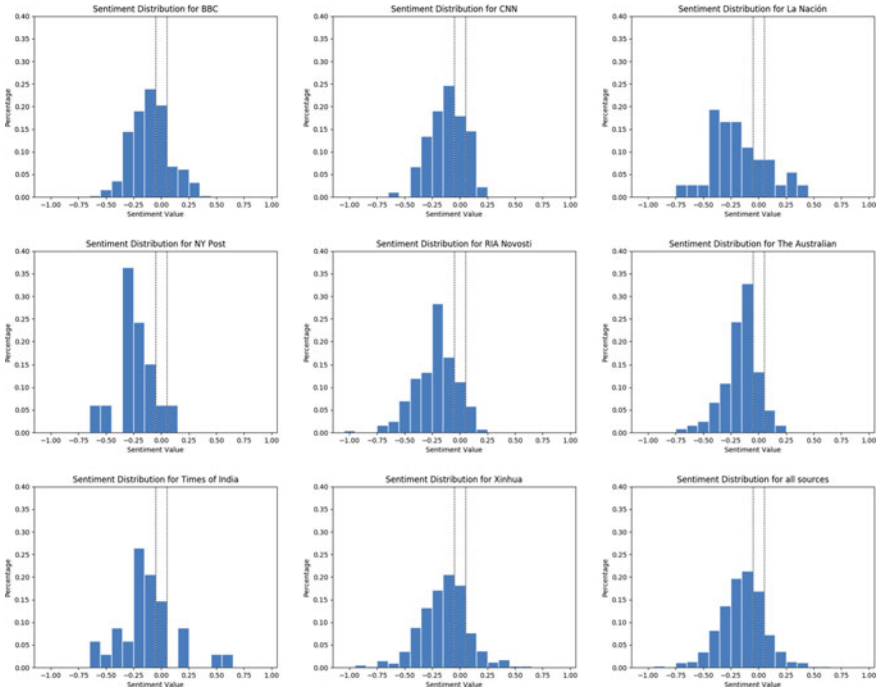


Fig. 3 Sentiment distribution for Fukushima nuclear disaster

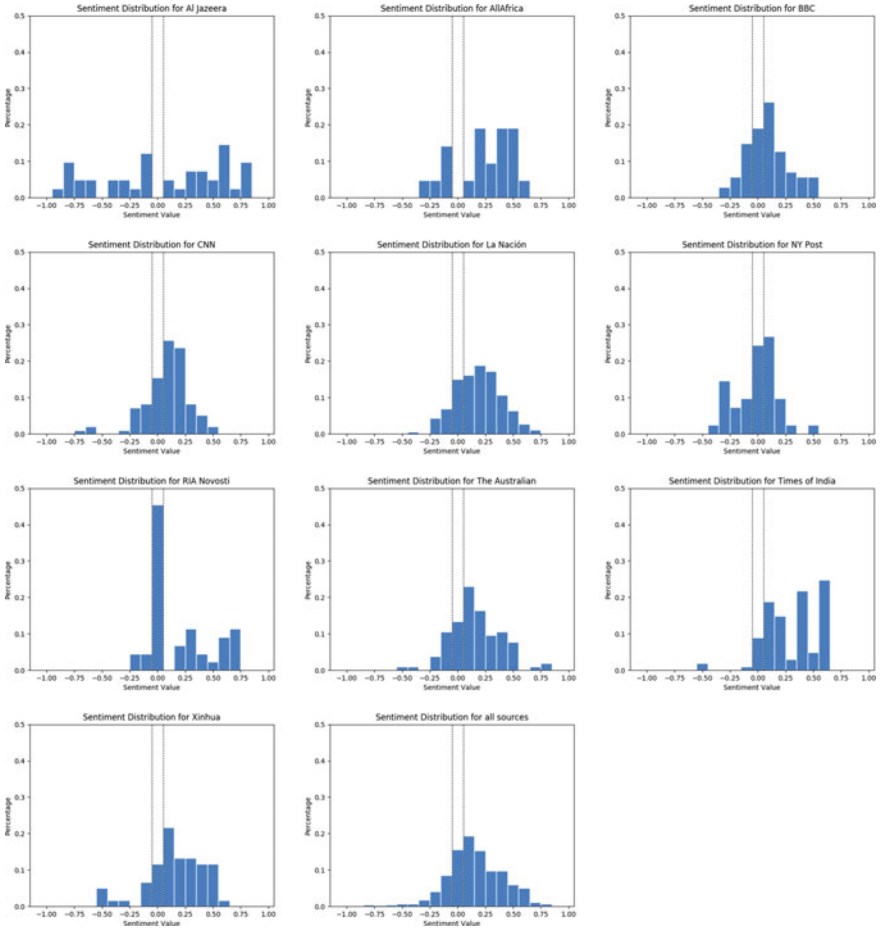


Fig. 4 Sentiment distribution for Trump's election

represents the sentiment compound value of the news while the Y-axis denotes the percentage of news with a corresponding sentiment compound value. We calculate this probability by dividing the number of news articles in the range of emotion value by the total number of news under this topic. For sentiment compound value, VADER has defined that if the compound score is less or equal to -0.05, then the text tested shows a negative sentiment and if it is higher or equal to 0.05, then it shows a positive sentiment. Any compound score which lies in between -0.05 and 0.05 will be regarded as neutral sentiment.

Figure 4 demonstrates the sentiment distribution for Fukushima Nuclear Disaster from different sources and in the last subgraph we also present the sentiment distribution from all sources. For news from all sources, the probability of negative sentiment in the news outweighs the probability of positive sentiment. Specifically,

69.47% of news showed a negative feeling, 13.68% revealed a positive feeling, and 16.86% of news conveyed a neutral sentiment.

The sentiment distribution for Donald Trump's Election, as shown in Fig. 4, yields greater positive sentiment occurring in a news articles than the probability for negative sentiment. In detail, 67.11% of news demonstrated a positive sentiment while 17.34% demonstrated a negative sentiment.

We provide box plots to demonstrate the spread of our data points and check if mean values were being affected by outliers. Information on how whiskers were calculated can be found in the documentation of Matplotlib's Boxplot function, default values were used for all arguments. Where IQR is the interquartile range $(Q3 - Q1)$, the upper whisker will extend to last data point less than $Q3 + 1.5 * IQR$. Similarly, the lower whisker will extend to the first data point greater than $Q1 - 1.5 * IQR$. Beyond the whiskers, data are considered outliers and are plotted as individual points.

The boxplots of sentiment by source demonstrate news publishers possess different viewpoints for each event. In addition, from the boxplot in Fig. 5, we observe that for publishers like BBC, Xinhua, and New York Post, many outliers are present, indicating that these publishers' viewpoints tend to vary widely. For example, for 2014 [Russia Annexes Crimea], news articles released by BBC are either quite positive or quite negative. Note, some news publishers, like Al Jazeera and The Times of India only report a few articles for some events, which provides less statical mass.

From Fig. 6 we can see for most of the events, publishers tend to have similar sentiment polarity toward them. However, for events like 2009 [President of Sudan, Omar Al-Bashir, Convicted for Crimes Against Humanity], sentiment diverges for different news publishers, where NY Post holds a positive view toward it the rest of the publishers generally hold a negative view. We also find that for events like 2014 [Russia Annexes Crimea], there are many outliers in the graph, showing that the overall sentiment for this event is polarized and was quite controversial during that period.

4.2 Comparison of Four Sentiment Analysis Tools

We tested the other three tools for the same dataset, trying to compare VADER with those tools. Since the sentiment scores calculated by VADER and TextBlob are between -1 and 1 while those calculated by Afinn and MATA are not, we normalized the sentiment scores from the latter libraries to -1 to 1 for comparison purposes.

For MATA, the return sentiment score is ranged between 0 and 1, making normalization straightforward. However, for Afinn, since it doesn't have the upper and lower bound of its return sentiment score, we normalized it in the following way: First, we try to calculate the average and standard deviation of the sentiment scores we have for Afinn. Then, we remove the data which is 15 standard deviations away from the average Afinn score. Finally, we divide all Afinn scores by the maximum absolute Afinn score we obtained from all news sources. In this phase, we removed

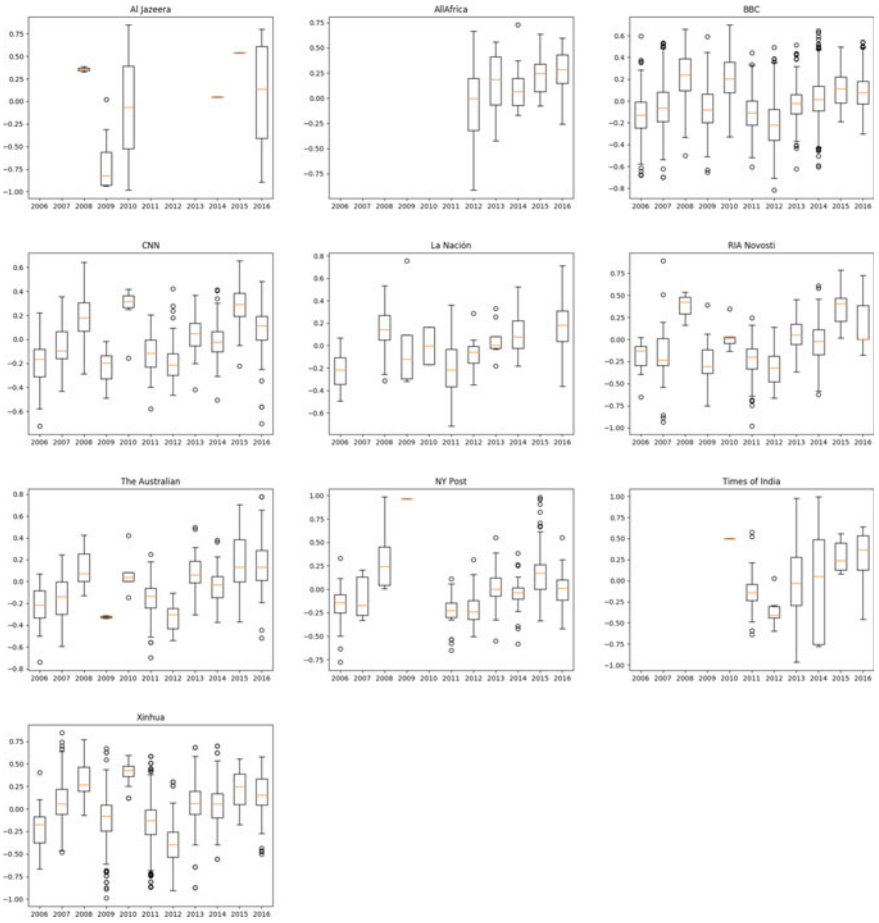


Fig. 5 The boxplots of sentiment based on source by VADER, paragraph-based

two news articles published by Al Jazeera: one for 2010 Kim Jong Un Succeeds, and another one for 2016 Trump elected.

Figures 7 and 8 show the calculated sentiment scores for different sources and events based on our chosen packages. From a practical perspective, there should be a center area that lies in between the positive and negative sentiment, which denotes neutral sentiment. For instance, although sentiment with score 0.01 may mean slightly positive while sentiment with score −0.02 may represent slightly negative, it is difficult to distinguish a difference. Thus, in our paper, we define sentiment score between −0.05 and 0.05 as neutral.

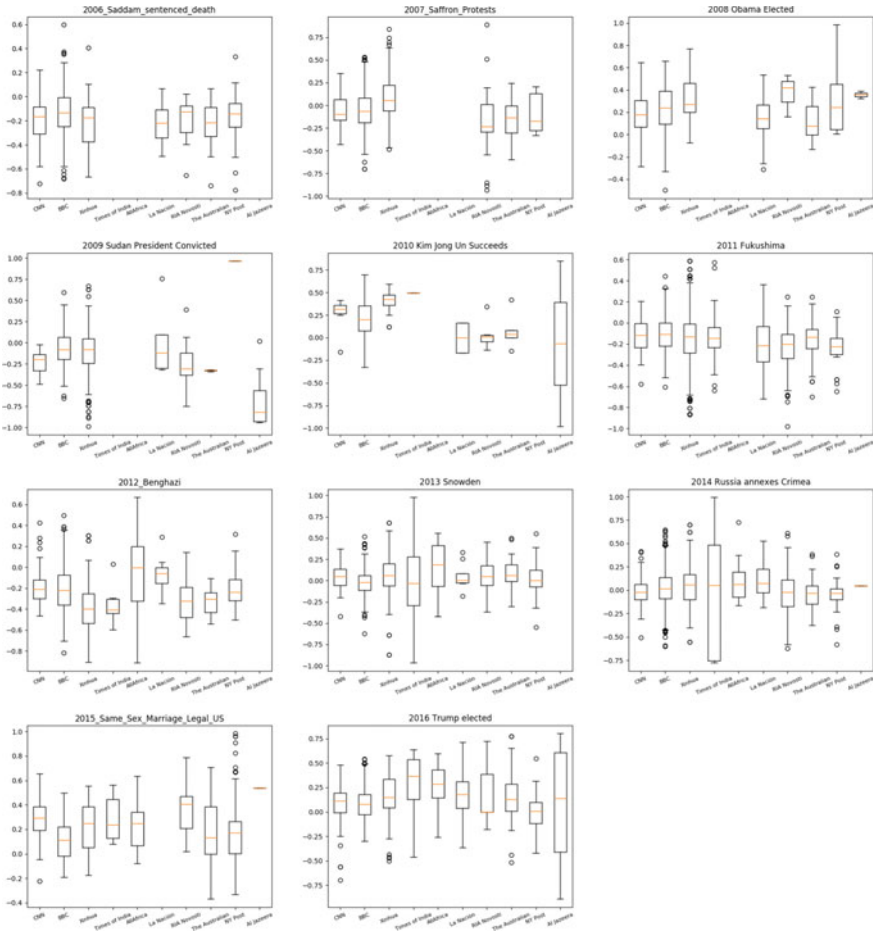


Fig. 6 The boxplots of sentiment based on topic by VADER, paragraph-based

From Fig. 8, we find that the sentiment for news published by a particular source for different events varies, indicating that publishers show different perspectives (sentiment) while reporting news versus consistently treating all news as negative, positive, or neutral. When looking into the detail, we find that for each event published by a specific publisher, the four tools tend to have an agreement concerning the sentiment polarity (when outside the neutral zone). For example, for news published by Al Jazeera concerning 2009 President of Sudan, Omar Al-Bashir, Convicted for Crimes Against Humanity, sentiment calculated by VADER, Afinn, and MATA shows a negative sentiment, while the sentiment calculated by TextBlob shows a neutral sentiment. This occurs throughout the events as demonstrated by both Figs. 8 and 9. We perceive this to be reasonable considering the difference of training data for each tool and sometimes it's difficult, even for humans, to discern the difference between

Fig. 7 Collective sentiment compound score from each source, paragraph-based

neutral and slightly positive or negative, since people tend to have different views on single news articles.

In Fig. 9, we group the sentiment compound score based on the event. For most events we choose, different publishers tend to have similar sentiment polarity. Surprisingly, for events like 2009 President of Sudan Indicted for Crimes Against Humanity, New York Post tends to have extremely positive sentiment while the rest of publishers hold negative or neutral-to-negative sentiment.

Overall, VADER is quite sensitive toward the sentiment underlying the news articles, as we can tell from Figs. 8 and 9, in general the compound score calculated by VADER is higher than the rest of the three and it can clearly detect the polarity of the news. However, TextBlob finds minimal sentiment magnitude variance in our study since most of the values it calculated lie in the range between -0.05 and 0.05.

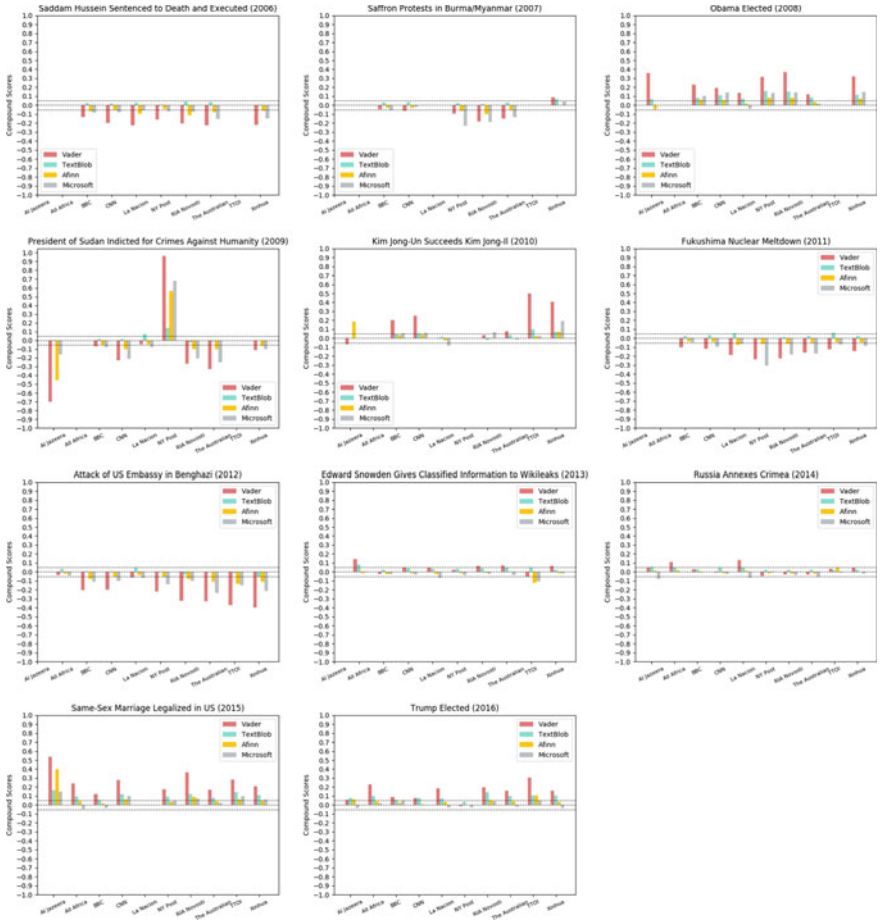


Fig. 8 Collective sentiment compound score for each event, paragraph-based

The performance of Afinn and MATA are fairly good regarding their ability to detect the polarity, but they are not as sensitive as VADER.

4.3 Three Article Processing Approaches

There are three common approaches to parsing text articles: sentence-based, paragraph-based, and article-based approaches. We tested all three approaches to process the news articles, and it turned out that the paragraph-based approach provided us with a relatively objective result which is neither too sensitive nor insensitive.

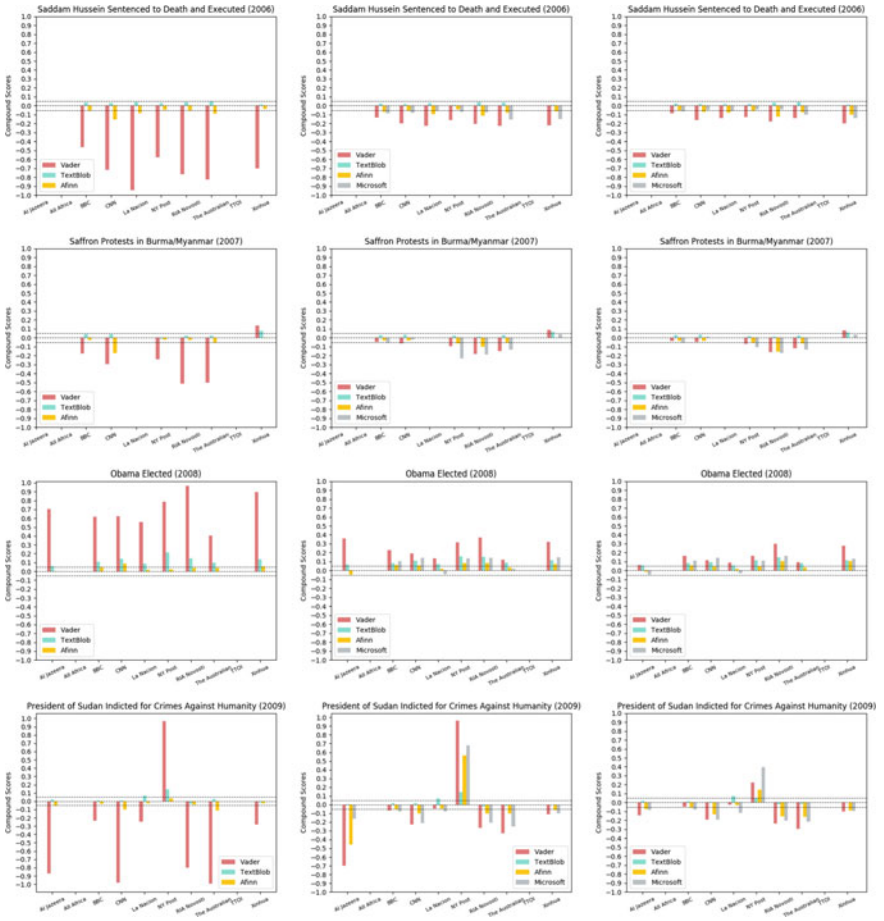


Fig. 9 Three article processing approaches. Column 1 = Article, Column 2 = Paragraph, Column 3 = Sentence

The article-based approach was easy to implement since we can directly put the text into the library to obtain its sentiment score. Unfortunately, since MATA has a limitation that the maximum size of a single document be less than 5120 characters and there were 1765 articles exceeding this limit, we didn't calculate the article-based approach for MATA.

It is also easy to calculate the sentiment score by paragraph-based approach, since the tag <p> in the LexisNexis XML files helps identify the paragraph. For sentence-based approach, however, there are many different cases that can be applied to determine a sentence. For example, if a text is ended by a period, exclamation mark, or question mark. Also, sometimes there are cases when a sentence ends with three continuous periods "…". Initially, we wanted to write a regular expression to extract the sentence out from a paragraph. However, given the linguistic difficulty

of handling all cases which can be used to form a sentence, we decided to use a well-trained tokenization package in the NLTK library to implement the sentence extraction task.

Our results in Fig. 9 show that VADER is more sensitive to the way we process the text. Nevertheless, for three different approaches calculated by VADER, although they have differences in the absolute value for the emotion score, they don't yield much difference in the polarity of the sentiment. Specifically, we find that VADER seems to be extremely sensitive to the emotion in the news articles processed by the article-based approach. For the result calculated by paragraph-based approach by VADER, it is less sensitive than the article-based approach, but is more sensitive than the sentence-based approach.

In a minority of cases, the sentiment polarity shifted as we alter the text processing approaches. For example, for news published by CNN for 2007 Saffron Protests in Burma/Myanmar, Afinn first possesses a negative view about it for article-based approach, but when we test the paragraph-based, the CNN then holds a neutral view toward it, though there is not much change in absolute sentiment score. Overall, the article-based approach is the most sensitive approach, followed by the paragraph-based approach and the sentence-based approach. Also for VADER, no matter which text processing approach we choose, it will rarely alter the sentiment polarity for our news. However, for the rest of tools, since it may alter the sentiment polarity when we use different approaches to process the text, we should think carefully concerning which approach works better before further application.

5 Discussion

The heightened sensitivity of VADER for different text processing approaches could be explained by its normalization approach. We know VADER uses a lexicon-based approach to obtain its sentiment score, where each sentiment-bearing word in a text will be mapped to a sentiment score which is given by human raters in a range of -4 to 4, and then we sum it up as the overall sentiment for this text. However, we notice that the return value the VADER provides us is between -1 and 1, indicating there has a normalization process before we get the sentiment score. Hutto and Gilbert do the normalization in the following ways, as it shows in VADER's source code:

$$normalizedScore = \frac{score}{\sqrt{score^2 + \alpha}}$$

where the score is the sum of the sentiment scores for each sentiment-bearing word, and the α, which is set to 15 as a default value, is used as a normalization parameter. For example, there is a sentence in one of the news articles concerning Trump's Election: "There are jurisdictions that fail to report hate crime statistics". In this sentence, three words get matched to the VADER's lexicon: fail, hate and crime, and their sentiment score are -2.5, -2.7, and -2.5, respectively. Thus, the sentiment

score before we normalize it should be -7.7, and when we replace the score in the above formula by -7.7, and let α be the default value 15, we get our normalized sentiment score -0.8934 for this sentence.

As we can observe from this formula, if the score becomes larger and larger, then the effect of α toward the denominator will decrease, and the whole equation will approach to -1 or 1. We know that the length of the text also matters for the performance of VADER. If the text is longer, then the drag effect of the α will be lesser, and the normalized score will be potentially higher, or more likely to get closer to -1 or 1.

Similar to VADER, the default implementation of sentiment analyzer in TextBlob is based on the pattern.en module [28], which relies on the hand-coded sentiment lexicon to calculate sentiment. However, this lexicon only contains 1,528 adjectives that frequently appear in movie reviews. For example, the sentence "It is fun and unforgettable." will yield a sentiment polarity 0.55 because the polarity of "fun" and "unforgettable" in our lexicon is 0.3 and 0.8, respectively, and after averaging these two we obtain 0.55.

One reason that yields the insensitivity of TextBlob is that the number of words in its lexicon is only 1,528, which is lesser than the VADER and Afinn, which contain 7,518 and 3,382 words, respectively. When we try to calculate the sentiment using TextBlob, especially in paragraph and sentence-based approaches, only a few adjectives in one sentence can match up with the adjectives in the lexicon, and in some paragraphs/sentences, there might exist some adjectives which cannot be mapped into the lexicon. After we average the polarity score of our matched adjectives, we still need to divide this average score by the number of paragraphs/sentences, and the final sentiment polarity we obtained becomes smaller.

Afinn also relies on the human-tagged sentiment lexicon and tries to map each sentiment-bearing word appearing in the text to the lexicon and then sum up all the sentiment scores to the final sentiment score. After that, we normalize the final result to the range of -1 to 1, in accordance with VADER and TextBlob. Since the number of words in Afinn's lexicon is as twice as the number of words in TextBlob, it performs better than TextBlob with regard to the sensitivity since it has higher possibility to match words in the text.

We do not know exactly how the commercial closed source MATA calculates sentiment.

6 Conclusion

We performed sentiment analysis on eleven different events by analyzing news articles from ten different international news sources. Our results demonstrated that most events produced similar sentiment polarities across all news sources, however, there was absolute variance among sentiment score magnitudes across news sources. For a few events, there was significant polarity shift for sentiment scores among news sources. Every news publisher expressed varied sentiment toward individual event,

versus a consistent trend of negative, positive, or neutral sentiment. We were pleased not to find major differences in sentiment polarity between major news sources for the same event which could have pointed to bias and drive tensions between the populations relying on these news outlets for objective reporting. We do acknowledge that this is a simple limited metric which fails to capture the specific differences in message of any two articles even if they share the same sentiment.

There are limitations present in our study. First, some events were not covered by all news sources mentioned in this study. Additionally, some events received little coverage in general (an event may have less than 100 articles in total dedicated to it for instance). In the future, we hope to expand this study to a larger dataset and possibly employ more advanced sentiment analysis techniques.

In addition, we hope to human validate the sentiment underlying the news articles. For instance, we can invite a group of people to read news articles and rate their sentiment toward them, to see the difference between the sentiment calculated by sentiment analysis tool and humans. Furthermore, we could label a part of news articles manually with their sentiment score, and try to apply machine learning techniques to train our own model and thus make predictions.

Acknowledgements The authors would like to thank the Notre Dame Center for Research Computing for financial support and Google who kindly provided a student research grant to utilize the Google Translation API cloud service for translating La Nacion articles into English. Finally, we would like to thank Mrs. Rachel O'Hare, Mr. Richard Jones, and Mrs. Victoria St. Martin for insightful discussions on perspective in journalism.

References

1. Mejova, Y., Srinivasan, P., Boynton, B.: Gop primary season on twitter: popular political sentiment in social media. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 517–526. ACM (2013)
2. Budak, C., Goel, S., Rao, J.M.: Fair and balanced? Quantifying media bias through crowdsourced content analysis. Public Opin. Q. **80**(S1), 250–271 (2016)
3. Stimson, J.A., MacKuen, M.B., Erikson, R.S.: Dynamic representation. Am. Polit. Sci. Rev. **89**(3), 543–565 (1995)
4. Page, B.I., Shapiro, R.Y.: Effects of public opinion on policy. Am. Polit. Sci. Rev. **77**(1), 175–190 (1983)
5. Seib, P.: Effects of real-time news coverage on foreign policy. J. Confl. Stud. **20**(2) (2000)
6. Zhang, C.: International coverage, foreign policy, and national image: exploring the complexities of media coverage, public opinion, and presidential agenda. Int. J. Commun. **6**, 20 (2012)
7. Robinson, P.: Theorizing the influence of media on world politics: models of media influence on foreign policy. Eur. J. Commun. **16**(4), 523–544 (2001)
8. Tambini, D.: Fake news: public policy responses. (2017)
9. Stratfor. Russia: Duma vote paves way for Russian internet. https://worldview.stratfor.com/situation-report/russia-duma-vote-paves-way-russian-internet. Accessed 26 February 2019
10. Distinguishing biased and impartial perspectives. https://www.canadianmysteries.ca/en/4179.php. Accessed 26 February 2019
11. Durrani, M., Ali, A., Khan, T., et al.: Sentiment analysis and the complex natural language. Complex Adapt. Syst. Model. **4**

12. Sentiment bias detection in support of news credibility judgment
13. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. Found. Trends® Inf. Retr. **2**(1–2), 1–135 (2008)
14. Abdul-Mageed, M., Diab, M., Kübler, S.: Samar: subjectivity and sentiment analysis for Arabic social media. Comput. Speech Lang. **28**(1), 20–37 (2014)
15. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., Belyaeva, J.: Sentiment analysis in the news (2013). arXiv preprint arXiv:1309.6202,
16. Bautin, M., Vijayarenu, L., Skiena, S.: International sentiment analysis for news and blogs. In: ICWSM (2008)
17. Kaya, M., Fidan, G., Toroslu, I.H.: Sentiment analysis of Turkish political news. In: Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, vol. 01, pp. 174–180. IEEE Computer Society (2012)
18. Hamborg, F., Donnay, K., Gipp, B.: Automated identification of media bias in news articles: an interdisciplinary literature review. Int. J. Digit. Librar. 1–25 (2018)
19. Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J.: ICEWS Automated Daily Event Data (2018)
20. Leetaru, K. Global database of events languages and tone. https://www.gdeltproject.org
21. News API. https://newsapi.org
22. LexisNexis. Nexis uni. https://www.lexisnexis.com/en-us/support/nexis-uni/default.page
23. Bobriakov, I.: Comparison of top 6 python nlp libraries. https://medium.com/activewizards-machine-learning-company/comparison-of-top-6-python-nlp-libraries-c4ce160237eb. Accessed 14 July 2018
24. Oelke, D., Geißelmann, B., Keim, D.A.: Visual analysis of explicit opinion and news bias in German soccer articles. In: EuroVA@ EuroVis (2012)
25. Hutto, C.J., Gilbert, E.: Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International AAAI Conference on Weblogs and Social Media (2014)
26. Ribeiro, F.N., Araújo, M., Gonçalves, P., Gonçalves, M.A., Benevenuto, F.: Sentibench—a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci. 5(1), 23 (2016)
27. Loria, S.: Textblob: simplified text. https://textblob.readthedocs.io/en/dev/
28. Computational Linguistics and Psycholinguistics Research Center. https://www.clips.uantwerpen.be/pages/pattern-en
29. Nielsen, F.: A new ANEW: evaluation of a word list for sentiment analysis in microblogs (2011). arXiv preprint arXiv:1103.2903
30. Microsoft Azure. Text analytics. https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics/
31. Araujo, M., Reis, J., Pereira, A., Benevenuto, F.: An evaluation of machine translation for multilingual sentence-level sentiment analysis. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing, pp. 1140–1145. ACM (2016)

An Agent-Based Model of Social Fabric Seen as an Emergent Behavior

Gamaliel Palomo, Mario Siller, Arnaud Grignard, and Luis Alonso

Abstract The study of cities and their social phenomena have been carried out historically from a "classical focused discipline approach" by areas such as sociology, psychology, anthropology, and urbanism. However, over the past years this has evolved to a multidisciplinary approach with other areas being involved including computational sciences, complex systems, and network science. For instance, an area of interest within this new approach is the study of social fabric in cities. In this work, we study this phenomenon and present a conceptual model to establish the relationship between social fabric and other city phenomena (segregation and encounters). Then by using a proposed three layer abstraction we are able to associate emergent behaviors to social agents and the physical city. Finally, we derive an agent-based model and a study case for the city of Zapopan in Mexico where problems related to the social fabric and social segregation has been historically identified. Based on simulations and empirical evidence we conclude that the perception of the physical city affects the decision-making in mobility and the consequent segregation both related to social fabric.

G. Palomo (✉) · M. Siller
Cinvestav Guadalajara, Av. del Bosque 1145, colonia el Bajío, 45019 Zapopan, JA, Mexico
e-mail: gapalomo@gdl.cinvestav.mx

M. Siller
e-mail: msiller@gdl.cinvestav.mx

A. Grignard · L. Alonso
MIT Media Lab, Cambridge, MA, USA
e-mail: agrignar@media.mit.edu

L. Alonso
e-mail: alonsolp@media.mit.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_10

1 Introduction

Cities have been studied by traditional disciplines mainly from social science areas. From a systemic point of view cities originally were conceived as mechanical machines with specific functions, linear behaviors with well-defined equilibrium states. However, as a better understanding was achieved this abstraction was limited to explain non-linear behavior phenomena observed in cities that characterized complex systems. In fact, this emerged as a new perspective and a multidisciplinary approach to study cities.

The complexity of cities have been addressed in a number of works, including Batty [1, 2], Allen [3, 4], White [5], Portugali [6], Siller [7], etc. For instance, emergent social phenomena have been studied from the perspective of complex systems. In this work, we focus in the social fabric phenomenon from such perspective.

In the presented study, we consider an emergent behavior as a global structure which appears from local interactions which are difficult to anticipate from the knowledge of individual's behavior [8]. The social fabric phenomenon meets this definition and takes place mainly in cities through different forms. This is because most human population live in urban areas. We believe that all these forms are and how this actually occurs are still opened questions that can be addressed from the perspective of complex systems theory and agent-based modeling. Additionally, the dynamics and particularities present in Zapopan, Jalisco, and Latin America as a region can further contribute to a better understanding of this phenomenon. In this work, we study social fabric and how it is related to and affected by other city phenomena such as segregation and encounters.

2 Social Fabric

The social fabric is a social phenomenon that refers to individuals in a society; it is the result of the relationship between them and the place where they live. The denomination of social fabric has its origin in terms imported from disciplines such as medicine, where a fabric is a set of cells, and these are the functional unit of humans. When dealing with the social aspect, these cells refer to individuals and the fabric is the relationships that form between them. Due mostly to the diversity of studies in the social field, this phenomenon has also been studied in works about social well-being [9, 10], urban prosperity [11, 12], and social segregation [13].

An aspect that stimulates the creation of social fabric is the physical encounters, carried out in the urban space and found in the literature of social sciences [14], government [11, 15], ONU-habitat [16]. In those works, it is common to find that the public spaces are the means to promote encounters that finally they result in interpersonal relationships. A phenomenon that difficults the creation of social fabric is the segregation, which is defined in[17] as

> Restrictions which place limits upon contact, communication and social relations, or limits on social interaction.

We differenciate two types of segregation in literature: territorial (addressen in works by [13, 18]) where it is addressed in works where the study focuses on the formation of groups of individuals socially, economic, and culturally motivated. The second one is the real-time segregation that we understand as that caused by the differences in patterns of activities and mobility trajectories. It is often related to the economic capacity of individuals, and has been studied at least in the last two decades (Netto in works related to segregation in mobility trajectories [19–22], Schnell and Yoav studing segregation in routine activity patterns [23–25]). Complex social phenomena in cities have been addressed in recent studies using different computational techniques. Agent-based modeling (ABM) is one of the most used for spacial-related social behaviors due to its capacity of abstraction through individual behavior's rules. Recent works include Grignard et al. [26–28], Collins et al. [29], etc.

In this work, we carry out a study of social fabric in an urban environment, modeling the encounters that take place in the public space (assuming that an encounter is the fact that two or more people are co-present within a distance where interaction becomes possible).

3 Modeling Social Fabric in Cities

3.1 Conceptual Model

Due to the fact that the phenomenon we study has an origin merely from the social sciences, there exist too many definitions of it. Hence, the first step is to reach a concrete conceptual terminology. As mentioned previously, there are two behaviors that directly affect the social fabric at a macro-scale: segregation and the generation of encounters (see Fig. 1). If we go to a lower level of abstraction we find micro behaviors, for this conceptualization we use a cognitive map (see Fig. 2).

3.2 Agent-Based Model

Using the previous conceptualization as a baseline, an ABM was constructed. A detailed review of the literature led us to identify three levels of abstraction that can help to systematically model social behaviors in a city: the physical city, level of social agents, and level of emergent phenomena. This is abstracted as the model shown in Fig. 3. In the following subsections, we describe the details using the ODD protocol [30].

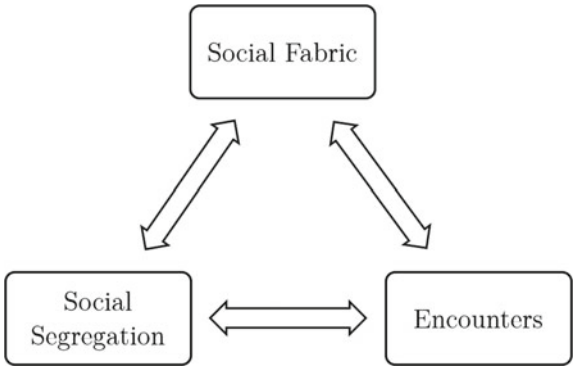


Fig. 1 Macro-level relationship between the phenomena of social fabric, social segregation and encounters

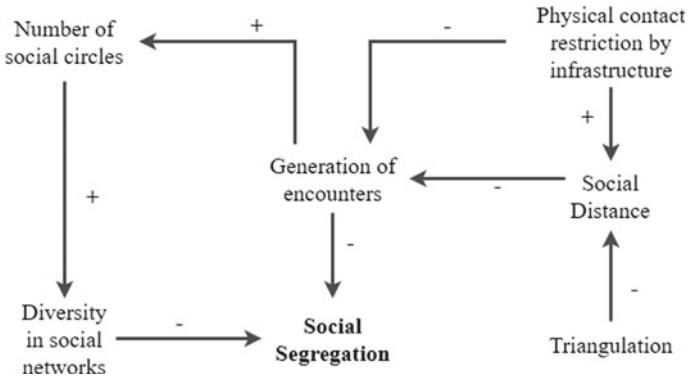


Fig. 2 Relationship between the factors that affect the generation of encounters and social segregation

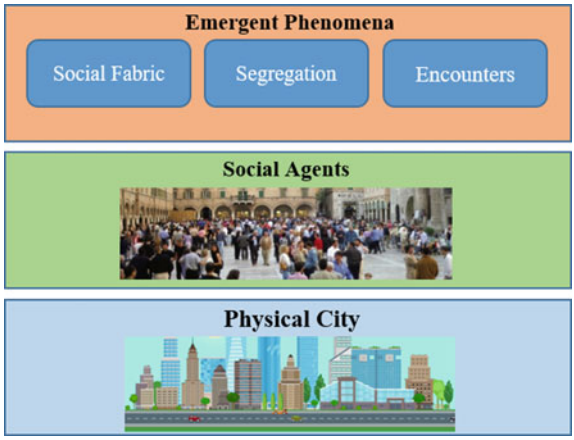


Fig. 3 Multilayer model composed by three levels of abstraction

3.2.1 Entities, State Variables, and Scales

We describe below the variables used in the model, separated according to our multi-layer abstraction.

Physical city. In terms of Multi-agent Systems, the physical infrastructure of a city provides an environment: spaces, roads, facilities, and restrictions for social agents to move and interact with. We use the entity **place of interest** to refer to a space within the city where an agent can move to and perform daily activities (home, work, market, school, gym, etc.). They are modeled as non-moving agents with geographic location. Also we have **roads** which are the ways of mobility used by people (streets, ways, footpaths). These agents are part of a georeferenced network, where the street intersections are the nodes, thus the road network R can be defined as a 2-tuple (E, V) where E is the set of road agents and V is the set of street intersections. Each physical agent owns environmental characteristics: lighting, paving type, accessibility, sidewalk.

Social agents. Social agents or city dwellers represent the entities from which emergent behaviors are originated and observed in the Physical City. In other words, the agents produce all urban dynamics observed in the city. For this, urban agents involve routines and plans as they do their daily activities. As a consequence, they are considered as proactive agents. The routines and planning process implicate some level of city perception. Their attributes include *location*, *routine:* a list of places that are part of a daily routine, *destination:* a current objective part of a routine, *shortest_path:* a calculated shortest path to achieve a routine using a perception value.

Emergent phenomena. At this level of abstraction, we study the emergence of three phenomena: a network of encounters, pedestrian patterns formed as a tissue of individual routes, and social fabric. Those correspond to the macro-level conception presented before (Fig. 1).

Network of Encounters

It is represented as a graph G where the nodes are agents, and the edges are the encounters, thus $G = (agents, encounters)$.

Routes

As a result of the agents planning process, paths to follow emerge. When a number of agents are considered, a large tissue of paths is constructed, and this could reveal information about segregation and the places that potentially could be places for encounters.

Social Fabric

For this model, the social fabric is studied as a phenomena directly affected by segregation and encounters. A city which does not have enough number of public spaces in its infrastructure will not promote encounters, arising as a result of a negative scenario for social fabric.

3.2.2 Process Overview and Scheduling

An urban agent A_i is a 2-tuple (PA_i, K_i) where

- Perception is defined as the cuantification process of how urban agents differentiate urban environment quality in terms of urban equipment. The lowest quality can be represented as 0 and the highest as 1. PA_i is the perception of agent A_i and it is represented as a set of vectors containing the knowledge about the roads, thus the length of PA_i is $|E|$.
- K_i is a routine carried out as daily activities. It is a sorted list of *places of interest* that A_i will visit.

As each agent has a determined routine, they build individual paths P_i which describe their mobility. For each A_i we use the sum of the components of PA_i as weights of the road network R. A shortest path is then applied to find a path for the routine. This process is shown in Algorithm 1.

For this paper, we use the copresence defined in [22] as agents *"positioned within a field where they can perceive the presence of another agent"*, using a circle built around an agent, an encounter is generated when two agents are into the *circle of copresence* of the other [31] as seen in Fig. 4. We assume no limit of the number of encounters an agent could have. It is evident that a variation in these circles' radius is directly reflected on the network's density.

Algorithm 1 buildPath

Require: A set PA_i, a sorted list of places K_i and a graph R and its weights.
Ensure: A path P_i
1: **for** $j = 1$ to $K_i.lenght - 1$ **do**
2: *add(shortestPath(R,K(j),K(j+1)))* to P_i
3: **end for**

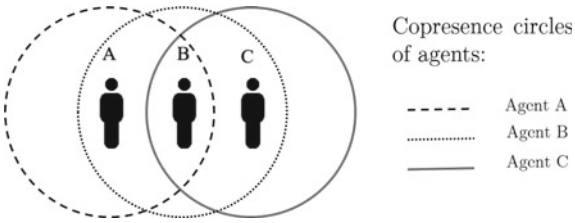


Fig. 4 Copresence circles. Agents A and B have an encounter as B and C also have. A and C do not have an encounter

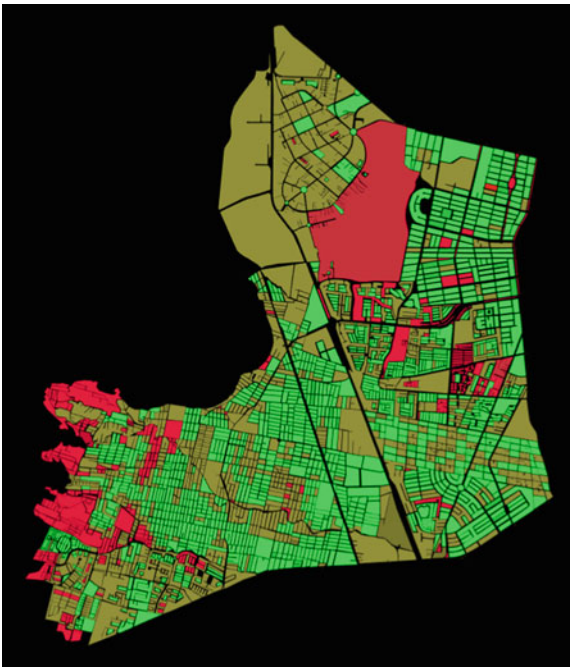


Fig. 5 Example of perception of an agent for the Zapopan case study. Red color represents a bad perception about that physical space and green good perception

3.2.3 Initialization and Input Data

Agents at the physical level were geographically distributed using actual data from government reports [11], from collaborations with the Public Spaces Office of the Zapopan Government [32], and from the National Institute of Statistics Geography and Informatics (INEGI) [33, 34]. We simulated a total of 500 social agents. The initial location (home) of each social agent was randomly fixed and the routine activities were selected. We consider that agents move all of them at the same speed which is 1.4 m/s. Figure 5 shows an example of the simulation environment used for this study and a social agent's perception implemented in GAMA modeling language [35]. For a better visualization, colors are used to distinguish the perception of an agent (red is a bad perception and green a good perception).

4 Case Study: The Miramar Region

The region of Miramar is located to the south of the city of Zapopan, in Jalisco, México. It maintains until 2010 a total of 28219 inhabitants [36]. Its selection as a case study is due to the high value of research, since it is a region with high rates of violence, gang formation, poverty, social segregation, and family fragmentation. These factors have historically affected the social fabric of the region.

4.1 Mobility Patterns and Its Relationship with Perception

An interesting result for this case study was observed through mobility patterns. Decisions in mobility are affected by a single perception about the physical city, and social agents' use of infrastructure is altered as a consequence. With this model, we can analyze specific parts of the physical city and understand which ones are mostly used by the agents in simulations under different city configurations. Running this case study we compared the impact of using a perception value and observed that infrastructure use changes, more streets are used when people are affected by this value. Mean number of streets used in all 100 experiments are shown in Fig. 6 as a boxplot. What these results show is that using a perception value for planning process in the social agents caused a more use of infrastructure for our case study.

4.2 Encounters

Encounters are affected by mobility. As described earlier in the conceptualization, social fabric is influenced by encounters, and mobility patterns may both promote or difficult this phenomenon. We also run simulations of this case study analyzing encounters derived from different perception configurations. Figure 7 summarizes our results using the Zapopan case study.

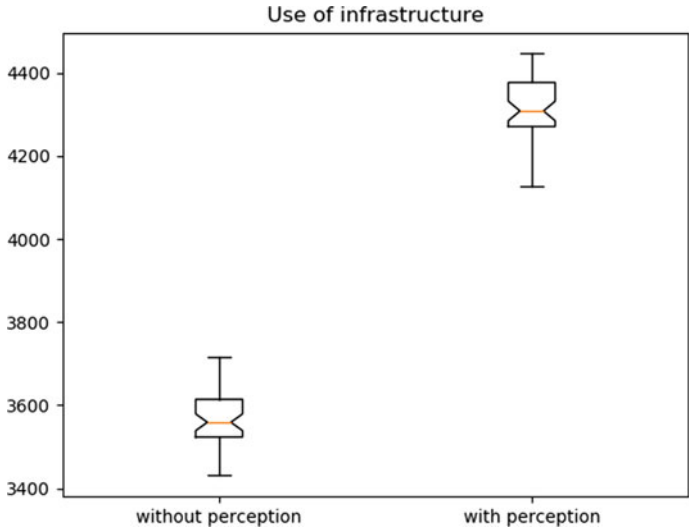


Fig. 6 Boxplot of the use of physical infrastructure affected by perception in social agents. Left values are the mean number of used streets for each of the 100 experiments

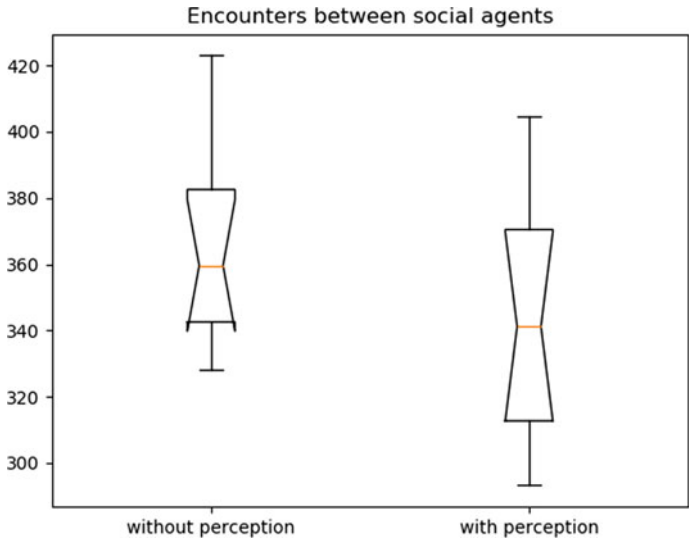


Fig. 7 Boxplot of encounters between social agents. Left values are the number of encounters carried out for each of the 100 experiments

5 Conclusion

The model presented in this work considers as a basis three levels: physical, social, and emergent. We think these could be used as a generic method to address other phenomena in cities. The level of social agents allows us to individually model the mobility decisions of people, while incorporating routine activities. Mobility decision-making is based on a perception factor that takes into account only characteristics of the physical layer, such as street lighting and condition. The model was implemented and simulated in GAMA and results of the case study about perception influence on mobility and encounters were presented. This work represents an attempt at modeling social fabric in cities through encounters and perception. We do not take into account the difference between positive and negative encounters. However, we believe this can be enhanced and strengthened in future works by considering more factors to agent's decision-making (not just infrastructure) such as collective knowledge and other phenomena such as crime, also a difference between positive and negative perception values can be considered. Although results by the simulations in the Zapopan case study seem to show viable scenarios for the promotion of encounters and social fabric, empirical knowledge about security aspects and its impact on mobility and generation of encounters tell us about another reality. Our model simulations take as input multiple random variables, this might impact the realism of any conclusions and it should be taken into account in future jobs where access to specific data is greater, improving the realism and validation of the simulation.

Acknowledgements This work has been partially supported by CONACYT, G. Palomo's Grant No. 613707.

References

1. Batty, M.: The New Science of Cities. Mit Press (2013)
2. Batty, M.: Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies, pp. 1041–1071. Springer New York, New York, NY (2009)
3. Allen, P.: Cities and regions as evolutionary complex systems. J. Geogr. Syst. **4**(01), 103–130 (1995)
4. Allen, P.: Cities and Regions as Self-Organizing Systems Models of Complexity. Taylor and Francis Group (1997)
5. White, R., Guy, E., Inge, U.: Modeling Cities and Regions as Complex Systems From Theory to Planning Applications. MIT Press (2015)
6. Portugali, J.: Complexity. Springer, Cognition and the City (2011)
7. Siller, M., Stibe, A.: Programmable cities: a new ICT approach. In: Short Paper Proceedings of the Workshop on Empowering Cities for Sustainable Wellbeing (ECSW 2016) Adjunct to the 11th International Conference on Persuasive Technology (PERSUASIVE 2016). Salzburg, Austria (2016). 5 April 2016
8. Boccara, N.: Modeling Complex Systems. Springer (2010)
9. Helliwell, J.F., Putnam, R.D.: The social context of well-being. Philosopical Trans. R. Soc. London. Ser. B, Biol. Sci., **359**(09), 1435–1446 (2004)
10. Diener, E.: Subjective well-being. Psychol. Bull. **95** (1984). SSRN: https://ssrn.com/abstract=2162125
11. López-Moreno, E., Velásquez, E.: Estrategia territorial para la prosperidad urbana zapopan **2030**(7) (1993)
12. FIABCI International and UN-Habitat.: City prosperity initiative—perception index (2019). Retreived from https://unhabitat.org/urban-initiatives/initiatives-programmes/global-public-space-programme/
13. Pfannenstein, B., Anacleto-Herrera, E., Sevilla-Villalobos, S.: La ciudad cerrada y su diversificación como reto del Área metropolitana de guadalajara, méxico. Revista de Geografía Norte Grande **68**, 163–184 (2017)
14. Murcia, E.I.T.: El sentido del tejido social en la construcción de comunidad. Polisemia **6:9**(03) (2013)
15. SOMA Estrategia Socioambiental and Proyectos Estratégicos de Zapopan.: Diagnóstico social de la colonia lomas del centinela. Technical Report 2, SOMA Estrategia Socioambiental and Proyectos Estratégicos de Zapopan **10** (2017)
16. UN-Habitat.: Global public space programme (2019). Retreived from https://unhabitat.org/urban-initiatives/initiatives-programmes/global-public-space-programme/
17. Freeman, L.C.: Segregation in social networks. Sociol. Methods Res. **6**(4), 411–429 (1978)
18. Lim, M., Metzler, R., Bar-Yam, Y.: Global pattern formation and ethnic/cultural violence. Science **317**(5844), 1540–1544 (2007)
19. Netto, V.M.: Socio-spatial networks—social segregation as a real-time phenomenon (2001)
20. Netto, V.M.: Social networks in the city, or the urban condition of coexistence. V!RUS (2010)
21. Netto, V.M.: Maíra Soares Pinheiro, and Roberto Paschoalino. Segregated networks in the city **39**(11), 1084–1102 (2015)
22. Vinicius, M.: Netto, João Vitor Meirelles de Miranda, Maíra Pinheiro, and Henrique Lorea. A Temporal Geography of Encounters, Cybergeo (2018)
23. Schnell, I., Yoav, B.: The sociospatial isolation of agents in everyday life spaces as an aspect of segregation. Ann. Assoc. Am. Geogr. **91**(4), 622–636 (2001)

24. Schnell, I., Benjamini, Y.: Globalisation and the structure of urban social space: the lesson from Tel Aviv. Urban Stud. **42**(13), 2489–2510 (2005)
25. Farber, S., O'Kelly, M., Miller, H.J., Neutens, T.: Measuring segregation using patterns of daily travel behavior: a social interaction based model of exposure. J. Transp. Geogr. **49**, 26–38 (2015)
26. Grignard, A., Drogoul, A.: Agent-based visualization: a real-time visualization tool applied both to data and simulation outputs. In: The AAAI-17 Workshop on Human-Machine Collaborative Learning (2017)
27. Grignard, A., Alonso, L., Taillandier, P., Gaudou, B., Nguyen-Huu, T., Gruel, W., Larson, K.: The Impact of New Mobility Modes on a City: A Generic Approach Using ABM: Proceedings of the Ninth International Conference on Complex Systems, vol. 07, pp. 272–280 (2018)
28. Grignard, A., Macià, N., Alonso Pastor, L., Noyman, A., Zhang, Y., Larson, K.: Cityscope andorra: a multi-level interactive and tangible agent-based visualization. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18, pp. 1939–1940. Richland, SC (2018). International Foundation for Autonomous Agents and Multiagent Systems
29. Collins, A.J., Cornelius, C.V., Sokolowski, J.A.: Agent-based model of criminal gang formation. In: SpringSim-ADS, p. 9. Society for Modeling and Simulation International (SCS) **04** (2017)
30. Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Visser, R.V.U., DeAngelis, D.L.: A standard protocol for describing individual-based and agent-based models. Ecol. Model. **198**(1), 115–126 (2006)
31. Hamill, L., Gilbert, N.: Social circles: a simple structure for agent-based social network models. J. Artif. Soc. Soc. Simul. **12**(2), 3 (2009)
32. Hidalgo-Luna, L.: Estado de la infraestructura de calles de zapopan (2018). GIS Data in vectorial shape format
33. National institute of statistics geography and informatics.: Manzanas inv2016 (2016). Data retrieved from INEGI, https://www.inegi.org.mx/app/descarga/default.html
34. National institute of statistics geography and informatics.: Frentes de manzana inv2016 (2016). Data retrieved from INEGI, https://www.inegi.org.mx/app/descarga/default.html
35. Grignard, A., Taillandier, P., Gaudou, B., Vo, D.A., Huynh, N.Q., Drogoul, A.: Gama 1.6: advancing the art of complex agent-based modeling and simulation. In: Boella, G., Elkind, E., Savarimuthu, B.T.R., Dignum, F., Purvis, M.K. (eds.) PRIMA 2013: Principles and Practice of Multi-Agent Systems, pp. 117–131. Berlin, Heidelberg (2013). Springer Berlin Heidelberg
36. Información sociodemográfica por colonia, municipio de zapopan (2010). Data retrieved from https://www.iieg.gob.mx

Deep Agent: Studying the Dynamics of Information Spread and Evolution in Social Networks

Ivan Garibay, Toktam A. Oghaz, Niloofar Yousefi, Ece Çiğdem Mutlu, Madeline Schiappa, Steven Scheinert, Georgios C. Anagnostopoulos, Christina Bouwens, Stephen M. Fiore, Alexander Mantzaris, John T. Murphy, William Rand, Anastasia Salter, Mel Stanfill, Gita Sukthankar, Nisha Baral, Gabriel Fair, Chathika Gunaratne, Neda B. Hajiakhoond, Jasser Jasser, Chathura Jayalath, Olivia B. Newton, Samaneh Saadat, Chathurani Senevirathna, Rachel Winter, and Xi Zhang

Abstract This paper explains the design of a social network analysis framework, developed under DARPA's SocialSim program, with novel architecture that models human emotional, cognitive, and social factors. Our framework is both theory and data-driven, and utilizes domain expertise. Our simulation effort helps understanding how information flows and evolves in social media platforms. We focused on modeling three information domains: cryptocurrencies, cyber threats, and software vulnerabilities for the three interrelated social environments: GitHub, Reddit, and Twitter. We participated in the SocialSim DARPA Challenge in December 2018, in which our models were subjected to an extensive performance evaluation for accuracy, generalizability, explainability, and experimental power. This paper reports the main concepts and models, utilized in our social media modeling effort in developing a multi-resolution simulation at the user, community, population, and content levels.

I. Garibay (✉) · T. A. Oghaz · N. Yousefi · E. Ç. Mutlu · M. Schiappa · S. Scheinert · C. Bouwens · S. M. Fiore · A. Mantzaris · A. Salter · M. Stanfill · G. Sukthankar · N. Baral · C. Gunaratne · N. B. Hajiakhoond · J. Jasser · C. Jayalath · O. B. Newton · S. Saadat · C. Senevirathna · R. Winter
University of Central Florida, Orlando, USA
e-mail: igaribay@ucf.edu

G. C. Anagnostopoulos · X. Zhang
Florida Institute of Technology, Melbourne, USA

J. T. Murphy
University of Chicago, Chicago, USA

W. Rand
North Carolina State University, Raleigh, USA

G. Fair
University of North Carolina at Charlotte, Charlotte, USA

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_11

1 Introduction

Emerging Online Social Networks (OSNs) have revolutionized the public information environment in an unprecedented way. Thus, it is crucial to study the process of the spread and evolution of online information to understand the reach and impact of news, ideas, and knowledge in OSNs. An accurate and scalable computational simulation of this process could potentially help combat misinformation campaigns by adversaries, efficiently deliver critical information to local populations during disaster relief operations, and contribute to social construction and policy designs that rely on information dissemination.

Despite progress in this field of research, current computational approaches to social and behavioral simulations have not been well positioned to uncover the underlying dynamics that explain the inner workings and reasons for the selection and diffusion of information in online social platforms. Current approaches to online social dynamic simulations fall into three main categories: (I) the statistical analysis and modeling of a particular phenomenon such as "information evolution" using a particular dataset and fitting a statistical model to the data, for instance, [1]; (II) the statistical physics approach using the Agent-Based Model (ABM) simulation as an extension of dynamic equation modeling; and (III) approaches using ABMs through "translating" a theoretical model into the agent-based framework [2]. Although the first approach can be used in econometrics to predict the economic outcomes a few months ahead, it fails to predict rich system dynamics (such as during a financial crisis) correctly, and does not take complex human dynamics into account, though it models a single dataset accurately. The second approach is typically based on a "Brownian agent" [3]), and applies agent-based models in a very different capacity than the more standard practice of using them as a complex systems modeling tool. The Brownian agents are restricted by the stochastic physics framework in which they are embedded, resulting in less capability in capturing complex dynamics behavior. Lastly, the third approach focuses on replicating a single phenomenon; the agents mirror a single set of equations that are focused on an observable macro-pattern instead of the deep cognitive mechanisms that drive human behavior. As a result of not modeling deep human emotional, cognitive, and social factors that determine social behavior outcomes, all three approaches lead to potential over-fitting on a single dimension of data.

Although frameworks like the Agent-Zero [4] and Homo Socialis [5] offer theoretical solutions to modeling the true complexity of human dynamics driven by deep neurocognitive underpinnings that are at the core of any human social activity, including the spread and evolution of information, these deep models are limited to modeling conceptual problems. Furthermore, simpler models are preferred for real-world problems as they allow for the parameters tuning directly associated with modeling a particular dataset, simpler models cannot simultaneously replicate multiple complex phenomena such as various aspects of human dynamics, including information cascading, gatekeeper's identification, information evolution, and persistent minorities.

To overcome the shortages in modeling the online social platforms, the Defense Advanced Research Projects Agency (DARPA) announced the "Computational Simulation of Online Social Behavior (SocialSim)" program to develop innovative technologies for high-fidelity computational simulation of online social behavior. Responding to this DARPA program, our team proposed and implemented a novel simulation framework that enables revolutionary advances in simulation of information spread and evolution on social media on a large scale. Our team accomplished this by (I) modeling social dynamics using a network of computational agents endowed with deep neurocognitive capabilities, (II) creating a family of plausible social dynamic models assembled from modularized sub-components, and (III) utilizing machine Learning algorithms and HPC Cloud Computing for model discovery, refinement, and testing.

This paper explains the design and concepts related to our framework and agent-based models for social network analysis using large volumes of data from GitHub, Reddit, and Twitter with the aim of better understanding of online social behaviors. We proposed the Deep Agent Framework (DAF), which operationalizes social theories of human behavior and social media into optimized generative simulation capabilities. Additionally, we developed Multiplexity-Based Model (MBM) which is an agent-based model designed based on concepts from graph theory, that simulates online social network evolution.

2 Challenge Problem Description

The challenge problem was designed to develop a multi-resolution simulation at the content, user, community, and population levels. Thus, a total of 57 accuracy measurements and metrics were designed to evaluate the participant models. The challenge evaluation procedure applied a combination of various metrics and measurements over four dimensions: accuracy, generalizability, explainability, and experimental power. Table 2 in the Appendix section contains the evaluation metrics and a performance comparison between the agent-based models for the community, content, population, and user-level interactions.

3 Methodology

Our generative model of social media (Fig. 1a) employs ABMs to simulate social dynamics via embedded agents as user profiles in OSN platforms, and "deep agents" as the platform users. The deep agent concept is adopted from the Agent-Zero [4] and the Homos Socialis [5] frameworks, which leads to deep agent modeling. Considering the Agent Zero framework, a cognitively plausible agent must account for three dimensions: Emotions (leading to ortho-rational behavior), Bounded Rationality, and Social Connectivity. Therefore, the social media users in our DAF framework

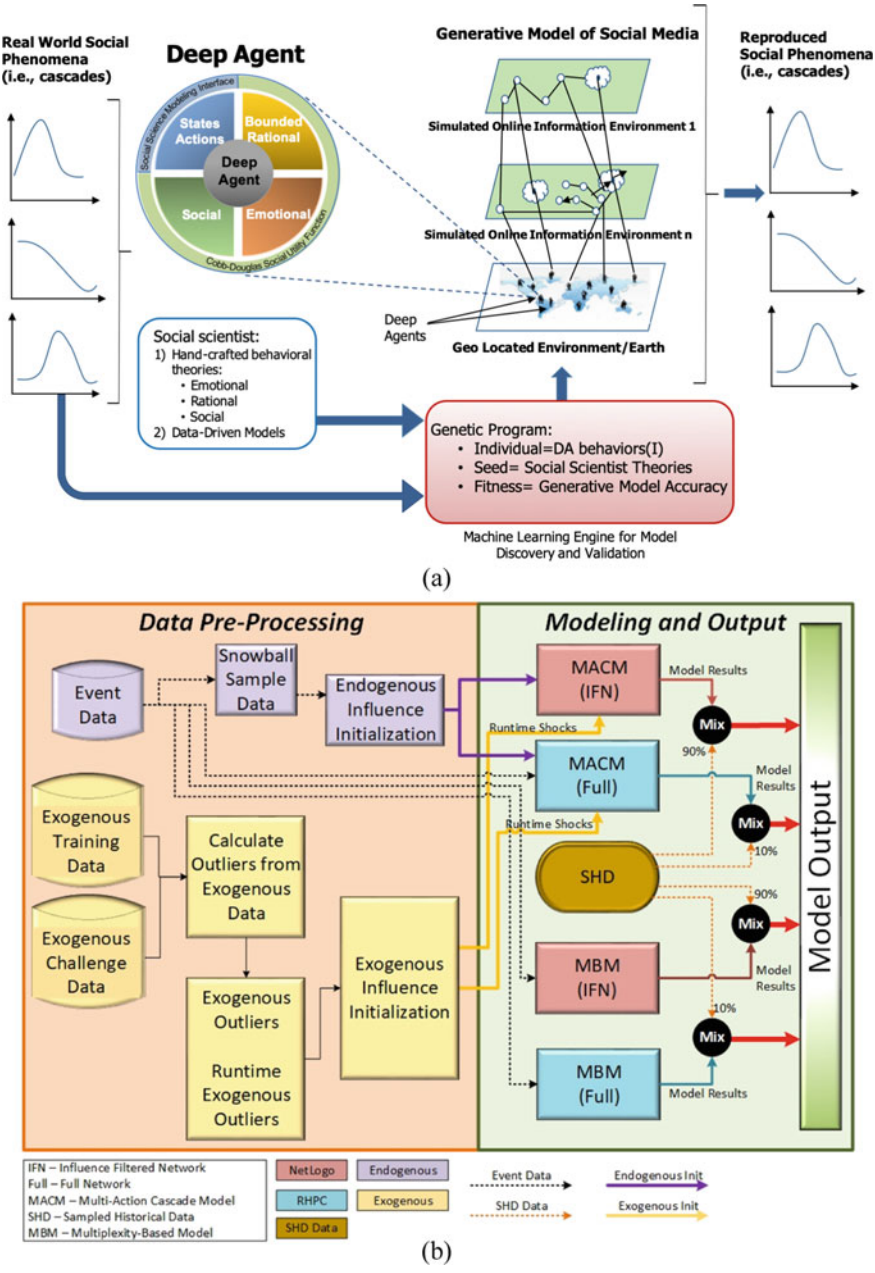


Fig. 1 Deep Agent Modeling Framework (DAF). **a** The framework maps the agents' rules of behavior to the real-world data using the hand-crafted behavioral theories (emotional, rational, and social) and the data-driven models to initialize the generative models. **b** Architecture of the DAF is depicted, which consists of the work-flow of the data process (endogenous and exogenous data), model simulation, model mixing, model evaluation and tuning

were modeled to possess these three dimensions, and are referred to as the deep agents in this paper. In contrast, "shallow" models are deficient in modeling deep human characteristics that would determine social behavior outcomes. Additionally, shallow models focus on fitting equations into a single phenomenon of interest, which results in models that are brittle and potentially over-fitted for a single dimension of the data.

The agents' interactions according to the embedded rules result in specific outcomes at the population, community, user, and content levels, and provide information regarding the agents' decision processes. The agents' relative actions are derivable by applying particular metrics, (addressed in Sect. 4.1) and the Appendix sections, which provide simple statistics for the cascade and group behaviors.

The designed Deep Agent Modeling Framework (DAF), depicted in Fig. 1b, allows to create a family of modular sub-components from which multiple plausible models can be systematically assembled, tested, and validated. Discovering every rule of behavior is possible through employing the genetic programming evolutionary model discovery method (red box in Fig. 1a) as in [6], which explores every possible space related to the set of agents' rules of behavior. This provides strong inferences of human behavior using computational simulations. The search can be guided by model accuracy, as measured by comparing model outputs with real-world social dynamics data. Assembling the framework with the explained pieces, this framework unleashes the power of combining massively parallel computing, data analytics of large datasets and, machine learning to assist in mixing and matching sub-models in a semi-automated way; this allows for exploration and testing that validates tens of thousands of models against a large set of target behaviors.

We developed the DAF simulation tool to help answering the questions related to properties of information exchange dynamics on online social media in population, content, user, and community levels. Additionally, we proposed the Multiplexity-Based Model (MBM), which captures social network evolution based on the preferential attachment, attention, and recency cognitive bias. We mix and compare our model with other theory-driven models designed by our team of researchers. These models include Multi-Action Cascade Model (MACM) and Sampled Historical Data (SHD). The mixing, evaluation, and comparisons are provided for the three models and their variations: MBM-Influenced Filtered Network, MBM, SHD, MACM-Influenced Filtered Network, and finally MACM. In this paper, we applied hand-crafted behavioral theories and data-driven models directly to feed the agent-based models.

3.1 Deep Agent Framework: Architecture and Analysis

The architecture of the framework (Fig. 1b) consists of the data pre-processing, and the modeling and output phases. The data pre-processing phase includes data sampling, extracting influential users through normalized transfer entropy, extracting external shocks to the system represented as outliers, and initializing the endoge-

nous and exogenous influences with the extracted users and shocks. The endogenous influence initialization involves the snowball sampling of the influential users and their relationships, to generate the static endogenous network, the network dynamics, and the network message information. Snowball sampling with normalized transfer entropy was used to extract the influential relationships from the event data, which was used beside the extracted activity disparity distributions of the endogenous relationships to build the static exogenous network. Using the extracted influential relationships, we used the most recent activities to build the initialized network dynamics and the last m messages to filter the network message information.

Each model has two variations, a full network simulation and an Influence Filtered Network (IFN) simulation. The full models take in the entire network of the event data in the initialization phase; however, the IFN models are initialized using only the filtered influential users to simulate the user interactions. We used all three outputs of the endogenous influence initialization phase as the inputs to the MACM and MACM-IFN models.

For the exogenous data initialization process, we extracted the outliers from the exogenous training and challenge data sources using different filtering methods. This process includes applying Fourier Transform (FT) on each different time-series, employing a moving window Magnitude filter and Butterworth filter on the FT of each different time-series to filter the anomalies, and the binary digitization of the anomaly time-series. We applied transfer entropy to the endogenous outliers and extracted the activity disparity distributions of the exogenous relationships from the filtered data to build the static exogenous network. Finally, we generated the network dynamics via extracting the last activity disparity of the exogenous relationships from the exogenous outliers. The generated static exogenous network and the network dynamics were fed into MACM and MACM-IFN models. The inputs to the MBM and MBM-IFN models were the entire network of the event data and the sampled data related to the last x weeks, respectively.

The modeling and output phase contains our models' simulations and mixing of the models, model evaluation, and model tuning. The five variations of the generative models were implemented using Netlogo and RHPC coding environments. The model mixing strategy refers to merging the simulation outputs of the MACM and MBM models with the simulation output from the SHD model. The mixing strategy combines the output of the Full models with 10% of the simulated user interactions using the SHD model. The IFN models take 90% of the simulated interactions using the SHD model as input combined with the model results. More information on mixing strategy is provided in Sect. 3.2.3.

3.2 Agent-Based Models

The agent-based models in this paper are generative rule-driven models, designed based on the social theory on Diffusion of Information (DoI) and the user actions in OSNs. Although each social media platform has specific user interactions, we can

Table 1 Table of symbols and definitions used in this paper

Variable definition	Symbol
Activity/action	A
Message content	M
Conversation	C
External shock	S
Fitness	F
Link/edge	L
Node degree	K
Stochastic noise	ε
Time-step/tick	t
Transfer entropy	T
User/node	U
Vote to a content	V

refer to four fundamental user activities observable in any OSN platform: Create, Post, Vote, and Follow. Along with this, there are four entities in any OSN environment: Actor, Content, Action, and Space. This viewpoint of the actions and entities allows behavior to be represented across multiple social media environments, referred to as the common language or the ontology of user actions and entities. The agent-based models in this paper are designed based on the four traditional DoI models: the Threshold Model (TM) [7–9]; the Independent Cascade Model (ICM) [10, 11], the Bass Growth Model [7, 12], Rand Agent-Based Model [13], and the Complex Contagion Model [14].

3.2.1 Multiplexity-Based Model

The MBM model simulates social network evolution by multiplex networks, which have multi-layer network structure with possible shared nodes among different layers [15]. As MBM is designed based on concepts from graph theory, we refer to OSN users as nodes and user interactions as links. The model consists of a directed bipartite graph with bipartite pairs of users-repositories for GitHub, users-subReddits for Reddit, and users-users for Twitter, distinguished by multiple layers. Each of the separate user actions in the platform generates a sub-graph and the combination of the actions generates the whole network structure. The set of user actions in this model are conversation creation, contribution, vote , and follow, which can be formalized as $(C_i \notin \{C\})$, $(C_i = C_j, M_i \notin \{M\})$, $(V_i \notin \{V\})$, and $(L_i \notin \{L_{U_j}\})$ respectively, where indices are representative of users that perform the action, and $\{C\}$, $\{V\}$, and $\{L_{U_j}\}$ refer to the sets of all conversations, votes, and links to followers of the user in the model up to the current time-step.

The cognitive factor of MBM refers to the information overload resulting in higher attention to recent activities and active users. In other words, MBM considers the recency bias affecting OSN users' decision-making processes to possible propagation of information. This concept has been designed in the model in terms of age and fitness values, such that the user's influence decays in time. Content targets that have been recently the object of actions, and the users that have recently acted, see their fitness decrease the least, whereas these values for inactive users are decreased the most. This results in paying higher attention to the influential users and targets, but allowing their fitness values to be reduced in popularity over time, and eventually to be supplanted by newer elements. Reaching a certain age leads to the node removal from the model node-set. As a result, the model's predictions are most affected by recent trending activities with higher attention to more active users. Accordingly, the driving forces of MBM are preferential attachment and preferential decay, both as functions of the node's degree $k \in \{1, \ldots, K\}$ and age $a \in \{t_0, \ldots, t_{max}\}$. The propagation of information from user U_j to neighboring user U_i can be represented as

$$(U_j, A_j, a_j) \rightarrow (U_i, A_i, a_i) \tag{1}$$

The model initially comprises $|U|$ nodes, with each node as a user U having L_U number of links. Each node in the graph is assigned a string fitness of $F = 1$ that models the node's influence on the growth of the network and decreases as a function of time, node age, and activity history [16]. Accordingly, node's age value can be calculated as

$$a_j \leftarrow a_j + \Big(1 - (t_{c_j} - t_{p_j}) * F_j)\Big) \tag{2}$$

where t_{c_j} is the current time and t_{p_j} refers to the previous activity time for user U_j, and F_j is user's current fitness value. The value of fitness for each node can be calculated as

$$F_j = \frac{|A_j|}{a_j} \tag{3}$$

where $|A_j|$ is the number of actions for user U_j and is equal to the number of user interactions or degree k_j. In other words, fitness is essentially a simulation of a user's productivity. User's fitness is increased by the diversity of activities, the shortness of time-span between its activities, and the fitness of all the interactions a user builds.

MBM network grows at each time-step by the successive addition of new nodes to the model node-set and new edges to the edge-set. Node addition ratio was extracted from the input data. New links emerge between the nodes with higher fitness values as a result of the preferential attachment.

In summary, the model can be broken down into three steps that are performed at each time-step: (I) Node selection: selecting a set of nodes from the set of all current model node-set and potential nodes that can be added to this set, referred to

here as $\{U_t\} \subset \{U\}$. (II) Interaction: building the interactions between the bipartite node pairs in $\{U_t\}$ and the rest of nodes in the model such that the pairs are matched according to a likelihood distribution weighted on the nodes' fitness values. A sub-graph associated with a behavior activity is assigned to the selected pair based on a likelihood distribution determined by the popularity of action types. The layer fitness score for each sub-graph gets re-calculated to predict how popular each action remains. (III) Update: updating the node and link fitness scores, local and global degree values, layer fitness scores, node and link ages, and removing the nodes and edges from the model according to fitness decay. In this step, the new age value for each node can be calculated by

$$a_j \leftarrow a_j + \Big(1 - (t_{c_j} - t_{p_j}) * (t_{c_j} + 1)\Big) \tag{4}$$

3.2.2 Multi-action Cascade Model

The MACM model [17, 18] is a cognition-based agent-based model that simulates the diffusion of information through the network using individual-scale probabilities of actions derived from the Independent Cascade Model. The cognitive factor of this model refers to the information overload resulted by vast amount of social media activity bombarding users' attention and affecting their decision-making processes through prioritizing and preference to possible propagation of information.

Using transfer entropy analysis on user-user and user-exogenous force influences, this model measures the probabilities of actions for user U_i related to endogenous and exogenous forces as

$$q = \mathbb{P}_t(U_i|U_j) = \mathbb{P}_{t-1}(U_i|U_j) + \frac{\varepsilon_{i,j}}{1 + T_{i,j}} \tag{5}$$

$$p = \mathbb{P}_t(U_i|S) = \bigcup_{s \in S}\Big(\mathbb{P}_{t-1}(U_i|s) + \frac{\varepsilon_{i,s}}{1 + T_{i,s}}\Big) \tag{6}$$

where neighboring user U_j is active in a conversation, $s \in S$ refers to external shock, $T_{i,j}$ is the transfer entropy from user U_j's action to user U_i's action, $T_{i,s}$ is the transfer entropy from external shock s to user U_i's action, and ε indicates noise relative to activity changes of the two users.

MACM considers the evidence that the internal and the external forces can define the dynamics of different event types causing the spread of information, and the evolution of a content as it spreads through the social network. The user actions in this model are conversation creation, contribution, sharing, and deletion, which can be formalized as $(C_i \notin \{C\})$, $(C_i = C_j, M_i \notin \{M\})$, $(C_i! = C_j, M_i = M_j)$, and $(C_i = C_j, M_i = \varnothing)$, respectively, where indices are representative of users that perform the action, and $\{C\}$ and $\{M\}$ refer to the sets of all conversations and contents in the model up to the current time-step. The propagation of information is modeled as a

message considering the influences from the neighboring nodes, the action type, the target conversation, and the content genome, such that

$$(U_j, A_j, C_j, M_j) \rightarrow (U_i, A_i, C_i, M_i) \tag{7}$$

where $Cont_A$ is the massage content in conversation C, A refers to the action, and U represents the user. Additionally, the user interactions are conditioned on the processing of the received messages from the connected nodes filtered down as a result of the cognitive overloading. Accordingly, MACM agents calculate the probability of performing action A_i in response to action A_j as the union of probabilities of actions based on the endogenous and exogenous forces as

$$\mathbb{P}_t(U_i|U_j, S) = \mathbb{P}_t(U_i|U_j) \cup \mathbb{P}_t(U_i|S) \tag{8}$$

3.2.3 Sampled Historical Data

The SHD model [19, 20] is a replay-based data mixture model designed based on the seasonality characteristic of the OSN user activities and the hypothesis that the users exhibit repetitive patterns. This model extracts the most recent activities from the training data to provide the information related to the user interactions and edge formations in the network, and predicts the future user interactions according to the same types of activities in the past. We employed the SHD model to simulate the less-active users that hold little influence on the network and have been removed in the filtering processes. The mixing strategy using the SHD model refers to the following: I) extract the active and less-active unique users from the social network data; II) predict the activities associated with the active users using MBM and MACM models; III) predict the activities associated with the less-active users using SHD; and IV) append the SHD simulated low-activity users to the events simulated by the MBM and MACM models.

4 Dataset Description

The challenge goal was to model social structures and their day-to-day changes, and accordingly, simulate the time-series network evolution of GitHub, Twitter, and Reddit social environments for the three domains of interest (CVEs, Cyber Threats, and Cryptocurrencies). The datasets for the challenge were provided by Leidos, and are explained in detail below.

The GitHub social network data contained information from the years 2015 to 2017. A total of 33,570 cryptocurrency-related repositories were associated with or included target coin names or keywords in their descriptions, and 1,193,370 events matched with those repositories. 5,505,496 cybersecurity repositories and 214,074,771 events were selected as well as 186,190 software vulnerability-related

repositories and 26,777,997 events. The Twitter social network included data for the years 2016 to 2017, and with a total of 7,382,724 cryptocurrency-related tweets, retweets, and quotes. These values were 30,704,025 and 74,074 for the cybersecurity and software vulnerability domains, respectively. The Reddit dataset included submissions and comments matching keywords for the years 2015–2017. The cryptocurrency-related data contained 299,401 submissions and 3,370,547 comments. These values were 2,442,942 and 33,629,588 for the cybersecurity domain, and 60,760 and 264,024 for the software vulnerability domain.

4.1 Evaluation Events and Metrics

Since each social environment has a unique set of events depending on the interactions in the platform, the evaluation events for each social environment were defined separately as the following: (I) GitHub: Commit Comment, Create, Delete, Fork, Issue Comment, Push, Watch, Pull Request, and Issues, (II) Twitter: Tweet (create original material), Retweet, Reply, and Quote, (III) Reddit: Comment and Post.

The evaluation measurements were applied to online social behaviors at the population, content, user, and community levels. Content-based measurements are any user interaction, that is, posting or replying in Reddit and Twitter platforms, or writing a comment in Github. User-level measurements were focused on user activities; for instance, the contribution counts of the user over time. Finally, the population-level measurements aggregated the events' and users' characteristics on a particular platform. Examples of the accuracy measurements and metrics include community burstiness and user Gini coefficient calculated by absolute percentage error, community Gini measured by absolute difference, user trustingness, user diffusion delay calculated by the Kolmogorov-Smirnov test (K-S test), and user popularity measured by Ranked-Biased Overlap (RBO). Additionally, "surprise measurements" were provided during the test event for competing campaigns in the cryptocurrency domain and competing attention in the cyber threats and CVE domains.

5 Experimental Results

We applied extensive analysis using our framework to mix and match and compare the models. We also calculated model performances for the user, community, population, and content levels for the cybersecurity, software vulnerability, and cryptocurrency domains. The training data input for the agent-based models were GitHub, Twitter, and Reddit network evolution over time for the three domains of interest, and the exogenous data for both training and test periods. Simulation inputs include the initialization data, the exogenous data, and the events data. There were two kinds of expected simulation outputs: (I) full network dynamics (event logs/data frames obeying specific formats that contain event-related information), and (II) direct output of the accuracy measurements.

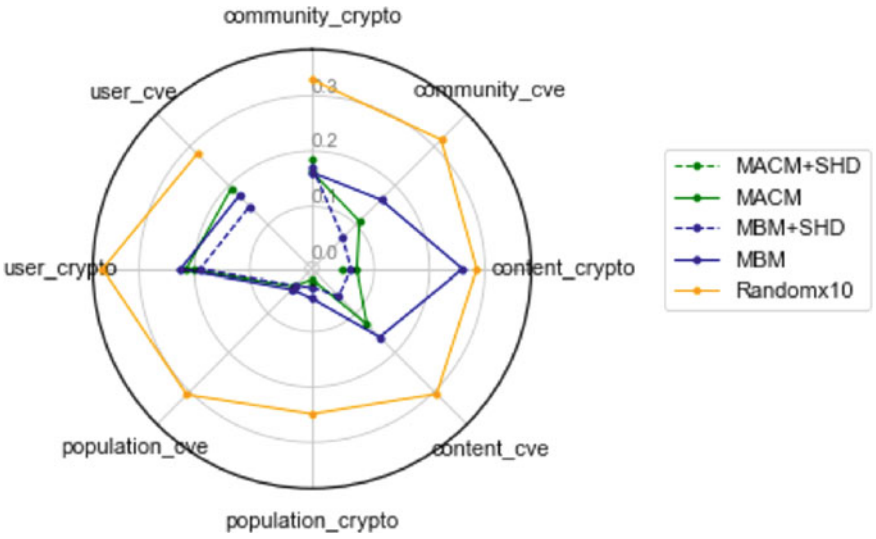


Fig. 2 The comparison of the normalized error ratio for the agent-based models and the ground truth based on the measurement types. A lower value corresponds to better model performance

The corresponding results for the benchmark comparisons are provided in Fig. 2, which represents the comparison of the model measurements versus the ground truth (Random) considering the measurement types. The Jensen-Shannon Divergence, Kolmogorov-Smirnov test, and absolute percentage error metrics were among the metrics used for this analysis. We normalized the model output results over the measurement group (user, community, population, and content), metric type, and the platform (GitHub, Twitter, Reddit) to calculate a value in the range 0 to 1, where the lower value corresponds to better model performance.

Figure 3a, b demonstrates the model-wise performance comparisons with and without the SHD strategy mixing, and for the two user influence pruning cases. The prefix "p" refers to the models initialized using the data pruned for the influential users, while the non-prefixed models relate to the models initialized employing the entire event data. We can observe that pruning influential relationships improved the MBM model performance; however, this strategy was not successful with the MACM model. The results indicate that social theory-based modeling may consider influential interactions inherently, and support that the more "explainable" a model, the higher the performance. Additionally, the results provide evidence that the mixing strategy helps improve the user and community level performances, and the single models are more successful in modeling the population and content level interactions of the influential users.

The metric-based performance comparison of the models is provided in Table 2 in the Appendix section, for the community, content, population, and user-level interactions, respectively. In these figures, the rows refer to the group-based measurement metrics for the models in rows. Each occupied cell indicates the best performance

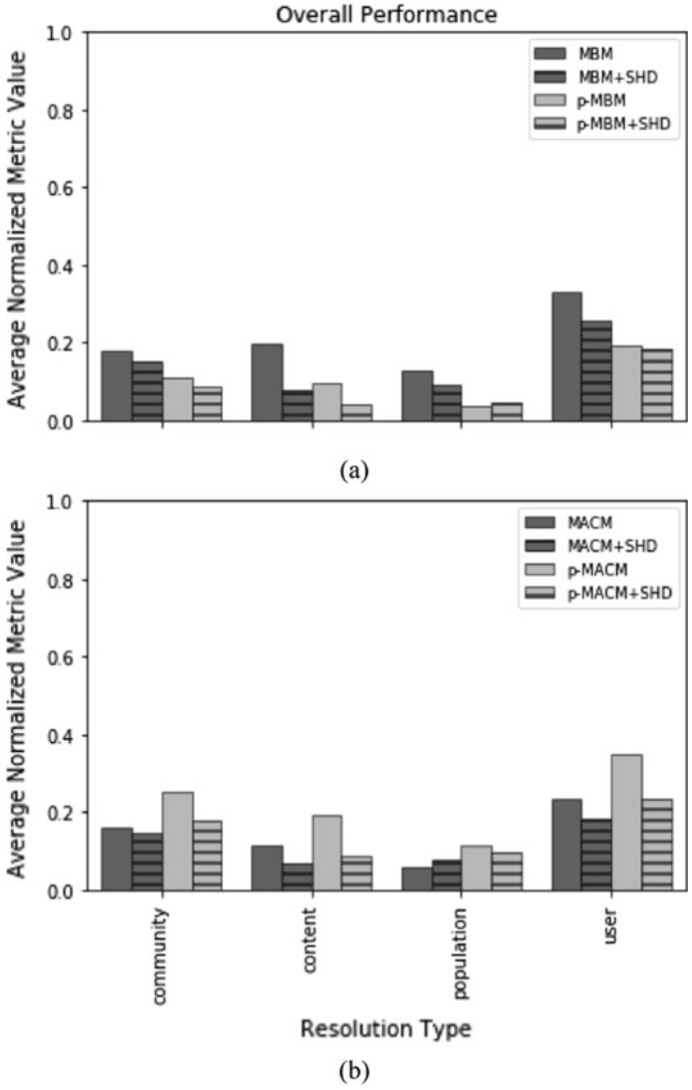


Fig. 3 Overall performance comparison over the **a** MBM and **b** MACM model variations over the community, content, population, and user levels, depicted in terms of normalized error ratio. A lower value corresponds to better model performance

Table 2 Performance comparison for the agent-based models across community, content, population, and user-level metrics

Category	Metric	MBM	MBM + SHD	pMACM	pMACM + SHD
Community	Burstiness	0.29	0.22	0.34	0.26
	Contributing users	0.31	0.24	0.35	0.26
	Event proportions	0.38	**0.20**	0.38	0.26
	Geo locations	0.22	0.23	0.33	**0.17**
	Gini coefficient	0.26	**0.18**	0.29	**0.20**
	Issue types	**0.00**	0.30	**0.00**	0.27
	User action counts	0.37	0.25	0.50	0.31
	Palma coefficient	0.55	0.31	0.53	0.40
	User account ages	**0.14**	**0.14**	0.54	**0.15**
	User burstiness	**0.18**	0.36	0.65	0.31
Content	Activity disparity Gini coefficient	**0.06**	**0.07**	**0.04**	**0.03**
	Activity disparity Palma coefficient	0.28	**0.12**	**0.09**	**0.16**
	Contributors	**0.06**	**0.10**	**0.11**	**0.00**
	Diffusion delay	0.22	**0.11**	0.63	**0.11**
	Event counts	0.29	**0.05**	**0.01**	**0.00**
	Daily event distribution	0.85	0.31	0.80	0.30
	Day of Week event distribution	0.78	0.22	0.75	0.20
	Growth	**0.07**	**0.12**	**0.16**	**0.01**
	Liveliness distribution	**0.15**	**0.06**	**0.15**	**0.01**
	Liveliness top K	0.90	0.52	0.86	0.38
	Popularity distribution top K	**0.10**	**0.03**	0.26	**0.06**
	Popularity top K	0.78	0.55	0.89	0.64
	User unique content	**0.06**	**0.07**	**0.10**	**0.05**
Population	Assortativity coefficient	**0.03**	**0.01**	**0.01**	**0.01**
	Average clustering coefficient	**0.01**	**0.02**	**0.01**	**0.01**
	Community modularity	**0.01**	**0.01**	**0.01**	**0.01**
	Degree distribution	0.81	0.68	0.34	0.54
	Density	**0.02**	**0.01**	**0.11**	**0.01**
	Max node degree	**0.03**	**0.02**	**0.04**	**0.03**
	Mean node degree	**0.01**	**0.01**	**0.02**	**0.01**
	Mean shortest path length	**0.01**	**0.01**	**0.00**	**0.01**
	Number of connected components	**0.06**	**0.10**	**0.01**	**0.02**
	Number of edges	**0.15**	**0.13**	**0.14**	**0.13**
	Number of nodes	**0.07**	**0.08**	**0.03**	**0.02**

(continued)

Table 2 (continued)

Category	Metric	MBM	MBM + SHD	pMACM	pMACM + SHD
User	Most active users	0.98	0.76	0.97	0.71
	Repository user continue proportion	**0.06**	**0.01**	**0.03**	**0.01**
	Subreddit user continue proportion	**0.02**	**0.01**	**0.00**	**0.00**
	Activity distribution	0.31	0.32	0.44	0.32
	Activity timeline	0.30	**0.18**	0.38	**0.12**
	Diffusion delay	0.47	**0.14**	0.35	0.46
	Gini Coefficient	**0.10**	**0.07**	**0.03**	**0.01**
	Palma Coefficient	0.26	0.26	0.29	0.30
	Popularity	0.82	0.69	0.85	0.67
	Trustingness	**0.00**	**0.08**	**0.00**	**0.08**
	User unique content	**0.05**	**0.07**	**0.19**	**0.07**

of the relative model for the specific metric. The values refer to the normalized sub-metrics averaged over 105 model runs. The content and user-level performances illustrate that the mixing strategy using the Standard Historical Data (SHD) improves the model performance in modeling the node level interactions. Finally, the population-level scores illustrate another benefit of the SHD mixing strategy in improving the performance for the degree distributions and the node level characteristics explanations.

6 Conclusion and Future Work

In this paper, we discussed how user interactions, behaviors, and complex human dynamics can be captured via combining massively parallel computing, data analytics of large datasets, and machine learning algorithms. We proposed the Deep Agent Framework (DAF), which operates beyond single models by mixing and matching sub-models in a semi-automated way. Our framework operationalizes social theories of human behavior and social media into optimized generative simulation capabilities that enable exploring information diffusion and evolution within the social media context. Our multi-resolution simulation at the user, community, population, and content levels, and our extensive analysis and results provide evidence that our framework is a powerful tool in modeling the diffusion and evolution of the information in variety of online social platforms. Although we applied hand-crafted behavioral theories and data-driven models directly to feed the agent-based models without employing the genetic program (red box in Fig. 1a), our results prove that our framework and our deep generative models are powerful in modeling online social network interactions.

Further improvements to the DAF framework can be made by employing the evolutionary model discovery to explore the space for the rule sets of behaviors related to the agents, which allows for the testing and validation of tens of thousands of models against large set of target behaviors as in [6]. Additionally, the future direction of our work serves to automatically introduce variants to all the models produced by different performers to obtain the overall best model.

Acknowledgements This work was supported by the Defense Advanced Research Projects Agency (DARPA) under grant number FA8650-18-C-7823. The views and opinions expressed in this article are the authors' own and should not be construed as official or as reflecting the views of the University of Central Florida, DARPA, or the U.S. Department of Defense.

Appendix

This section provides a list of measurement metrics across community, content, population, and user levels in Table 2. Each occupied cell indicates the relative performance of the model in terms of error (in columns) reported for a specified metric (in rows). Values in the cells refer to the error ratio assessed via first normalizing over each sub-metric, followed by averaging over the results for the model runs. In this table, a lower value corresponds to better model performance and the error ratios below 0.2 are marked in bold.

References

1. Adamic, L.A., Lento, T.M., Adar, E., Ng, P.C.: Information evolution in social networks. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 473–482. ACM (2016)
2. Rand, W., Herrmann, J., Schein, B., Vodopivec, N.: An agent-based model of urgent diffusion in social media. J. Artif. Soc. Soc. Simul. **18**(2), 1 (2015)
3. Milovan Šuvakov, D.G., Schweitzer, F., Tadić, B.: Agent-based simulations of emotion spreading in online social networks. *arXiv preprint* arXiv:1205.6278 (2012)
4. Epstein, J.M.: Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science, vol. 25. Princeton University Press (2014)
5. Gintis, H., Helbing, D., et al.: Homo socialis: an analytical core for sociological theory. Rev. Behav. Econ. **2**(1–2), 1–59 (2015)
6. Chathika, G., Ivan, G.: Alternate social theory discovery using genetic programming: towards better understanding the artificial Anasazi. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 115–122. ACM (2017)
7. Bass, F.M.: A new product growth for model consumer durables. Manage. Sci. **50**(12_supplement), 1825–1832 (2004)
8. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociol. **83**(6), 1420–1443 (1978)
9. Watts, D.J.: A simple model of global cascades on random networks. Proc. Natl. Acad. Sci. **99**(9), 5766–5771 (2002)
10. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark. Lett. **12**(3), 211–223 (2001)

11. Goldenberg, J., Libai, B., Muller, E.: Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata. Acad. Mark. Sci. Rev. **9**(3), 1–18 (2001)
12. Bass, F.M.: A new product growth for model consumer durables. Manag. Sci. **15**(5), 215–227 (1969)
13. Rand, W., Rust, R.T.: Agent-based modeling in marketing: guidelines for rigor. Int. J. Res. Market. **28**(3), 181–193 (2011)
14. Centola, D., Macy, M.: Complex contagions and the weakness of long ties. Am. J. Sociol. **113**(3), 702–734 (2007)
15. Basu, P., Dippel, M., Sundaram, R.: Multiplex networks: a generative model and algorithmic complexity. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 456–463. IEEE (2015)
16. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Ann. Rev. Sociol. **27**(1), 415–444 (2001)
17. Gunaratne, C., Senevirathna, C., Jayalath, C., Baral, N., Rand, W., Garibay, I.: A multi-action cascade model of conversation. In: 5th International Conference on Computational Social Science. http://app.ic2s2.org/app/sessions/9kXqn5btgKKC5yfCvg/details (2019)
18. Gunaratne, C., Baral, N., Rand, W., Garibay, I., Jayalath, C., Senevirathna, C.: The effects of information overload on online conversation dynamics. Comput. Math. Organ. Theory (2020)
19. Bidoki, N.H., Schiappa, M., Sukthankar, G., Garibay, I.: Predicting social network evolution from community data partitions. In: 2019 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (2019) (In–press)
20. Saadat, S., Gunaratne, C., Baral, N., Sukthankar, G., Garibay, I.: Initializing agent-based models with clustering archetypes. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp 233–239. Springer (2018)

Electoral College: Emergent Battlegrounds An Agent-Based Model of Campaign Behavior Change with District Allocation of Electors

John Silver

Abstract The model presented in this paper explores whether reforming the Electoral College to a district model would have an impact on campaign activity in so-called "battleground" states. Using an agent-based model to estimate election behavior of both voters and campaigns, results are analyzed for net change in frequency of states being targeted based on expected gains from very slight perturbation of election results in battleground states where margin of winning between parties is very small. Using data from the 2016 United States presidential election as a baseline and allowing for the configuration of empirical voter demographics and competitiveness thresholds for targeted campaign activity, the model explores the possible net change in battleground priorities, with the results indicating familiar state targets of focus such as Iowa, New Hampshire and Ohio, may lose prominence to larger traditionally "safe" states such as California, Texas, and Virginia where large numbers of electoral votes previously unavailable to one party or the other would come into play.

Keywords Agent-based model · Electoral College · Voting · Campaign · Election

1 Introduction

That there is inherent bias in the Electoral College has been evident since it was conceived by Alexander Hamilton in 1788. It was never intended to be directly representative of aggregate popular will [4], on the contrary, it was intended to subvert it, directly, deliberately, and with the expressed reason to ensure only qualified individuals were elevated to the office. Nonetheless, its efficacy and fairness have been in question ever since. This paper and attendant model seek to explore the consequences of the most likely reform to the Electoral College—on account of it already

J. Silver (✉)
George Mason University, 4400 University Drive, Fairfax, VA, USA
e-mail: jsilver9@gmu.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_12

being in place in two states—the proportional assignment of electors as it impacts campaign activity to sway narrow-margin "battleground" states and districts.

2 Background

Langley and Dana [6] have identified the various biases in the Electoral College, namely, the lack of unit apportionment that is the mixing of a fixed two electors combined with a variable number apportioned by population—equivalent to seats in the House of Representatives plus the seats in the Senate. These manifest as more power being given to more populous states, but more effective voting power being given to residents of less populous states, owing to the minimum 2 + 1 electors. Turner [7] presents the case of Nebraska and Maine as evidence of the efficacy of a district model of reform. In such a model, the popular vote winner in each state is assigned two electors, and the winner of each congressional district is assigned one. This does not remove the bias of the Electoral College completely, but it approaches proportional allocation. There are no constitutional barriers, only the inertia of political will in the states is required.

Turner identifies seven qualities that drive this solution. It should of course improve on reflecting the public will and increase interest in voting, but it also alters the incentives and feedback in gathering support for "down-ticket" races, changes the electoral favor of certain groups—not least geographically, which is the focus of this paper—as well as the opportunities for third parties, exploitation, and fraud. Lastly, it must be reasonably likely to be adopted. This model attempts to analyze only the change in electoral favor as engaged by the two major parties. It is the expectation that more states will be engaged, that they will be more equally engaged, and that much of the movement will be from smaller states to larger population centers generally considered "safe."

That campaigns seek to influence elections by targeting battleground states is well known. Hill et al. [5] analyze this activity and determine that despite minimal effects, the "battleground" state activity is still sought with vigor as the payoffs are large for only a slight movement in the electoral outcome. This may be positive, seeking to enhance the candidate's position, or negative, seeking to degrade the position of an opponent.

This is the behavior being modeled. Based on identifying low-margin districts and states, campaigns direct activity that in net has either a positive or deleterious effect, simultaneously garnering advantage for one's self and exacting damage on one's opponent.

3 Methodology

An agent-based model is implemented in NetLogo 6.1 [8] with a statistical export for offline analysis. Empirical data is loaded using the 2016 statewide presidential results from the DailyKos [2] congressional district-level results mapping (Washington, DC results are missing in that dataset, but have been added from the District of Columbia Bureau of Elections [3]), with voters instantiated at 1/10000 scale and party affinity assigned proportional and probabilistically to the recorded vote totals. While this dataset necessarily includes interpolation of precinct-level returns aggregated at the county level and subsequently overlaid into the corresponding congressional districts, for this purpose it is sufficiently precise. As these data represent active voters, we assume our entire population is thus eligible and by default intent on voting. Third-party results are consolidated as independent voters. District-level population density data were acquired from ArcGIS [1].

A total of 226 runs were performed for each of traditional Electoral College tallying and district allocation of electors at a battleground margin threshold of 5%. This was chosen purely at author's discretion as a reasonable median—10% or more is reasonably "safe" by practical definition, and <1% is not analytically useful. The model allows this to be set arbitrarily. For these configurations, the output of the model documents the frequency of states and districts appearing as "battlegrounds" where the margin between the two primary parties is less than the configured threshold. Finally, the differences between the traditional Electoral College frequencies and the proportional assignment were compared, presenting a rank-order of net position changes by state. These results were then aggregated into total frequency and per-district spend by state.

The expectation is that familiar "battlegrounds" will be swapped for equally familiar non-battlegrounds in recent memory. The reasoning for this expectation is that large "safe" states will now have potential for securing electors without the need to carry the entire state in order to gain advantage. Proportional assignment of electors makes it possible to seek smaller local gains (Fig. 1).

4 The Model

The interface is divided into three sections, one for setup and configuration, the agent view, and various instrumentation. On setup, the agents are initialized and assigned a probabilistic party affinity based on the 2016 election results for Republican, Democratic, and independent parties.

If "continuous" updates are selected campaign target selection can be seen as districts flash their borders to indicate activity. With "on tick" updates, this behavior is masked, though it is still occurring in the background.

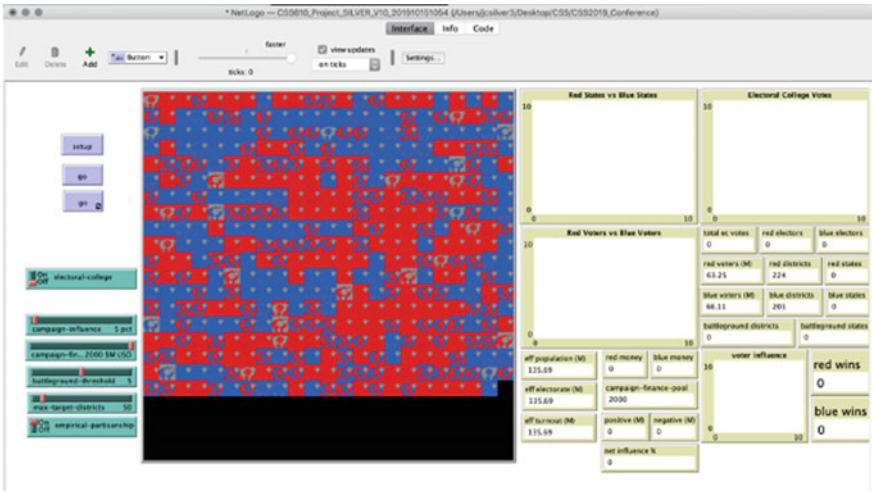


Fig. 1 Model immediately after setup initialization showing 538 available electors and 2016 partisan affinity

For the analytical purposes of this paper, it is the "battleground districts" and "battleground states" values that are of interest, though other behaviors can be observed. For instance, as described by Turner [7], when switching-off the Electoral College toggle, the resulting (now proportional) votes become much narrower in variance. Switching back to Electoral College causes wild swings in apparent win/loss margins to appear. This is the perceived "mandate" Turner describes.

The "Voter Influence" graph displays the net position of all campaign activity. Where campaigns operate alone, they will have nothing but positive effects. Where they come in contact, the lesser-funded candidate will suffer damage. Depending on how budgets have been allotted, this may be fairly severe punishment that will appear as a net negative aggregate influence (Fig. 2).

5 Validation

In order to calibrate the model to ensure there are not programmatic artifacts, disabling empirical partisanship results in an even probability between two parties. Running 100 elections, we get red/blue performance with statistically insignificant differences, as illustrated in the STATA codebook and box plot outputs below.

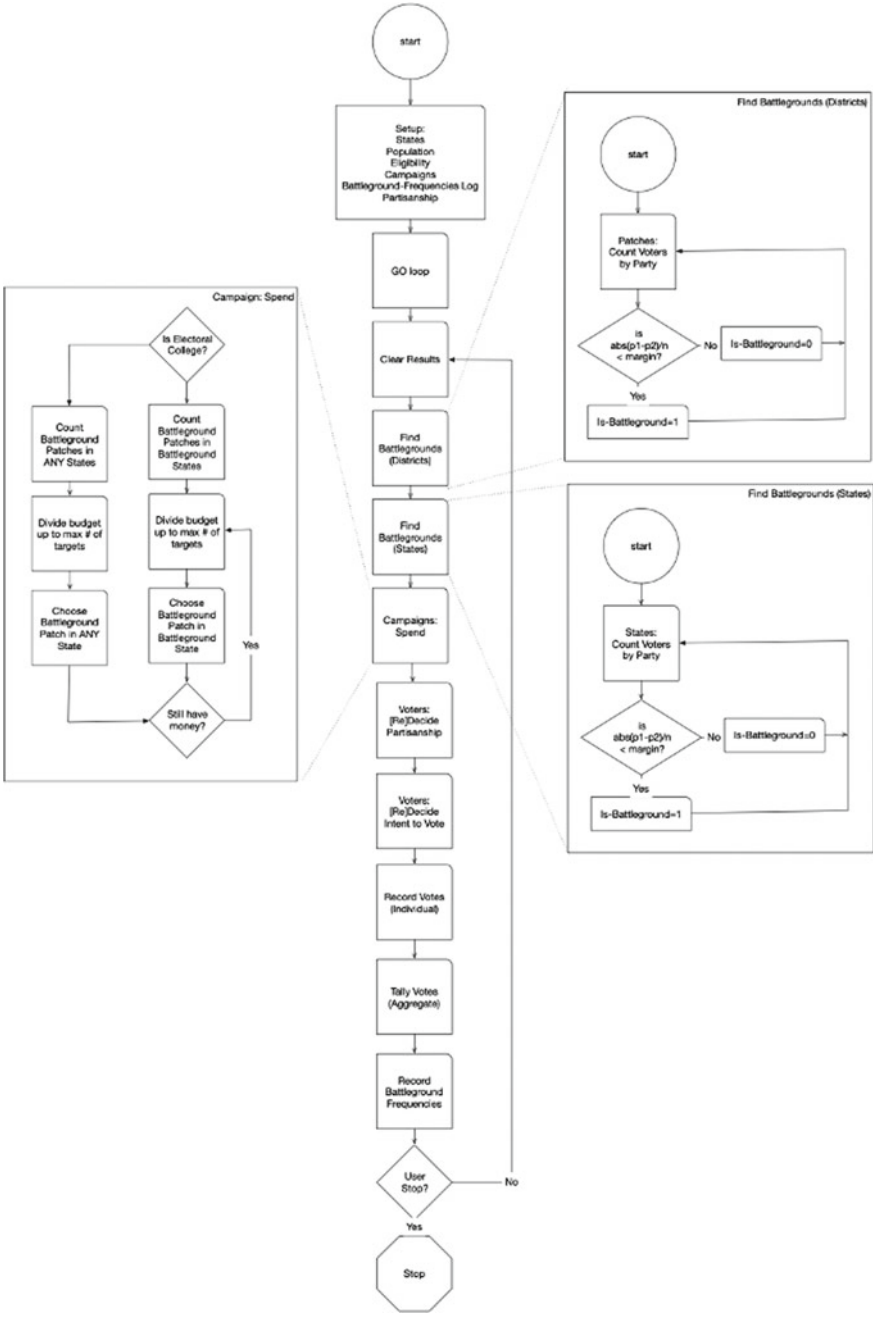


Fig. 2 Model program flow

```
bluewins
                  type:  numeric (byte)
                 range:  [38,61]   units:  1
         unique values:  24    missing .:  0/100
                  mean:     48.63
             std. dev:   5.21973

percentiles:
10%       25%       50%       75%       90%
41        45        49        51        56

redwins
                  type:  numeric (byte)
                 range:  [37,60]   units:  1
         unique values:  23    missing .:  0/100
                  mean:     48.95
             std. dev:     5.058

percentiles:
10%       25%       50%       75%       90%
42        46        49      52.5        56
```

Using the empirical population under a district model we get the following results. Notable here is the narrow range and standard deviation, indicating the district model may indeed meet Turner's first criterion to provide more accurate results in reflecting public will (Figs. 3 and 4).

```
bluewins
                  type:  numeric (byte)
                 range:  [4,18]    units:  1
         unique values:  14    missing .:  0/100
                  mean:      9.92
             std. dev:    2.9531

percentiles:
10%       25%       50%       75%       90%
6.5         8        10        12      14.5

redwins
                  type:  numeric (byte)
                 range:  [79,95]   units:  1
         unique values:  16    missing .:  0/100
                  mean:     87.97
             std. dev:   3.10573
```

```
percentiles:
10%       25%       50%       75%       90%
84        86        89        90        92
```

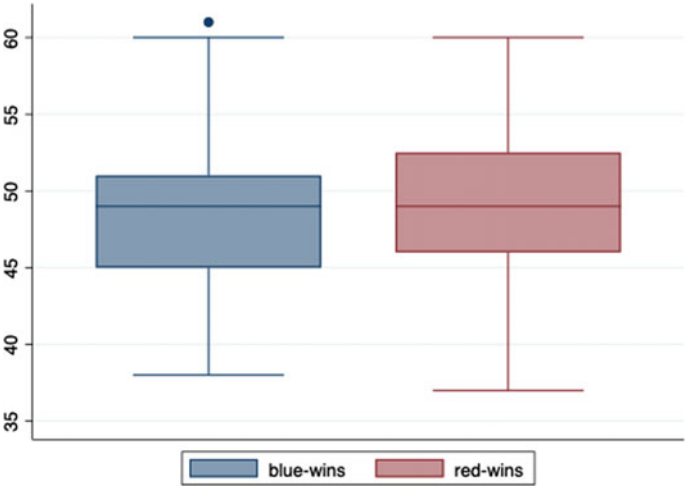


Fig. 3 Box plot of baseline 5050 probability for district model, N = 100. Differences between bluewins and redwins are attributed to randomness but are not statistically significant, as expected

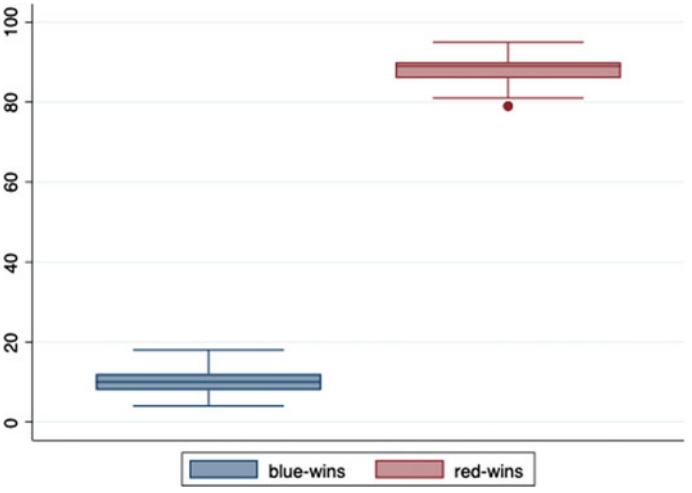


Fig. 4 Box plot of 2016 probability for district model, N = 100. Differences between bluewins and redwins roughly match probabilistic expectations from the 2016 results

The winner-take-all Electoral College model is similarly validated at 50/50 affinity (Fig. 5):

```
bluewins
                  type:  numeric (byte)
                 range:  [37,63]   units:  1
         unique values:  25    missing .:  0/100
                  mean:      49.7
             std. dev:   5.45227

percentiles:
10%       25%       50%       75%       90%
42.5      46        49        53.5      57

redwins
                  type:  numeric (byte)
                 range:  [36,62]   units:  1
         unique values:  26    missing .:  0/100
                  mean:      49.6
             std. dev:   5.43464

percentiles:
10%       25%       50%       75%       90%
42        46        50        53        57.5
```

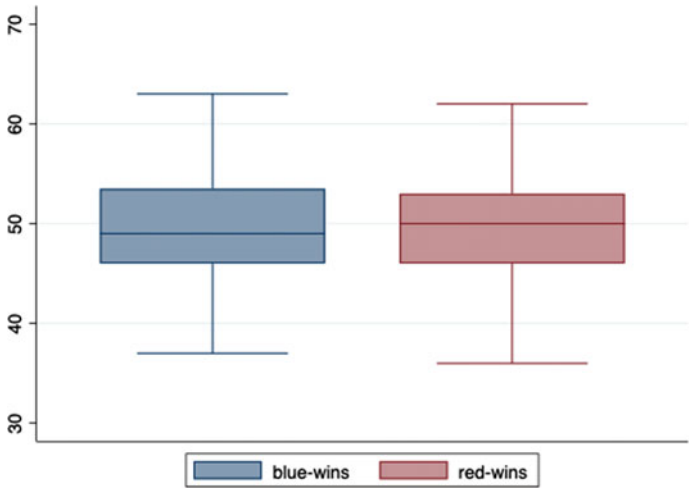


Fig. 5 Box plot of baseline 50/50 probability for winner-take-all model, N = 100. Differences between bluewins and redwins are attributed to randomness but are not statistically significant, as expected

The empirical distribution, winner-take-all results, note the larger range and standard deviation compared to the district model (Fig. 6):

```
bluewins
                  type:  numeric (byte)
                 range:  [19,41]   units:  1
         unique values:  20    missing .:  0/100
                  mean:     30.18
             std. dev:    4.38173

percentiles:
10%       25%       50%       75%       90%
25        27        30      33.5        36

redwins
                  type:  numeric (byte)
                 range:  [59,80]   units:  1
         unique values:  19    missing .:  0/100
                  mean:     68.42
             std. dev:    4.27863

percentiles:
10%       25%       50%       75%       90%
62.5      65        69        71        74
```

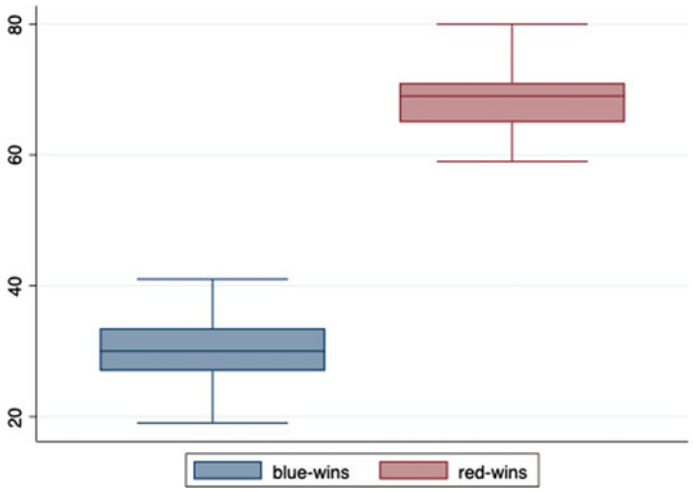


Fig. 6 Box plot of 2016 probability for winner-take-all model, N = 100. Range and standard deviation increased from district model

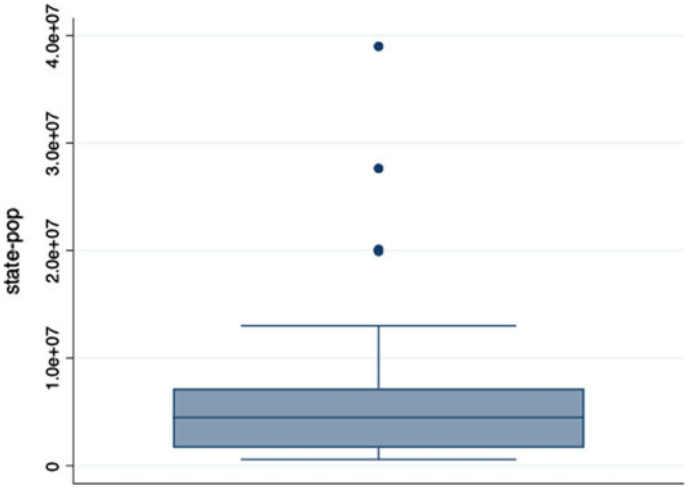


Fig. 7 Box plot output for 2016 state population distribution. All states, outliers are California, Texas, and Florida

To assess whether state size is a factor in any differences, we establish the baseline state population distribution, with three noted outliers California, Texas, and Florida (Fig. 7):

```
Statepop overall:
statepop
                  type:  numeric (long)
                 range:  [599928,38986172] units:  1
         unique values:  51    missing .:  0/51
                  mean:   6.3e+06
             std. dev:   7.2e+06

percentiles:
10%       25%       50%       75%       90%
872884    1.7e+06   4.5e+06   7.2e+06   1.3e+07
```

6 Results

The following results are from the state-level aggregate dataset (226 runs each winner-take-all and district model). At the highest level, we do notice a change in the state population distribution between the winner-take-all Electoral College model and the district model (Figs. 8 and 9).

```
codebook statepop if spendec > 0
statepop
                  type:  numeric (long)
                 range:  [644954,27637153] units:  1
         unique values:  42    missing .:  0/42
                  mean:   6.1e+06
             std. dev:   5.8e+06

percentiles:
10%       25%       50%       75%       90%
1.0e+06   1.9e+06   4.6e+06   7.2e+06   1.3e+07


codebook statepop if spenddm > 0
statepop
                  type:  numeric (long)
                 range:  [644954,38986172] units:  1
         unique values:  47    missing .:  0/47
                  mean:   6.8e+06
             std. dev:   7.3e+06

percentiles:
10%       25%       50%       75%       90%
1.1e+06   2.1e+06   4.9e+06   8.5e+06   1.3e+07
```

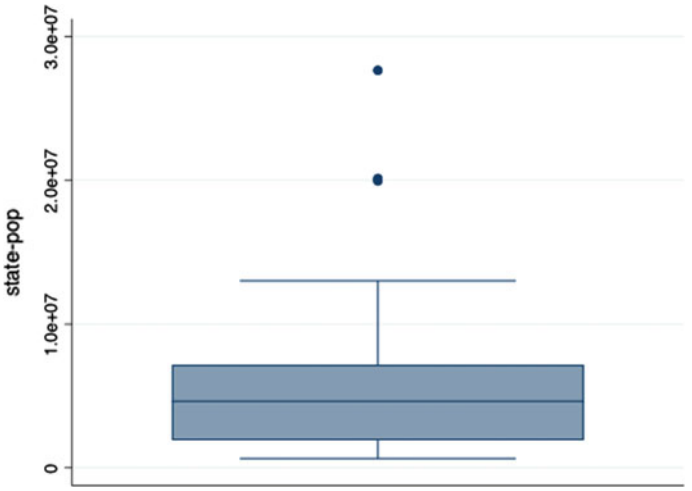


Fig. 8 Box plot output for 2016 state population distribution, only states receiving campaign spending under winner-take-all rules

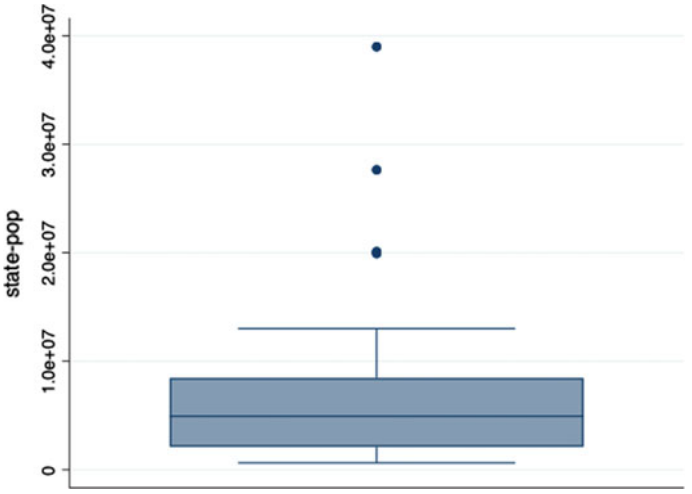


Fig. 9 Box plot output for 2016 state population distribution, only states receiving campaign spending under district model rules

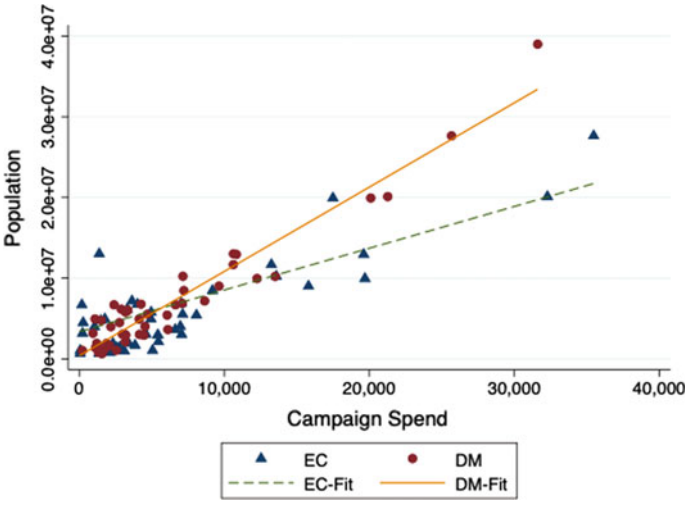


Fig. 10 Two-way scatter of state population to simulated campaign spend under winner-take-all Electoral College rules and District Model rules

While the district model is much more widely spread (std. dev 7.3e+06 vs. 5.8e+06) and we see greater outliers at the extreme end of the population of California, there is not a significant difference in means. So, while we can observe there is a tendency of lowered spending in smaller states and higher spending in larger states under district rules than winner-take-all, at this time it does not appear to be significant—most of the difference being attributable to gains by California (Fig. 10).

In terms of net per-district spending change between the winner-take-all Electoral College system and the district model, the movement is only approximately ±3%. Movements are generally as expected, with gains in California, Washington and Illinois, losses in Delaware, New Hampshire, Rhode Island, New Mexico, Idaho and Utah, and no change at all in Wyoming, South Dakota, and Alaska (Figs. 11 and 12).

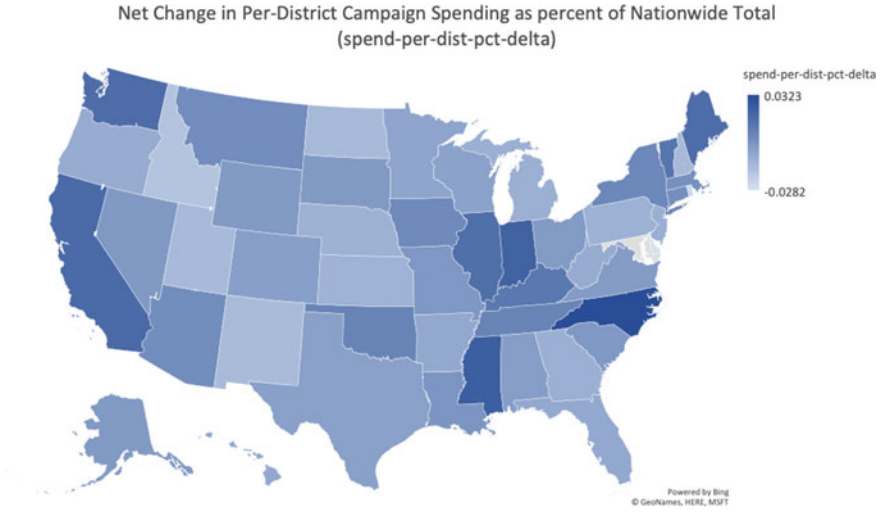


Fig. 11 Mapped results of net spending change per district as percent of all spending

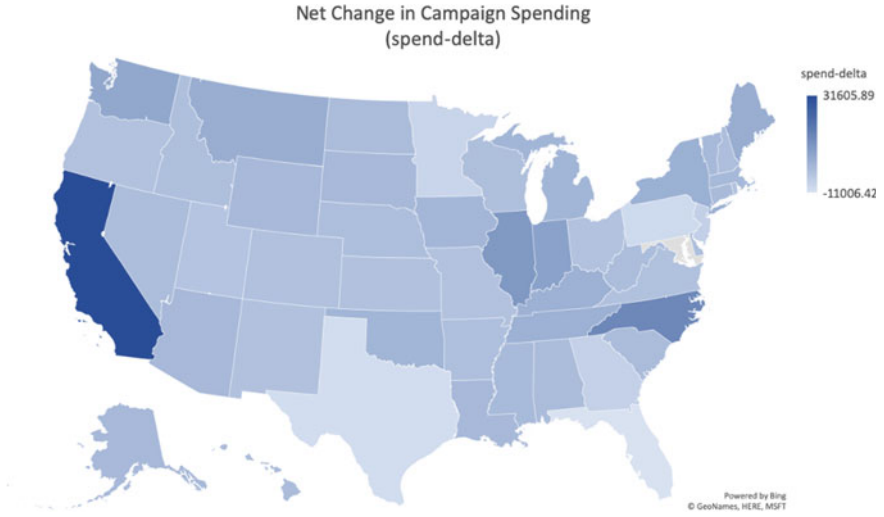


Fig. 12 Mapped results of net spending change per state

More striking is the overwhelming shift of campaign dollars to California and North Carolina, with losses in traditional battlegrounds of Florida and Pennsylvania.

7 Discussion and Conclusions

This model is abstract and the decision structure of the campaign behavior under investigation relatively simplistic, but this is not unrealistic. The results are superficially within broad expectations, but that may reflect that the behavior being modeled really is predictable as described and operating on the same or similar parameters. In that sense, organized political behavior, especially at the national level, is a fairly top-down business driven by statistical observation, so it is plausible that while the input parameters and decision trees are far more complex than presented here, they probably are not much more complicated. The addition of more varied district demographics and a more robust analysis within battleground choice may suffice to adequately model the campaign behavior and pass more than face validation. However, with the addition of refined logic and greater heterogeneity, I would not expect the fundamental patterns observed in this model to change. The precise members of the output may be ordered differently, but the nearly zero-sum game of changing battleground campaign targets would likely be expected to be similar.

Appendix 1: Results—Electoral College

	State	Population	Districts	Spend	Per District	Percent
1	AK	752680	1	0.00	0.00	0.000
2	AL	4958020	7	1766.72	252.39	0.006
3	AR	3050229	4	4598.53	1149.63	0.028
4	AZ	6877018	9	7034.10	781.57	0.019
5	CA	38986172	53	0.00	0.00	0.000
6	CO	5425480	7	8081.75	1154.54	0.028
7	CT	3641077	5	6622.72	1324.54	0.032
8	DC	659110	1	66.67	66.67	0.002
9	DE	953628	1	3115.67	3115.67	0.075
10	FL	20108441	27	32276.38	1195.42	0.029
11	GA	10237835	14	13587.31	970.52	0.023
12	HI	1435363	2	2784.93	1392.46	0.034
13	IA	3173649	4	228.76	57.19	0.001
14	ID	1686061	2	3848.39	1924.20	0.046
15	IL	13022489	18	1382.85	76.83	0.002

	State	Population	Districts	Spend	Per District	Percent
16	IN	6698147	9	182.16	20.24	0.000
17	KS	2960265	4	5414.78	1353.70	0.033
18	KY	4497004	6	268.60	44.77	0.001
19	LA	4766833	6	1495.99	249.33	0.006
20	MA	6780933	9	4011.56	445.73	0.011
21	MD	6038124	8	0.00	0.00	0.000
22	ME	1367696	2	0.00	0.00	0.000
23	MI	9954630	14	19696.15	1406.87	0.034
24	MN	5529373	8	7130.91	891.36	0.021
25	MO	6158939	8	3308.32	413.54	0.010
26	MS	3057449	4	0.00	0.00	0.000
27	MT	1047855	1	66.67	66.67	0.002
28	NC	10220673	13	0.00	0.00	0.000
29	ND	776579	1	2123.40	2123.40	0.051
30	NE	1917345	3	3403.03	1134.34	0.027

	State	Population	Districts	Spend	Per District	Percent
31	NH	1360762	2	2613.33	1306.67	0.031
32	NJ	9008703	12	15806.45	1317.20	0.032
33	NM	2127466	3	5457.75	1819.25	0.044
34	NV	2924856	4	5427.12	1356.78	0.033
35	NY	19934508	27	17488.57	647.72	0.016
36	OH	11685719	16	13249.60	828.10	0.020
37	OK	3990593	5	1054.67	210.93	0.005
38	OR	4029969	5	6972.86	1394.57	0.034
39	PA	12941864	18	19612.63	1089.59	0.026
40	RI	1060080	2	5056.39	2528.20	0.061

	State	Population	Districts	Spend	Per District	Percent
41	SC	4943188	7	4942.27	706.04	0.017
42	SD	872884	1	0.00	0.00	0.000
43	TN	6698360	9	0.00	0.00	0.000
44	TX	27637153	36	35461.17	985.03	0.024
45	UT	3029950	4	7062.58	1765.65	0.043
46	VA	8462750	11	9182.08	834.73	0.020
47	VT	644954	1	1360.05	1360.05	0.033
48	WA	7177995	10	3639.19	363.92	0.009
49	WI	5795428	8	4954.11	619.26	0.015
50	WV	1914436	3	2330.08	776.69	0.019
51	WY	599928	1	0.00	0.00	0.000

Appendix 2: Results—District Model

	State	Population	Districts	Spend	Per District	Percent
1	AK	752680	1	0.00	0.00	0.000
2	AL	4958020	7	1093.41	156.20	0.005
3	AR	3050229	4	2984.12	746.03	0.022
4	AZ	6877018	9	7102.75	789.19	0.023
5	CA	38986172	53	31605.89	596.34	0.017
6	CO	5425480	7	6061.27	865.90	0.025
7	CT	3641077	5	6112.65	1222.53	0.036
8	DC	659110	1	0.00	0.00	0.000
9	DE	953628	1	1639.74	1639.74	0.048
10	FL	20108441	27	21269.96	787.78	0.023
11	GA	10237835	14	7149.90	510.71	0.015
12	HI	1435363	2	1879.17	939.59	0.027
13	IA	3173649	4	948.87	237.22	0.007
14	ID	1686061	2	2081.74	1040.87	0.030
15	IL	13022489	18	10627.38	590.41	0.017

	State	Population	Districts	Spend	Per District	Percent
16	IN	6698147	9	6640.15	737.79	0.021
17	KS	2960265	4	3204.99	801.25	0.023
18	KY	4497004	6	2777.68	462.95	0.013
19	LA	4766833	6	1527.72	254.62	0.007
20	MA	6780933	9	4254.52	472.72	0.014
21	MD	6038124	8	3358.34	419.79	0.012
22	ME	1367696	2	1124.58	562.29	0.016
23	MI	9954630	14	12260.36	875.74	0.025
24	MN	5529373	8	4693.39	586.67	0.017
25	MO	6158939	8	2902.91	362.86	0.011
26	MS	3057449	4	3123.90	780.97	0.023
27	MT	1047855	1	210.93	210.93	0.006
28	NC	10220673	13	13510.36	1039.26	0.030
29	ND	776579	1	1339.78	1339.78	0.039
30	NE	1917345	3	1881.22	627.07	0.018

	State	Population	Districts	Spend	Per District	Percent
31	NH	1360762	2	1329.75	664.87	0.019
32	NJ	9008703	12	9634.53	802.88	0.023
33	NM	2127466	3	3173.07	1057.69	0.031
34	NV	2924856	4	4457.37	1114.34	0.032
35	NY	19934508	27	20099.57	744.43	0.022
36	OH	11685719	16	10623.51	663.97	0.019
37	OK	3990593	5	2182.46	436.49	0.013
38	OR	4029969	5	4538.39	907.68	0.026
39	PA	12941864	18	10829.01	601.61	0.017
40	RI	1060080	2	2557.92	1278.96	0.037

	State	Population	Districts	Spend	Per District	Percent
41	SC	4943188	7	4131.48	590.21	0.017
42	SD	872884	1	0.00	0.00	0.000
43	TN	6698360	9	2402.24	266.92	0.008
44	TX	27637153	36	25667.61	712.99	0.021
45	UT	3029950	4	4149.75	1037.44	0.030
46	VA	8462750	11	7219.24	656.29	0.019
47	VT	644954	1	1566.37	1566.37	0.045
48	WA	7177995	10	8645.50	864.55	0.025
49	WI	5795428	8	3199.05	399.88	0.012
50	WV	1914436	3	1204.00	401.33	0.012
51	WY	599928	1	0.00	0.00	0.000

Appendix 3: Results—Delta

	State	Population	Districts	Spend	Per District	Percent
1	AK	752680	1	0.00	0.00	0.000
2	AL	4958020	7	−673.31	−96.19	−0.002
3	AR	3050229	4	−1614.42	−403.60	−0.006
4	AZ	6877018	9	68.65	7.62	0.004
5	CA	38986172	53	31605.89	596.34	0.017
6	CO	5425480	7	−2020.48	−288.64	−0.003
7	CT	3641077	5	−510.07	−102.01	0.004
8	DC	659110	1	−66.67	−66.67	−0.002
9	DE	953628	1	−1475.93	−1475.93	−0.027
10	FL	20108441	27	−11006.42	−407.64	−0.006
11	GA	10237835	14	−6437.41	−459.81	−0.009
12	HI	1435363	2	−905.75	−452.87	−0.006
13	IA	3173649	4	720.1	1180.03	0.006
14	ID	1686061	2	−1766.65	−883.33	−0.016
15	IL	13022489	18	9244.52	513.58	0.015

	State	Population	Districts	Spend	Per District	Percent
16	IN	6698147	9	6457.98	717.55	0.021
17	KS	2960265	4	−2209.80	−552.45	−0.009
18	KY	4497004	6	2509.09	418.18	0.012
19	LA	4766833	6	31.72	5.29	0.001
20	MA	6780933	9	242.97	26.99	0.003
21	MD	6038124	8	3358.34	419.79	0.012
22	ME	1367696	2	1124.58	562.29	0.016
23	MI	9954630	14	−7435.79	−531.13	−0.008
24	MN	5529373	8	−2437.53	−304.69	−0.004
25	MO	6158939	8	−405.41	−50.68	0.001
26	MS	3057449	4	3123.90	780.97	0.023
27	MT	1047855	1	144.27	144.26	0.005
28	NC	10220673	13	13510.36	1039.26	0.030
29	ND	776579	1	−783.61	−783.62	−0.012
30	NE	1917345	3	−1521.81	507.27	−0.009

	State	Population	Districts	Spend	Per District	Percent
31	NH	1360762	2	1329.75	664.87	0.019
32	NJ	9008703	12	9634.53	802.88	0.023
33	NM	2127466	3	3173.07	1057.69	0.031
34	NV	2924856	4	4457.37	1114.34	0.032
35	NY	19934508	27	20099.57	744.43	0.022
36	OH	11685719	16	10623.51	663.97	0.019
37	OK	3990593	5	2182.46	436.49	0.013
38	OR	4029969	5	4538.39	907.68	0.026
39	PA	12941864	18	10829.01	601.61	0.017
40	RI	1060080	2	2557.92	1278.96	0.037

	State	Population	Districts	Spend	Per District	Percent
41	SC	4943188	7	4131.48	590.21	0.017
42	SD	872884	1	0.00	0.00	0.000
43	TN	6698360	9	2402.24	266.92	0.008
44	TX	27637153	36	25667.61	712.99	0.021
45	UT	3029950	4	4149.75	1037.44	0.030
46	VA	8462750	11	7219.24	656.29	0.019
47	VT	644954	1	1566.37	1566.37	0.045
48	WA	7177995	10	8645.50	864.55	0.025
49	WI	5795428	8	3199.05	399.88	0.012
50	WV	1914436	3	1204.00	401.33	0.012
51	WY	599928	1	0.00	0.00	0.000

References

1. ArcGIS: 2016 Population Density by Congressional District (n.d.). http://hub.arcgis.com/datasets/esri::2016-population-density-by-congressional-district/data. Cited 13 Oct 2019
2. DailyKos: Daily Kos Elections' presidential results by congressional district for 2016, 2012, and 2008 (n.d.). http://www.dailykos.com/stories/2012/11/19/1163009/. Cited 14 Oct 2019
3. District of Columbia Board of Elections: General Election 2016—certified results (2018). http://electionresults.dcboe.org/election_results/2016-General-Election. Cited 26 Oct 2019
4. Hamilton, A., Madison, J., Jay, J., Dunn, J., Horowitz, D., Botting, E.: The Federalist Papers. Yale University Press, New Haven, CT (2009). http://www.jstor.org/stable/j.ctt5vm398.71. Cited 26 Oct 2019
5. Hill, J., Rodriquez, E., Wooden, A.: Stump speeches and road trips: the impact of state campaign appearances in Presidential Elections. PS: Polit. Sci. Polit. **43**(2), 243–254 (2010). https://doi.org/10.1017/S1049096510000077
6. Longley, L., Dana, J.: New empirical estimates of the biases of the Electoral College for the 1980s. West. Polit. Q. **37**(1), 157–175 (1984). https://doi.org/10.2307/448331
7. Turner, R.: The contemporary presidency: do Nebraska and Maine have the right idea? The political and partisan implications of the district system. Pres. Stud. Q. **35**(1), 116–137 (2005). https://doi.org/10.1111/j.1741-5705.2004.00238.x
8. Wilensky, U.: NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL (1999). http://ccl.northwestern.edu/netlogo/. Cited 26 Oct 2019

Social Viscosity, Fluidity, and Turbulence in Collective Perceptions of Color: An Agent-Based Model of Color Scale Convergence

Juan Salamanca and Santiago Núñez-Corrales

Abstract Social flow, viscosity, and turbulence increasingly help explain observations of collective social systems in which self-organization is driven by norms and beliefs. We propose a simple agent-based model of self-organization of human agents, adapted from seminal color sorting experiments of individual perceptions of color proximity, as representative of a fundamental class of social phenomena involving convergence toward a stable collective social structure. We define *inverse social viscosity* as the measure of the difference between agent situated beliefs and perceptions as the driver of collective action flow. We study convergence and reversions using a particular form of the equation describing nucleation processes in phase transition theory. Our analysis suggests that tolerance to imperfect compliance with norms and a degree of tolerance with own beliefs decrease coordination efforts. In addition, our research suggests that social viscosity is a proxy measure for the cost of social organization, which can in turn be used to inform the design of socio-technical systems. Breaking social isolation is a successful strategy to foster self-organization.

1 Introduction: Social Viscosity, Fluids, and Turbulence

Fluid spatiality is one of the topologies used to describe the constant flux of the social. It serves to frame social action as a fluid, embracing both boundless variation and transformation without discontinuity [27]. The constituent molecules of such fluid are social actants[1] bounded by the repulsion and attraction that determine their

[1] Following Latour [24], we use the word actant to refer to social actors that could be either human or nonhuman.

J. Salamanca (✉)
School of Art+Design, University of Illinois at Urbana-Champaign, Champaign, USA
e-mail: jsal@illinois.edu

S. Núñez-Corrales
Illinois Informatics and National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Champaign, USA

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_13

relationships—whether interactions or subscriptions.[2] The structure of a social fluid can also be thought of as a constantly evolving network in which nodes correspond to actants, edges to their interactions, and following the reasoning from fluid dynamics [41], viscosity to the number of interactions that lock agents in place. In such a network, the interactions are ontologically contingent on the actants whose identity and programs of action are, in turn, co-shaped by the number and kind of relationship they hold with their peers. Instead of being rigid, the structure of the network changes with the flux of meanings and actions eliciting a sort of viscosity that permeates the space of possible future actions, sparing no one. In former studies, we defined such social viscosity as the resistance to action flow exerted by actors performing concurrent actions. It is thus the manifestation of mutual disturbances produced by the interactions that bind actors together [33].

Understanding global aspects of such complex systems requires going beyond the analysis of the properties of their constituent actants or their relationships in isolation, observing them both from within and from afar: social fluids are emergent. Summers-Affler's [39] suggests an empirical investigation based on systems theory (e.g., [25]) to study how social action disperses and coalesces the flow of social life, generating temporary turbulent flows and vortexes of attraction. Social fluidity becomes manifest in the construction of identity from the individuals to the collective [29], the characterization of modernity as a tension between that which is fixed and the increasing presence of that which is fluid [42], the constant restructuring of social and spatial mobilities [19], and the prevalent continuity in the evolution of values and norms within societies and organizations [21]. In all of these, social perceptions of actants play a pivotal role: the combination of actant traits and their local circumstances determine how accurate their perception of social events is. We believe, motivated by these observations, that simple models of emergent social self-organization capable of capturing fundamental properties of social fluidity can be constructed [8].

Our approach here is simultaneously systemic and grounded in a phenomenological account of meaning across social life with a simple underlying principle: the consistency of shared social meaning has a direct impact on how social viscosity materializes and acts on social systems. For Schutz [35], the constitution of social meaning includes both the meaning attributed by the actor, deemed *subjective*, and the many others inferred by the observers, deemed *objective*. We are inclined to believe that turbulence and vortexes of attraction emerge, consolidate and collapse when the subjective and objective meanings of a critical mass of agents either converge or diverge in overt streams of action. Assuming that human actors are intrinsically driven to enact their programs of action, the fluidity of their action flow depends on their constant effort to reconcile their objective meanings derived from actions observed in other (third-party) actants, and the subjective meaning underlying their

[2] In our research vocabulary, relationships encompass both interactions and subscriptions. Interactions occur between agents of the same kind, whereas subscriptions occur between agents of different kinds. Both could be either directed or undirected. In the research presented herein we refer exclusively to interactions between humans.

subsequent action. The higher the reconciliation effort—i.e., the need to reduce the impedance between strongly established, viscous internal truths and incoming conflicting information—the more turbulent the action flow.

We report here recent research efforts toward building and analyzing a simple model of emergent social organization where the effect of interactions, constrained by social norms and inner beliefs, is captured by a metric associated with social viscosity. We define *inverse viscosity* (IV) as the measurable difference that emerges when reconciling present internal beliefs with future perceptions of reality; this measure can be defined *locally* with respect to individual actants (LIV), or *globally* as an average (GIV). The social system chosen to test our approach is a self-organizing group acting under the premise of perceived proximity to one another. Our model is an adaptation of Shepard and Cooper's seminal color perception experiment originally intended to demonstrate the isomorphism between the similarity space of our internal representation of colors and the similarity space of the perception of the same colors [37]. Our sorting mechanisms do not require metric magnitudes: subjective representations of similarity or dissimilarity suffice to classify and arrange our world experiences.

The social norm used here is the actant's need to reduce color distance by comparing with other actants, and the inner belief is given by a tolerance value representing how comfortable an actant is to imperfect placement. All the entities of the social fluid act as mutual observers—nonhuman actors were excluded—that analyze the transformation of the system from one state to the next while their acting is supported or constrained by the forces of the network in which they partake. We study the social viscosity by modeling empirical observations and re-enacting them in an agent-based model and simulation (ABMS). The ABMS affords to assess agents trajectory and viscosity to further analyze convergences and reversions. We utilize principles and concepts from nucleation in phase transitions [2] to numerically characterize the outcomes of our experiments.

2 Social Meaning and the Analysis of Proximity

Shepard and Cooper's empirical method asked a group of 37 participants with six types of vision conditions to judge similarities between stimuli in two situations: when those stimuli are actually present and when those stimuli are only named. The task involved arranging four decks of 36 cards by similarity, one at a time, each with two squares of distinct color pairs. Cards were split into colors-only and names+color for present stimuli; names-only and braille-only for named stimuli. The resulting sorted sequences were used to compute matrices, each cell containing the similarity rank between pairs of colors, and averaged by combined condition. Researchers analyzed the data by computing correlations, and performing a series of hierarchical clusterings and non-metric multidimensional scalings (NMDS). We highlight two of their conclusions that apply to color vision subjects: (i) there is a high positive correlation between the similarity of perceived and only-imagined colors,

and (ii) Newton's color wheel emerged for the NMDS analysis of color-normal subjects whether the colors were either perceived or only imagined.

Based on Shepard and Cooper findings, we designed a participatory simulation [28] to gain insights toward building an agent-based model of color proximity that assumes (a) that people are able to map their estimation of color proximity into spatial distances in the world, and (b) that they naturally share Newton's chromatic space. Thus, if we ask people to impersonate one unique color and attempt to position themselves near or far away from others with similar or dissimilar colors, respectively, Newton's color wheel should emerge after some finite number of iterations. In organizing themselves, a fluid topology would emerge and we would be able to observe and analyze the trajectories of action, and assess the social viscosity of the system.

We ran the participatory simulation described above with 18 graphic design students. To validate if they instinctively carry Newton's color space, they were asked to individually sort 20 color cards by similarity. The colors were a selection of distinctive hues from Munsell color chart (5RP, 10RP, 5R, 10R, 5YR, 10YR, 5Y, 10Y, 5GY, 10GY, 5G, 10G, 5GB, 10BG, 5B, 10B, 5PB, 10PB, 5P, 10P). The resulting color spaces were consistent with the visual spectrum of light grading from red to purple, with some exceptions exhibiting an inverted sequence. In terms of shape, some groups of participants formed a circle, whereas others a line. This has relevant implications for the calculation of proximity since purple and red are perceptually closer to each other in a circular color space, and green is farthest from both of them. Conversely, purple and red are farthest in a linear space, and green sits at the middle of the spectrum. Then, students were assigned one unique color card from the sample of 20 distinct hues, asked to clip that card to the two—or one—adjacent, and move to a separate empty room. Once at the second room, each participant chooses a random place and, on facilitator command, attempted to organize by perceived color similarity.

As expected, the participants converged to a final circular arrangement sorted in the same sequence as Newton's color wheel (Fig. 1). Initially, participants wandered around displaying and comparing their color cards. Groups of two or three rapidly formed (Fig. 1, Left). After some internal validation, those who did not feel comfortable among their local neighbors quickly switched groups. Groups with high internal similarity persisted, whereas others dissolved. All groups held constant inter-

Fig. 1 Sequence of circular organization as participants assess their color proximity to one another. Left: formation of early groups. Center: Groups fan out to connect with others. Right: convergence on Newton's color wheel. Still images from a zenithal video recording at lab experiments

nal negotiation when sorting the direction of their color progression. At second 40, three groups prevailed: yellow-orange-reds, yellow-green-blues, and blue-purples. Gradually, groups fan out forming semicircles facing the center of the room (Fig. 1, Center). The smaller groups connected to the larger ones rearranging themselves according to the dominant color sequence. Eventually, after 2:40 min the configuration stabilized in a closed circle where purple and red ends came together (Fig. 1, Right).

2.1 Meaning and Viscosity in Color Agents

A careful look at the dynamics of the participatory simulation revealed a pattern of action flow that we synthesize in a perception-action cycle seeking to preserve the richness of interactions. Schutz's definition of social meaning served us to discretize the perception-action cycle in the following stages:

1. **Stage 1**: Participants chose their own set of interactants. They usually interact with up to five of the nearest participants in their visual perception field. This set can change in the next iteration if others are in proximity.
2. **Stage 2**: Participants perceive nearby colors and estimate the current distance to all their interactants. At that point the objective meaning of the current state of the system is inferred, e.g., if Blue stepped away from Purple, then Purple infers that Blue believes she is too close to him.
3. **Stage 3**: Participants internally translate their color proximity model into an expected distance and contrast it with the current distance to all its interactants. In doing so, each participant assembles the subjective meaning of her following action, e.g., Purple wants to move near to Blue to convey her proximity.
4. **Stage 4**: Participants enact their subjective meaning as they move themselves to a position that minimizes the difference between the expected and the current distance between all their interactants. At that point the system has transitioned into a new state and a new perception-action cycle starts.

Social *(local) inverse viscosity* (LIV) per participant at each instant is the discrepancy between their expected status of the world and the current, collectively enacted status. The perception-action cycle is therefore a constant reconciliation of objective and subjective meanings explicit in the winding course to stability. Correspondingly, *global inverse viscosity* (GIV) of the social fluid is then the average of all the individual viscosities. In this particular model, the global inverse viscosity tends to convergence because all participants share a similar color mental model, yet it varies as interactions unfold. Stabilized groups share areas of lower inverse viscosity fluid spatiality, whereas disintegrating groups sort their trajectories in turbulent spots.

3 An Agent-Based Model of Color Proximity

Based on the perception-action cycle derived from observations of the participatory simulations, we developed an ABMS[3] in a custom made multi-agent programmable modeling environment purposefully built to run on modern browsers. The tool is entirely programmed in JavaScript using primarily P5.js[4] and Chroma.js[5] libraries.

In the ABMS, the objective and subjective meanings inferred from social action become explicit. That is something possible in actual experiments only by asking participants to verbalize their thinking on the fly. The model carries many details from the observations from the participatory simulation to preserve the validity of the perception-action model as much as possible. We expect that the conclusions derived from this simulation to cast a new light on social interaction research.

We present here a brief account of the interaction process. In the appendix, we offer a detailed description of the model following the *Overview, Design concepts, Details* (ODD) protocol.

The model proceeds in discrete time instants (ticks) spaced at 100 ms each. During every tick agents store their current position in a trajectory collection, and filter out the set of agents with whom to interact according to the experimenter's interaction rule choice. If there are no interactants in the collection the interaction is terminated; else, they estimate the magnitude and direction of next step by adding all the anticipated vectors toward each interactant. If the length of the estimated step is greater than the agent's tolerance threshold, they adopt the step's heading and execute the step. Else the step is ignored and the interaction terminated. When all the agents ignore the step, the group has converged in a stable phase.

4 Case Study: Nucleation in Circular Color Scales

4.1 Methods

We performed two experiments using our agent-based model of color proximity. To interpret model outcomes, we resorted to convergence and reversions as primary observables, both responsive to model parameters. *Convergence* characterizes the time required for a group of agents to reach a stable color scale. *Reversions* quantify the number of discrete instants when social inverse viscosity increases. Convergence relates microscale descriptions (i.e., actants) to macroscale property (i.e., structure or function), while reversions represent microscale fluctuations.

---

[3] The multi-agent programmable modeling environment is available at www.smartartifact.com/ColorAgents and the source code is available at https://github.com/SocialViscosityLab/ColorAgents.

[4] See: https://p5js.org.

[5] See: https://vis4.net/chromajs.

We found phase nucleation [20] to effectively describe the process of self-organization in our model. In a nutshell, nucleation is the progressive time-dependent organization of microscale entities into distinguishable stable macro-structures, known as phases. The seminal observations made by Avrami [2] suggest that the appearance of growth nuclei constitute an effective; model for phase formation. Growth nuclei are initial small remnants from prior phases whose stability under those conditions was granted in virtue of their size. They may even appear due to fluctuations or topological defects in the prior system state. When conditions in the context change favorably, growth nuclei trigger the formation of larger aggregates whose stability, under fixed and somewhat homogeneous conditions, depends proportionally on the number of coalesced nuclei. This behavior is well known experimentally, and bears relation to how self-organization into large-scale structures maximizes entropy production and propagation [4].

Convergence is imperfect in individual nucleation experiments due to process fluctuations, yet convergence smoothens at the thermodynamic limit in ensemble experiments. We note that in our model as prescribed by statistical mechanics [3], the self-organization of agents into a color wheel is collectively a stochastic process even when agents make deterministic, local decisions. The latter becomes even more so when the unit of analysis is the *ensemble of runs* rather than individual ones. In the color proximity model, nucleation represents convergence to a correctly ordered color scale as determined by color distance. To quantitatively capture convergence as phase nucleation, we used an alternative form of the Johnson-Mehl-Avrami-Kolmogorov equation [2] that is used to estimate the volume fraction of microstates that have nucleated; our equation captures the decreasing volume fraction $X(t)$ of microstates that have yet to nucleate at time t, subject to the effects of increasing social viscosity as a stable color scale arises:

$$X(t) = \exp\{-kt^n\} \tag{1}$$

k may be interpreted as a rate of convergence and n is associated with the dimensionality of the space where distance is measured.

Our analysis process proceeded as follows. All experiments use a chordal proximity model. For each experiment, a fixed value was established per parameter set, and $K = 5$ repetitions per parameter set were performed. Only GIV and reversions per repetition were considered as dependent variables. We recorded the initial and final GIV after 450 discrete time steps. Values for k and n were estimated by curve fitting through optimization using SciPy [18] (version 1.3.0, Python 3.6) after GIV curves had been normalized against the maximum GIV value observed during each simulation. Finally, average values for k, n, and the number of reversions were computed and used in each experiment class to perform MANOVA and ANOVA tests, respectively. ANOVA tests were performed using Python's `statsmodel` library [36], and MANOVA tests were performed using R [32]. Our analysis is fully reproducible.[6]

[6] GitHub repository: https://github.com/snunezcr/social-viscosity-analysis.

Table 1 Descriptive statistics for color proximity convergence experiment with all-to-all interactions. Mean and standard deviation values (in parenthesis) are provided for $K = 5$ samples and $t \in [0, 450]$. Nomenclature: **IV**: initial GIV, **FV**: final GIV, **R**: reversion count

τ	IV	FV	k	n	R
0.0	7.18 (0.72)	0.15 (0.00)	1.11×10^{-3} (4.31×10^{-4})	1.43 (0.06)	87.00 (8.63)
0.2	6.65 (0.97)	0.15 (0.00)	1.21×10^{-3} (6.28×10^{-4})	1.43 (0.08)	1.20 (1.30)
0.4	6.95 (0.57)	0.15 (0.01)	1.13×10^{-3} (2.06×10^{-4})	1.42 (0.03)	0.40 (0.89)
0.6	6.71 (1.00)	0.23 (0.13)	3.01×10^{-3} (4.29×10^{-3})	1.33 (0.19)	1.20 (1.79)
0.8	6.98 (0.57)	0.26 (0.13)	1.68×10^{-3} (6.22×10^{-4})	1.34 (0.06)	0.60 (1.34)
1.0	7.29 (0.44)	0.42 (0.20)	2.14×10^{-3} (1.06×10^{-3})	1.29 (0.11)	0.40 (0.55)

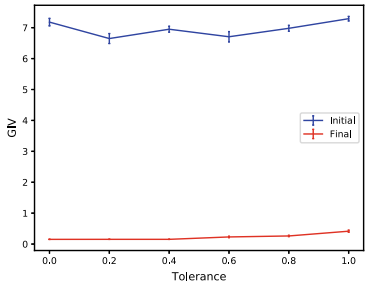
4.2 All-to-all Interactions

All-to-all simulations have tolerance τ to imperfect proximity as the principal parameter. Initial and final GIV values correspond to their maximum and minimum ones respectively. All runs converged to a single color scale ($R < 2.00$). Table 1 contains relevant descriptive statistics for the experiment.

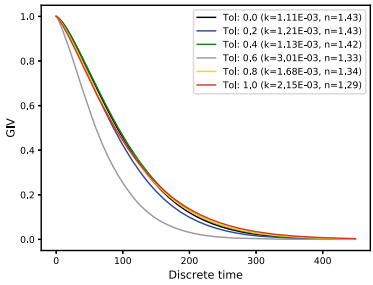
Next, curves were computed for each value of τ and visualized (Fig. 2). While initial GIV appears to depend only on the initial number of agents and their random placement, final GIV increases with tolerance values (Fig. 2a). While most curve fittings yield similar parametrizations, convergence is faster with $\tau = 0.6$ (Fig. 2b). In general, τ appears to play a very significant role in reversions (Fig. 2c): relaxing distance restrictions drastically decreases the number of reversions required for convergence. We also note that reversions increased significantly at $\tau = 0.6$ compared to at $\tau = 0.4$ and at $\tau = 0.8$, suggesting that an analogue to annealing takes place.

We performed a one-way ANOVA in order to determine whether τ induced a significant difference on reversions (Table 2) using ordinary least-squares regression (adjusted $R^2 = 0.987$). The analysis indicated conclusively that τ explains differences in reversions. We also performed a Tuckey HSD test (FWER = 0.05) and found that differences were maximized between $\tau = 0.0$ and all other values (adjusted $p < 0.001$), while no significant differences were found between all other possible pairs (adjusted $p < 0.9$).

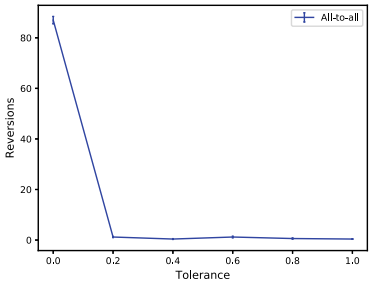
Finally, a MANOVA was performed in order to understand the simultaneous effect of τ over k and n (Table 3). Results indicate that τ explains differences in curve parametrization observed in Fig. 2b. However, the effect over k and n differ per factor (Table 4). Only n appears to be significantly impacted by τ. One possible interpretation is that each value of τ yields a different model of distance, thereby



(a) Initial (blue) and final (red) GIV values for all-to-all circular nucleation.



(b) Time-dependent scale nucleation trends per value of τ (colored lines).



(c) Average reversions in all-to-all scale nucleation.

Fig. 2 Outcomes of simulation experiments with circular color proximity model

Table 2 One-way ANOVA for reversions as a function of τ (adjusted $R^2 = 0.987$)

	SS	df	F	p
τ	30,992.27	5.0	451.89	6.49×10^{-23}
Residual	329.20	24.0	–	–

altering (possibly in non-linear fashion) the metric space in which agents attempt to optimize their positions.

4.3 N Nearest-Neighbor Interactions

We performed a two-factor experiment using the N nearest-neighbor interactions model. Each agent coordinates with its N most proximal agents to gauge its distance. For each value of N, the full range of τ was explored, except for $N = 0$ since no

Table 3 Total MANOVA results for all-to-all interaction experiments

	df	Pillai	Approx F	num df	den df	p
τ (all factors)	1	0.37	7.93	2	27	1.95×10^{-3}
Residuals	28	–	–	–	–	–

Table 4 Per-factor response for MANOVA on all-to-all interaction experiments

	df	SS	MS	F	p
τ (k)	1	5.09×10^{-6}	5.09×10^{-6}	1.57	0.22
Residuals	28	9.10×10^{-5}	3.25×10^{-6}	–	–
τ (n)	1	7.34×10^{-2}	7.34×10^{-2}	7.75	9.52×10^{-3}
Residuals	28	0.27	9.47×10^{-3}	–	–

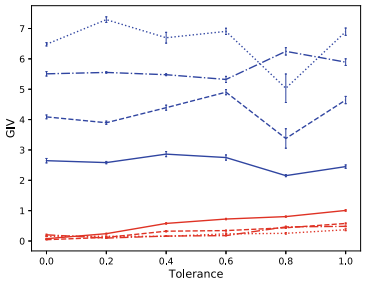
convergence is possible. Results reveal various significant differences with respect to an all-to-all interaction model (Table 5). No longer initial and final GIV correspond to the maximum and minimum GIV, respectively, for all cases (Fig. 3a vs. 3b); a small N induced observed differences. No longer can convergence be guaranteed ($R > 2.0$) except for $N = 20$, similar to the all-to-all interactions model (Fig. 3c). Restricting the number of possible interactions in actants creates a tension between the local and global convergence: the smaller the world, view, the easier it becomes for actants to reach a local minimum and partially converge while experiencing slow (or no) global convergence [30].

Moreover, N appears to drastically parameterize the geometry of the metric space specified by n. Convergence to partial or complete nucleation depends on having sufficient neighbors, at least 50% of the total number of actants in our analysis (Fig. 4b–d). As N decreases, curves spread wider along τ values (Fig. 4b) due to relaxation of constrains experienced when computing distances. We observed an extreme case at $N = 5$ and $\tau = 1.0$ (Fig. 4a, red line): actants only perform a few minimization steps before becoming fixed at the local minimum that takes least effort to reach.

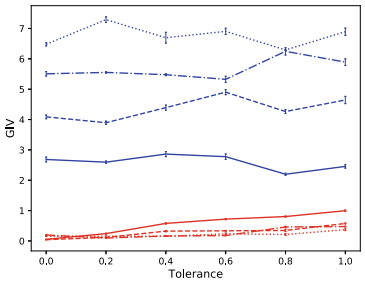
We performed a two-way ANOVA in order to determine whether N and τ induced significant differences on reversions (Table 2) using ordinary least-squares regression. After comparing the fitting of models with (adjusted $R^2 = 0.882$) and without interactions (adjusted $R^2 = 0.876$), our analysis chose the former one. Intuitively, we hypothesized that τ would couple with the number of nearest neighbors to impact reversions, similar to how reversions lead progressively toward lower local minima as in annealing. Significant differences were found (F = 39.71, $p = 3.95 \times 10^{-39}$), τ being the most significant factor followed by N; however, their interaction failed to be significant over the complete dataset except for the $N = 5$ subset. We performed two Tuckey HSD tests (FWER=0.05), one for N and one for τ. Significant differences were found between $\tau = 0.0$ and all other τ values (adjusted $p < 0.001$)

Table 5 Descriptive statistics for color proximity convergence experiment with N nearest-neighbor interactions. Mean and standard deviation values (in parenthesis) are provided for $K = 5$ samples and $t \in [0, 450]$. Nomenclature: IV: initial GIV, FV: final GIV, MV: maximum GIV, mV: minimum GIV, R: reversion count

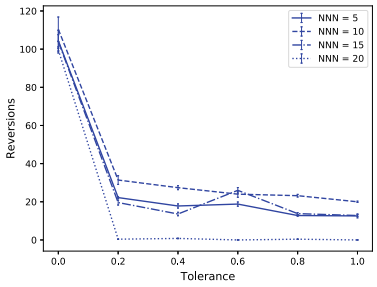
N	τ	IV	FV	MV	mV	k	n	R
5	0.0	2.65 (0.44)	0.08 (0.07)	2.68 (0.48)	0.05 (0.03)	8.64×10^{-3} (8.64×10^{-3})	1.04 (0.17)	104.40 (32.81)
	0.2	2.58 (0.23)	0.24 (0.08)	2.60 (0.23)	0.24 (0.08)	1.28×10^{-2} (1.28×10^{-2})	0.97 (0.16)	22.20 (2.28)
	0.4	2.86 (0.50)	0.58 (0.12)	2.86 (0.50)	0.58 (0.12)	5.48×10^{-2} (5.48×10^{-2})	0.65 (0.18)	17.80 (6.57)
	0.6	2.75 (0.52)	0.72 (0.13)	2.78 (0.55)	0.72 (0.12)	4.85×10^{-2} (4.85×10^{-2})	0.59 (0.09)	18.80 (6.61)
	0.8	2.15 (0.20)	0.80 (0.10)	2.20 (0.21)	0.80 (0.10)	9.93×10^{-2} (9.93×10^{-2})	0.42 (0.05)	12.80 (2.17)
	1.0	2.45 (0.33)	1.01 (0.14)	2.46 (0.33)	1.00 (0.14)	2.42×10^{-2} (2.42×10^{-2})	-9.47 (9.26)	12.60 (6.07)
10	0.0	4.09 (0.37)	0.04 (0.02)	4.09 (0.37)	0.04 (0.02)	3.50×10^{-3} (3.50×10^{-3})	1.24 (0.11)	110.40 (38.59)
	0.2	3.90 (0.32)	0.12 (0.06)	3.90 (0.32)	0.12 (0.06)	4.93×10^{-3} (4.93×10^{-3})	1.13 (0.02)	31.40 (13.72)
	0.4	4.39 (0.50)	0.32 (0.12)	4.39 (0.50)	0.32 (0.12)	7.18×10^{-3} (7.18×10^{-3})	1.08 (0.16)	27.40 (6.31)
	0.6	4.90 (0.50)	0.34 (0.08)	4.90 (0.50)	0.33 (0.08)	4.81×10^{-3} (4.81×10^{-3})	1.15 (0.11)	24.00 (9.80)
	0.8	3.38 (1.93)	0.44 (0.11)	4.26 (0.38)	0.34 (0.22)	7.80×10^{-3} (7.80×10^{-3})	1.01 (0.10)	23.20 (4.15)
	1.0	4.64 (0.72)	0.58 (0.09)	4.64 (0.72)	0.58 (0.09)	1.62×10^{-2} (1.62×10^{-2})	0.88 (0.17)	20.00 (2.00)
15	0.0	5.51 (0.43)	0.20 (0.12)	5.51 (0.43)	0.20 (0.11)	3.01×10^{-3} (3.01×10^{-3})	1.25 (0.08)	104.00 (21.49)
	0.2	5.55 (0.14)	0.10 (0.00)	5.55 (0.14)	0.10 (0.00)	2.15×10^{-3} (2.15×10^{-3})	1.33 (0.11)	19.60 (7.83)
	0.4	5.48 (0.19)	0.16 (0.05)	5.48 (0.19)	0.16 (0.05)	2.07×10^{-3} (2.07×10^{-3})	1.33 (0.07)	13.60 (5.46)
	0.6	5.32 (0.60)	0.18 (0.03)	5.33 (0.60)	0.18 (0.03)	4.32×10^{-3} (4.32×10^{-3})	1.17 (0.10)	26.00 (8.69)
	0.8	6.25 (0.72)	0.46 (0.14)	6.25 (0.72)	0.46 (0.14)	3.31×10^{-3} (3.31×10^{-3})	1.19 (0.08)	13.80 (2.86)
	1.0	5.89 (0.65)	0.48 (0.07)	5.89 (0.65)	0.47 (0.07)	6.30×10^{-3} (6.30×10^{-3})	1.08 (0.09)	12.80 (2.77)
20	0.0	6.48 (0.33)	0.16 (0.00)	6.48 (0.33)	0.15 (0.00)	1.32×10^{-3} (1.32×10^{-3})	1.40 (0.07)	99.60 (9.66)
	0.2	7.29 (0.56)	0.16 (0.00)	7.29 (0.56)	0.16 (0.00)	1.10×10^{-3} (1.10×10^{-3})	1.42 (0.05)	0.40 (0.55)
	0.4	6.70 (1.07)	0.15 (0.01)	6.70 (1.07)	0.15 (0.01)	1.22×10^{-3} (1.22×10^{-3})	1.41 (0.05)	0.80 (0.84)
	0.6	6.91 (0.62)	0.23 (0.15)	6.91 (0.62)	0.23 (0.15)	1.42×10^{-3} (1.42×10^{-3})	1.37 (0.05)	0.00 (0.00)
	0.8	5.03 (2.82)	0.25 (0.14)	6.29 (0.11)	0.21 (0.18)	2.12×10^{-3} (2.12×10^{-3})	1.31 (0.10)	0.40 (0.55)
	1.0	6.90 (0.72)	0.37 (0.18)	6.90 (0.72)	0.37 (0.18)	2.89×10^{-3} (2.89×10^{-3})	1.26 (0.15)	0.00 (0.00)



(a) Initial (blue) and final (red) GIV.



(b) Maximum (blue) and minimum (red) GIV.

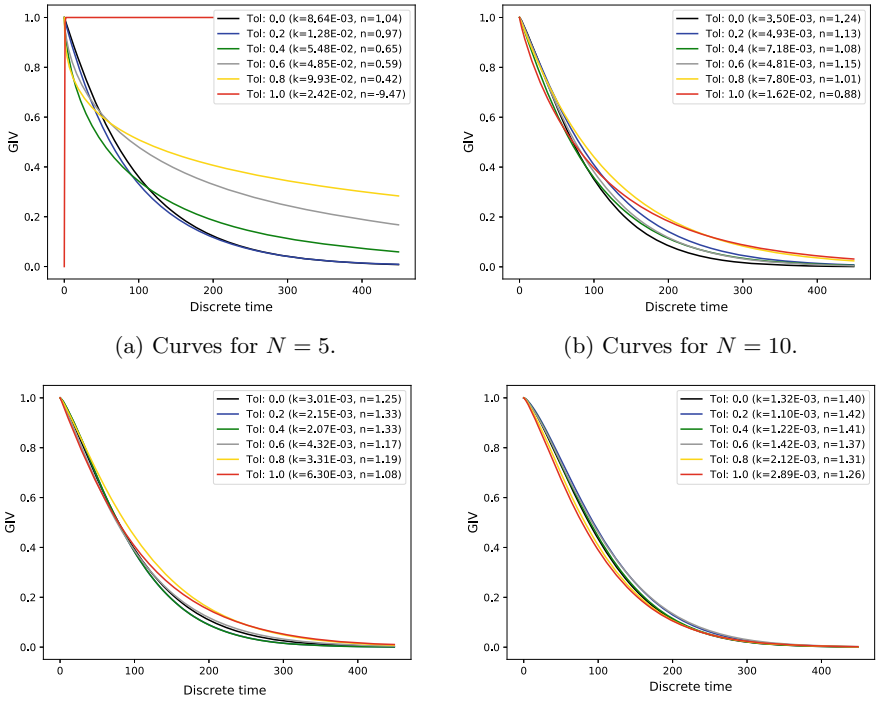


(c) Average reversions in N nearest-neighbor circular nucleation.

Fig. 3 Outcomes of simulation experiments with circular color proximity model. Line styles in (**c**) apply to (**a**) and (**b**)

while other differences between pairs were not significant (adjusted $p < 0.6$). Relaxing FWER to 0.1 produced a significant difference between $N = 10$ and $N = 20$. This outcome may result from the drastic geometry shifts produced by $N = 5$ and $\tau = 1.0$. A restricted two-way ANOVA without $N = 5$ only marginally improved model fitting (adjusted $R^2 = 0.896$) without yielding changes in either Tuckey HSD tests. A two-way ANOVA without $\tau = 1.0$ produces a slightly worse fit (adjusted $R^2 = 0.876$), and the significant difference between $N = 10$ and $N = 20$ at FWER = 0.1 disappears (Table 6).

As with the all-to-all model, we performed a MANOVA using N and τ over k and n (Table 7). τ, N and their interaction explain differences in curve parametrization observed in Fig. 4a–d. As in the prior experiment, the effect over k and n differs per factor (Table 8). We hypothesized that both curve parameters should be impacted by both τ and N, since convergence depends on how τ and N configure the possibilities of each actant; we expected this to be particularly so when the number of nearest neighbors is small. In order of significance, our analysis indicates that τ, its interaction



(a) Curves for $N = 5$. (b) Curves for $N = 10$.

(c) Curves for $N = 15$. (d) Curves for $N = 20$.

Fig. 4 Time-dependent scale nucleation trends per τ value (colored lines) and N nearest neighbors

Table 6 Two-way ANOVA for reversions as a function of τ to imperfect convergence and number of nearest neighbors. Adjusted $R^2 = 0.882$

	SS	df	F	p
N	7963.67	3.0	16.84	7.26×10^{-09}
τ	134874.57	5.0	171.08	3.42×10^{-46}
τ:N	1162.03	15.0	0.49	0.94
Residual	15136.40	96.0	–	–

with N ($\tau * N$) and finally N by itself are significant in explaining observed curve differences. We then proceeded to analyze per-factor responses (Table 8). Differences in k are explained mostly by variations in N followed by τ and $\tau * N$, consistent with the fact that the horizon of differences between perceptions and observations in an actant scales with the number of neighbors: having few of them imposes severe restrictions for nucleation of the color scale. n presented an interesting case: the most relevant factor is $\tau * N$, suggesting that the coupling between τ and N yields a significantly different geometry for the metric space of distances in color space; N and τ follow afterwards.

Table 7 Total MANOVA results for N nearest-neighbor interaction experiments

	df	Pillai	Approx F	Num df	Den df	p
N (all factors)	1	0.47	50.71	2	115	2.20×10^{-16}
Tolerance (all factors)	1	0.21	15.18	2	115	1.41×10^{-6}
N:Tolerance (all factors)	1	0.26	19.74	2	15	4.26×10^{-8}
Residuals	116	–	–	–	–	–

Table 8 Per-factor response for MANOVA on N nearest-neighbor interaction experiments

	df	SS	MS	F	p
N (k)	1	0.02	0.02	51.85	6.44×10^{-11}
τ (k)	1	0.00	0.0	7.70	6.45×10^{-3}
N:τ (k)	1	0.00	0.00	8.29	4.76×10^{-3}
Residuals	116	0.05	0.00	–	–
N (n)	1	76.39	76.39	13.44	3.73×10^{-4}
τ (n)	1	61.75	61.75	10.86	1.30×10^{-3}
N:τ (n)	1	91.62	91.62	16.12	1.06×10^{-4}
Residuals	116	–	–	–	–

5 Discussion

In our simulations, convergence and reversions model two classes of socially accessible evidence in the color proximity experiment: observables that describe the degree of success of perception-dependent social coordination [10], and observables that describe fluctuations (or rather imperfections) in the process conducing to one or many unexpected externalities [31]. Both classes of observables were successfully captured by a seemingly unrelated phenomenon, that of nucleation in phase transition theory, one with an apparently striking ability to synthesize the core elements of social coordination and its various degrees of success. However, the connection between coordinated perceptions of color and phase nucleation lies deep in their mutual relation to thermodynamic systems [34] and fluids [9], both known for their encompassing descriptive universality. Our research exemplifies how the analysis of social situations where coordinated perception occurs can provide new conceptual and theoretical insights backed by existing mathematical physics principles.

When the accessible number of neighbors at any moment is unconstrained, the combination of tolerance to imperfect arrangements and reversions led to an unexpected benefit: convergence occurs more rapidly when actants were less sensitive to imperfect arrangements while remaining sufficiently exposed to novelty (e.g., convergence of partially sorted scales). Essentially, we confirmed the intuition that

efficient coordination requires some degree of tolerance to approximation. When coordination is too strict, the rate of social convergence may not improve. If we conceptualize coordination as an outcome of self-regulation, and if a cost metric per coordination step is added to actants, strictness may produce high-maintenance interactions that can even impair convergence itself [12]. In most occasions and complex social systems, reaching an approximate equilibrium in finite time has much higher social value than achieving perfect coordination [1]. Our simulation suggests that extensions to the model can be made to include cost, and thus to attempt to simulate and gain insights into their dynamics and structure. We anticipate these types of simulation to have future impact in the emergent development of social recommender systems [17], including those for smart cities and urban traffic interventions where dealing more efficiently with uncertainty is critical.

A striking fact uncovered by our simulation is the effect of absolute tolerance to any imperfect arrangement ($\tau = 1.0$) when coupled with the density of social interactions. Only when the number of neighbors was sufficiently small did actants proved resilient to convergence. We believe our observations and analyses provide a formal correlate for the concept of social indifference: in the same fashion in which nucleation in entities with small coordination numbers is unlikely at high temperatures (i.e., agitated states of motion), indifference drives the structure of social arrangements—in our case, the final constitution of the social scale—toward those in which coordination remains always possible, and thus the structure is permanently open to change [13]. Reinterpreted within the framework of causal entropic forces [43], actants collectively maximize their of future freedom of action, represented by the number of possible worlds to which it is possible to converge in the future. *This is still convergence, but of a rather counter-intuitive type*. In hindsight, our experiment appears to satisfy the common intuition that indifference leads to inaction. Understanding this intuition at a deeper level in connection with theories of mind for agents with some form of rationality [16], however, remains an open question. For the moment, we speculate that social convergence toward desirable behaviors may be fostered by interventions that seek to increase the number of inter-personal connections despite prior indifference in actants. In these scenarios, we wish to find values of N that suffice to move k and particularly n from landscapes of inaction to some coordination, even in the midst of indifference.

Reversions provide a powerful window into the mechanics of social turbulence. Turbulence arises when two media with the right viscosity differences meet [7]. In social coordination, this translates to the encounter of smaller, semi-stable scales and their later merging and reconfiguration. Reversions quantify the number of *belief updates* among actants: since actants are memory-less, their only mechanism to decrease local distances (i.e., improve global convergence) is to change places to attempt gaining higher social viscosity. We are then justified in viewing selective spatial translations as a type of language used by agents to minimize the uncertainty of future interactions [5]. Only then, we may reify social turbulence as those configurations in space and time in which new interactions between groups increase uncertainty for them when those groups had prior lowered it their internally. Reversions constitute an imprint of uncertainty reduction in social systems.

Finally, our research suggests that, at least for social systems that can be abstracted as some form of nucleation through phase transition theory, the description of convergence through parameters that determine distances is equivalent to the notion of a social architecture, albeit simple in our case. We use architecture in a generalized manner to refer to the systems of constraints and opportunities in spaces of interaction that drive the functioning of societies and groups. Constraints manifest as patterns of coordination, which together form a vocabulary of interaction [15] capable of allowing us to reason about consequences derived from spatial restrictions. Opportunities arise as actants use their internal machinery to accomplish goals, and in some cases, to learn about the world; all forms of learning presuppose some basic, communicable conceptual scaffolding capable of maximizing advantageous moves and minimizing either fruitless or risky [6] paths. Our simulation also has led us to explore space as machinery [14]: without memory, measuring distances in color space is the primary tool involved in self-organization of the actants. Moreover, the higher the number of interactions the faster the convergence to a stable organization at the cost of a higher social viscosity. Broader social implications of this final point deserve further investigation.

6 Conclusions and Further Work

In this study, we have demonstrated how a simple agent-based model of self-organization based on color proximity is able to capture some fundamental aspects of how social structures converge in the presence of social norms and individual beliefs. While our results are encouraging, we have just begun to understand the implications of our model and experiments in the larger context of social viscosity, fluidity, and turbulence. We have identified two central lines of future work, one around abstract fundamental issues and the other around concrete methodological needs.

At the fundamental level, our intention seeks to deepen the application of statistical physics principles to social situations and permeate the metaphors as far as possible. We wish to explore partial convergence (i.e., phases), the consequences of multiple actants with the same assigned color and the role of small groups in the overall convergence process toward single scales. Landau theory [22, 23] constitutes a promising research direction in this sense. To further investigate this, new types of interactions (including stochastic ones, and mediated by non-humans) become desirable as a means to simulate imperfections in the machinery of an actant and study cyber-physical systems more realistically.

We wish to extend the number and classes of systems for which convergence remains a meaningful descriptor and understand in which social systems our approach is limited. One particular point pertains to verifying and validating the model presented here. While the results of Shepard and Cooper's model are well known, further validation of this process may be achieved by running ensembles of experiments with human participants where an additional adaptive color key tool is provided to simulate tolerance to imperfect color matching. The tool consists of a complete color

wheel masked by a cover wheel concealing all colors but those in a specific segment, and a notch that participants can use to center on their own color within the exposed color range. If the peer's color belongs to the unmasked section, then the participant is instructed to be indifferent to that peer. Clearly, the length of the unmasked segment is proportional to the tolerance factor we have described, and several such tools can be constructed and distributed to all participants per tolerance level. Outcomes of these experiments would then be contrasted against simulated data in a similar fashion as performed in this study.

We also note that the uniformity imposed by having a single color palette across actants is perhaps the most prominent source of convergence. Action meaning is constructed and interpreted from a standard for the whole population. Mixing agents whose color models are the traditional additive or subtractive with agents bearing alternative color perception models rooted in non-western cultures [26], would help elucidate the challenges of proximity-based intercultural interaction. The sensibility function in the model used to estimate color proximity is linear, but it could be fine-tuned to respond accurately to our mechanisms of similarity assessment, inspired by Stevens' psycho-physical power law [38] that explains various ways in which the strength of the stimulus maps to the magnitude of the sensation aroused. In the same vein, multidimensional perception mechanisms could better account for actual social behaviors such as our tendency to favor kinship and familiarity. Learning functions would allow exploring clean slate convergence scenarios (e.g., [11, 40]).

Methodologically, we have various immediate and long-term goals. Additional parametrizations in our model need to be explored, including the use of radial distance functions and field of view as well as linear and exponential distance functions. To this end, we plan to improve our statistics by increasing the number of repetitions ($K \geq 25$). Long term, we wish to extend our experimental platform in three ways. First, we plan to include the ability to perform parametric sweeps. Second, we seek to abstract the existing code to facilitate expressing other models of self-organization with similar features. Finally, we plan to gather the specification of the color scale convergence model as case study for ongoing research aimed at constructing a Generalized Theory of Interactions (GToI).

Acknowledgements S. Núñez-Corrales wishes to thank J. Salamanca for introducing him to color proximity models and the pragmatics of self-organization in social systems, as well as to Prof. Eric Jakobsson for his insights into statistical physics principles and his continued support. He dedicates this work to Les Gasser† (School of Information Science, National Center for Supercomputing Applications and Beckman Institute, UIUC) *in memoriam*, whose advances in agent-based systems have profoundly informed our research. This research has been partially funded by Illinois Informatics at UIUC and the SIGHPC ACMIntel Fellowship for Computational and Data Science (2017).

J. Salamanca wishes to thank D. Gómez for his insights about the participatory experimental design and for introducing him to Roger Shepard's work on non-metric multidimensional scaling of color proximity.

We have no conflict of interest to declare.

Appendix: Description of the Agent-Based Model of Color Proximity

Purpose

The purpose of the model is to replicate the dynamics observed in the participatory simulation of the self-organizing group of people arranging themselves by perceived color proximity. In particular, we want to measure and study the evolution of social viscosity when agents share the same color mental model.

State Variables and Scales

Agents represent normal vision people and are characterized by the following state variables:

- ID: a unique identifier for this agent,
- Position: a vector with spatial coordinates,
- Color value: the unique color value 'impersonated' by this agent. By default the model uses the HSB color space.
- Tolerance: an scalar threshold between 0 and 1 to determine whether or not to move to a new self-estimated position.
- Color mental model: an ordered collection of colors particular to each agent in the world.
- Spatial mental model: a mechanism to convert values from other mental models into spatial distances. For example, it serves to map the perceived proximity between two colors into a spatial proximity. See Submodels section for details.
- Visual perception angle: the scope of visual perception of each agent. It opens toward the direction agent is heading. By default it is set to $\pi * 3/4$.
- Shortest: a scalar used by the spatial mental model to determine how near the agent wants to be from the most similar agent.
- Farthest: a scalar used by spatial mental model to determine how far away the agent was to be from the most dissimilar agent.

The topology of the space where agents move is continuous, unbounded, and unwrapped. Both the temporal and spatial resolution of the model is arbitrary.

Process Overview and Scheduling

The model proceeds in tick intervals of 100 milliseconds. At each tick agents store their current position in a trajectory collection, and filter out the set of agents with whom to interact according to experimenter's rule. If there are no interactants in the

set, the agents set a boolean variable to 'done' and the interaction is terminated; else, they estimate the magnitude and direction of next step by adding all the anticipated vectors toward each interactant. See section Submodels for details on how the step is calculated.

If the estimated step is greater than the agent's tolerance threshold, they set the new bearing, set themselves to 'not-done', and execute the step. If the step magnitude falls below the tolerance threshold they ignore the move, set themselves to 'done' and terminate the current interaction. The boolean variable done/not-done is used to highlight who is inactive in the visualization.

Design Concepts

Emergence: groups emerge from agents with proximal colors. Depending on the initialization settings groups may progressively merge together into a single one with circular shape.

Sensing: Agents perceive both the colors of their neighbors and the spatial distance to all of them. They have functions to estimate color proximity, distance proximity and translate color proximity into distance proximity. See Submodels section for details.

Observation For every tick the data collected for analysis from each agent are its spatial position, and a matrix of their interactions with a detailed description of each interactant. See Submodels section for the details of how social viscosity is computed from each matrix.

Initialization

The experimenter selects four parameters before running the model. First, the color palette from the Color Factory. That defines not only the color palette to be used by agent's color mental model but the size of the agent population. There are as many agents as color in the color palette. Second, the Interaction Rule sets the mechanism used by agents to choose their interactants. Third, a Sensibility function that defines how agents estimate their color proximity to each of their interactants. Fourth, the Tolerance for all agents.

In this research, we used the following parameters: *Color palette*: Munsell color space with 20 colors (5RP, 10RP, 5R, 10R, 5YR, 10YR, 5Y, 10Y, 5GY, 10GY, 5G, 10G, 5GB, 10BG, 5B, 10B, 5PB, 10PB, 5P, 10P) converted into HSB values using the library Chroma.js.[7] *Interaction Rule* either All-to-all and N-nearest. *Sensibility*: Chordal distance. *Tolerance*: six values (0, 0.2, 0.4, 0.6, 0.8, 1.0).

[7] https://vis4.net/chromajs/.

Submodels

Perceived color proximity: returns a value between 0 and 1, where 1 is the farthest perceived distance. The function needs the target and reference color. In the case of *chordal distance* sensibility function, the proximity is estimated by $P = |i - j| * (2\pi/N)$, where i and j are the indexes of each colors in the default color palette, and N is the number of colors in the palette. To normalize the output, if P is grater than π then $P = (2\pi - P)/\pi$, else $P = P/\pi$.

Spatial mental model: a mechanism to map values from other mental models into spatial distances. The distance is estimated by $D = s + (f - s) * v$, where s and f are the agent's shortest and farthest parameters and v is the value within the range between 0 and 1 to be mapped.

Step calculation: An agent's step is the result of the sum of all the corrected spatial vectors to its interactants. More formally, $\sum_{i=1}^{n} \mathbf{v_i} * \Delta_i$, where $\mathbf{v_i}$ is the unit spatial vector to interactant i, and Δ_i is the difference between the current euclidean distance to i and the expected distance to i. The expected distance is estimated using the Spatial mental model.

Local Inverse viscosity (LIV): is estimated at each simulation tick by $\sum_{n=1}^{i} |(c_t - e_t)|/e_t$, where c_t is the current distance between agents at tick t and e_t is the forecast distance between the same agents estimated with the information available at tick t. The result is then normalized by the total number of interactants i at tick t.

Global inverse viscosity (GIV): is the average of the social viscosity sv of all agents i at time t, expressed also as $\sum_{n=1}^{i} sv_t/i$.

References

1. Anshelevich, E., Sekar, S.: Approximate equilibrium and incentivizing social coordination. In: Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 508–514 (2014)
2. Avrami, M.: Kinetics of phase change. I General theory. J. Chem. Phys. **7**(12), 1103–1112 (1939)
3. Avrami, M.: Kinetics of phase change. II Transformation-time relations for random distribution of nuclei. J. Chem. Phys. **8**(2), 212–224 (1940)
4. Bak, P., Tang, C., Wiesenfeld, K.: Self-organized criticality: An explanation of the 1/f noise. Phys. Rev. Lett. **59**(4), 381 (1987)
5. Berger, C.R., Bradac, J.J.: Language and Social Knowledge: Uncertainty in Interpersonal Relations, vol. 2. Hodder Education (1982)
6. Bogenrieder, I.: Social architecture as a prerequisite for organizational learning. Manage. Learning **33**(2), 197–212 (2002)
7. Campbell, I.H., Turner, J.S.: Turbulent mixing between uids with different viscosities. Nature **313**(5997), 39 (1985)
8. Casciaro, T.: Seeing things clearly: Social structure, personality, and accuracy in social network perception. Soc. Netw. **20**(4), 331–351 (1998)
9. Chang, Y.-F.: Social thermodynamics, social hydrodynamics and some mathematical applications in social sciences. Int. J. Modern Soc. Sci **2**(2), 94–108 (2013)

10. Di Paolo, E.A.: Social coordination and spatial organization: Steps towards the evolution of communication. In: Fourth European Conference on Artificial Life, vol. 4, p. 464. MIT Press (1997)
11. Fagyal, Z., et al.: Centers and peripheries: network roles in language change. Lingua **120**(8), 2061–2079 (2010)
12. Finkel, E.J. et al.: High-maintenance interaction: ineffcient social coordination impairs self-regulation. J. Personal. Soc. Psychol. **91**(3), 456 (2006)
13. Hetherington, K., Lee, N.: Social order and the blank figure. Environ. Planning D: Soc. Space **18**(2), 169–184 (2000)
14. Hillier, B.: Space is the machine: a configurational theory of architecture. Space Syntax (2007)
15. Hillier, B., Hanson, J.: The Social Logic of Space. Cambridge University Press (1989)
16. Hollis, M., Sugden, R.: Rationality in action. Mind **102**(405), 1–35 (1993)
17. Jiang, M. et al.: Social contextual recommendation. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 45–54. ACM (2012)
18. Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: Open source scientific tools for Python. Online; accessed <today>] (2001). http://www.scipy.org/
19. Kaufmann, V., Montulet, B.: Between social and spatial mobilities: The issue of social uidity. In: Tracing Mobilities: Towards a Cosmopolitan Perspective, pp. 37–55 (2008)
20. Kelton, K., Greer, A.L.: Nucleation in Condensed Matter: Applications in Materials and Biology, vol. 15. Elsevier (2010)
21. Kolstad, I.: The evolution of social norms: with managerial implications. J. Socio-Econ. **36**(1), 58–72 (2007)
22. Lev Davidovich Landau: On the theory of phase transitions. Ukr. J. Phys. **11**, 19–32 (1937)
23. Landau, L.D.: On the theory of phase transitions. II. Zh. Eksp. Teor. Fiz. **11**, 627 (1937)
24. Latour, B.: Pandora's Hope: Essays on the Reality of Science Studies, p. 324. Harvard University Press, Cambridge, Mass (1999)
25. Luhmann, N.: Social Systems. Stanford University Press, Stanford, Calif. (1995)
26. Maclaury, R.: Color and Cogntion in Mesoamerica: Constructing Categories as Vantages, p. 616. The University of Texas Press, Austin, TX (1997)
27. Mol, A., Law, J.: Regions, networks and fluids: anaemia and social topology. Soc. Stud. Sci. **24**(4), 641–671 (1994). https://doi.org/10.1177/030631279402400402
28. North, M.J., Macal, C.M.: Managing Business Complexity. Discovering Strategic Solutions with Agent-Based Modeling and Simulation. Oxford University Press, New York (2007)
29. Onorato, R.S., Turner, J.C.: Fluidity in the self-concept: the shift from personal to social identity. Eur. J. Soc. Psychol. **34**(3), 257–278 (2004)
30. Pineda, E., Crespo, D.: Microstructure development in Kolmogorov, Johnson-Mehl, and Avrami nucleation and growth kinetics. Phys. Rev.B **60**(5), 3104 (1999)
31. Prigogine, I.: Order through uctuation: self-organization and social system. In: Jantsch, E. (ed.) Evolution and Consciousness: Human Systems in Transition, pp. 93–130. Addison-Wesley, Reading MA (1976)
32. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria (2017). https://www.R-project.org/
33. Salamanca, J.: Smart artifacts mediating social viscosity. In: Vertesi, J., Ribes, D. (Eds.) DigitalSTS: A Handbook and Field-Guide, Chap. 32, pp. 497–509. Princeton University Press, Princeton, NJ., USA (2019)
34. Scafetta, N., Hamilton, P., Grigolini, P.: The thermodynamics of social processes: the teen birth phenomenon. Fractals **9**(02), 193–208 (2001)
35. Schutz, A.: The Phenomenology of the Social World. 1st Paperback. Northwestern University Studies in Phenomenology and Existential Philosophy, p. 255. Northwestern University Press, Evanston, Ill. (1972)
36. Seabold, S., Perktold, J.: Statsmodels: Econometric and Statistical Modeling with Python. In: 9th Python in Science Conference (2010)
37. Shepard, R., Cooper, L.: The representation of colors in the blind, color-blind, and normally sighted. Psychol. Sci. **3**(2), 97–113 (1992)

38. Stevens, S.S.: On the psychophysical law. Psychol. Rev. **64**(3), 153–181 (1957). ISSN: 1939–1471(Electronic),0033–295X(Print). https://doi.org/10.1037/h0046162
39. Summers-Affer, E.: Vortexes of involvement: social systems as turbulent flow. Philos. Soc. Sci. **37**(4), 433–448 (2007)
40. Swarup, S., Lakkaraju, K., Gasser, L.: Learning a common language through an emergent interaction topology. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1381–1383. ACM (2006)
41. Thien, N.P., Tanner, R.I.: A new constitutive equation derived from network theory. J. Non-Newtonian Fluid Mech. **2**(4), 353–365 (1977)
42. Turner, B.S.: Social uids: Metaphors and meanings of society. Body Soc. **9**(1), 1–10 (2003)
43. Wissner-Gross, A.D., Freer, C.E.: Causal entropic forces. Phys. Rev. Lett. **110**(16), 168702 (2013)

Inside the Mind of the Nonfiler: An Agent-Based Modeling Approach

Melissa Vigil, Thi Nguyen, and Ellen Badgley

Abstract Nonfilers are taxpayers who are required to file a tax return but fail to do so, complicating the mission of tax authorities by requiring additional outreach to ensure that tax responsibilities are accurately assessed and collected. In this paper, we describe an in-process effort to develop an extensible agent-based model (the Nonfiler Decision Model, or NDM). This agent-based modeling approach is meant to complement ongoing nonfiler research efforts (econometric modeling/surveys) by the Internal Revenue Service's (IRS) Research, Applied Analytics and Statistics branch (RAAS). Taxpayer data spanning tax years 2006 through 2015 was prepared for calibration of the model. Preliminary high-level results from the agent-based model (prior to calibration/validation against the data set) showed periodicity in nonfiling behavior, interestingly with less effect observed when agents participated in shared social networks than when they acted independently. Future work includes model calibration/validation, more sophisticated representations of social networks and their effects, and more fine-grained temporal modeling to account for week-to-week variation in filing over a simulated tax year.

Disclaimer

The views and opinions presented in this paper reflect those of the authors. They do not necessarily reflect the views or the official position of the Internal Revenue Service.

M. Vigil · T. Nguyen
Internal Revenue Service (IRS), Research, Applied Analytics and Statistics (RAAS), Washington, DC, USA
e-mail: Melissa.Vigil@irs.gov

T. Nguyen
e-mail: Thi.T.Nguyen@irs.gov

E. Badgley (✉)
The MITRE Corporation, McLean, VA, USA
e-mail: ebadgley@mitre.org

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_14

1 Introduction

Nonfilers—taxpayers who are required to file a tax return but fail to do so—are an important segment of taxpayers selected for outreach by tax authorities such as the United States' Internal Revenue Service (IRS). As with other types of tax noncompliance, nonfilers affect overall revenue. The IRS has decreased nonfiler interventions such as Automated Substitute for Return (ASFR) or more invasive methods such as audits, both of which were relatively effective at converting late filers (those who file after the deadline of April 15 without an extension to file) to on-time filers in the past. Research has revealed that U.S. taxpayers exhibit tax compliance rates much higher than neoclassical economics models predict [1], and also demonstrate relatively high tax morale—the moral obligation or intrinsic motivation to pay taxes [2]. Nevertheless, the number of nonfilers has doubled from 10 million in 2005 up to 20 million in 2017, and it is suspected that the decrease in IRS engagement with late filers may have converted them into nonfilers instead of on-time filers [3]. The IRS would like to understand what actions they could take to decrease the rate of nonfiler behavior without incurring the significant costs of auditing or undertaking as many expensive field studies.

This paper presents background on the nonfiler problem, and the institutional challenges of treating nonfilers in light of current budget constraints; a summary of current IRS nonfiler research efforts; and an agent-based modeling (ABM) framework developed by the IRS's Research, Applied Analytics, and Statistics branch (RAAS) as a basis for further investigation of the drivers and dynamics of nonfiling behavior, and potential levers for IRS to convert nonfilers into filers. Of particular interest are the following research questions:

- What are the impacts of *social networks* on nonfiling behavior?
- How can we characterize the effects of *population segmentation* on nonfiling behavior?
- How does the IRS's *enforcement rate* affect nonfiling behavior?

1.1 Brief Overview of Nonfilers and IRS Enforcement/Outreach

Taxpayers in the US are required (under most circumstances) to file a tax return every year. A nonfiler is a taxpayer who has a requirement to file a return but has failed to file for a given tax year. Nonfiling status is determined on a year-by-year basis: a taxpayer may be a filer for multiple years and a nonfiler for others. Theoretically, a nonfiler may at any point change into a late-filer by filing a late tax return.

A recent survey and focus groups conducted on late filers and nonfilers revealed various reasons taxpayers file late or not at all. At least seventy-five percent of survey respondents fail to file because of financial constraints, possibly combined with tax complexity: they either do not have the resources to pay a balance due or cannot

prepare their tax returns on their own and cannot afford to pay a preparer to do so on their behalf. Others become nonfilers because filing is not salient: after experiencing adverse life events (death in the family, illness, unemployment, divorce, etc.), they are distracted from the task of preparing and filing their tax returns. Still, others are nonfilers because of lack of information: they do not receive the tax documents or supporting information from a third party necessary to file their returns. Regardless of the reason, a substantial portion of nonfilers apply for an extension to file (indicating an intention to file their taxes but after the April 15 deadline), but fail to follow through [4, 5].

IRS Outreach to Nonfilers

If IRS does not receive a tax return for a given taxpayer, it initiates an outreach/enforcement process aimed at encouraging the taxpayer to file. The first step is to determine if the taxpayer is actually required to file. To do so, IRS calculates income from taxpayers' information documents (e.g., W-2s, 1099s). If the taxpayer had income sufficient to have a filing requirement, but a tax return has not been filed, IRS will flag the taxpayer as a potential nonfiler and may use multiple processes to encourage the taxpayer to file usually beginning with a series of notices, and in some cases assessing tax using a Substitute for Return (SFR) process.

The notice process is a series of notices sent by the IRS to encourage the taxpayer to file. The first notice (CP59) is softer in tone, but if a return remains unfiled after eight weeks, up to two additional notices (CP516 and/or CP518) are issued. These notices convey that IRS may determine the taxpayer's taxes for them, warns of potential accrual of penalty and interest, and warns that IRS may hold refunds for subsequent tax years until the current tax year's issue is resolved.

Because IRS tries to be certain that these taxpayers are truly nonfilers, the notice process typically begins about nine months after the filing deadline of April 15: the first wave in late November/early December and the second wave in March. Waiting so long to receive a notice about a late return is not ideal because the prior year's tax filing may no longer be salient to the taxpayer and the taxpayer may be preoccupied with the current year's return. Evidence from an early-outreach study suggests that the psychological burden of filing potential penalty and interest on a past return concurrently with the current year's return may suppress the response rate and discourage the taxpayer from filing the current year's return as well [6]. In fact, data show that one of the strongest indicators of nonfiling behavior is the incidence of a prior year's nonfiling [4].

If the taxpayer does not respond to these notices, the case may be referred to Taxpayer Delinquency Investigation (TDI). Some TDIs are assigned to various compliance processing such as the Automated Substitute for Return (ASFR) process, campus or Field Examination processes, or to Field Collection.

The Substitute for Return (SFR) is, as the name indicates, an IRS-produced substitute for the taxpayer's tax return, where tax is assessed using income associated with the taxpayers' information returns. Because IRS does not have the information or the authority to determine certain credits and deductions, those are omitted from the nonfiler's SFR, and (without information about itemized deductions) the standard

deduction is assumed. The resulting estimated taxable income on the SFR can be a substantial *overstatement* of tax due and may further discourage the taxpayer from filing their return once they receive this estimate from IRS. However, taxable income can also be *understated* since IRS does not have tax documents to match with self-employment income or other income not provided on information returns. A review of late tax returns secured after contact indicates that more than one-third of these taxpayers were due a refund of about $3,000, while more than half had a balance due of about $5,500 [4]. The remainder paid their balance exactly throughout the year via withholding and neither had a balance due or a refund.

1.2 Current IRS Nonfiler Research

IRS/RAAS is actively conducting two pilots investigating whether earlier contact with taxpayers is more effective at preventing nonfiling. Both are randomized control trials, and both test the effectiveness of moving up the timing of notices. Based on a statistical analysis of past nonfilers and their characteristics, IRS can predict with reasonable accuracy who is likely to be a nonfiler before the filing season begins. Both pilots test a series of postcards and letters mailed—the first wave is sent just before April 15, another just prior to the extension due date of October 15, and the third in mid-December.

The first pilot [6] was fielded on the tax year 2017 taxpayers who were identified as likely to be nonfilers. Preliminary results from this effort show a small, but a promising reduction in nonfilers. The pilot also showed that postcards were not as effective as letters, that timing of notices matters, that reminding the taxpayer of more than one year depresses filing for the current year, and that letters are more effective than postcards.

The second pilot [5] will focus on the tax year 2019 taxpayers and will also test the timing of notices and compare the effectiveness of postcards vs. letters. However, this pilot will rely on results from focus groups and cognitive interviews conducted on segmented groups of past nonfilers, such as retirees, previously unemployed, chronic underwithholders, and past extension filers, and will test the use of messaging concepts around typical behavioral concepts, like loss aversion, regret minimization, salience/risk aversion, and social norms.

1.3 Past Uses of Agent-Based Models (ABMs) at RAAS

This is not the first investigation by IRS into the use of agent-based modeling for tax compliance studies, but it is the first attempt by RAAS to develop a model specifically focused on nonfilers. Previous models include the Tax Compliance Simulator (TCS), a general model designed to test taxpayer reporting behavior in response to varying enforcement scenarios [7]; the Individual Reporting Compliance Model

(IRCM), a large-scale Agent-Based Model (ABM) that simulates a community of 85,000 taxpayers and incorporates taxpayer learning, social networks, and enforcement measures [8]; and the Small Business Tax Compliance Simulator (SBTCS), which models the evolution of tax compliance strategies among a population of 10,000 taxpayers with business income and a range of tax behavior archetypes [9].

2 Methods

2.1 Agent-Based Model Overview

The Nonfiler Decision Model (NDM) is an agent-based model implemented in NetLogo [10] for exploring the dynamics of nonfiler behavior. Specifically, the model seeks to understand how IRS enforcement rates and activities elicit taxpayer response and subsequent (possible) change of behavior.

The Appendix, included with this paper, provides a complete description of the model in Overview, Design Concepts and Details, and Decision Making (ODD + D) protocol format [11].

The NDM models two types of entities: taxpayers and IRS. These agents interact around the decision to file/not file taxes (on the part of the taxpayers) and to enforce/not enforce treatment on nonfilers (on the part of IRS). As depicted in Fig. 1, taxpayers and IRS interact on an annual basis: taxpayers determine, based on their knowledge and beliefs, whether or not to file their tax return, and then IRS selectively applies an enforcement action to a proportion of nonfilers. Based on whether or not enforcement was experienced, taxpayers update their knowledge/beliefs, which affects their decision to file during the next filing season. Taxpayers are also embedded

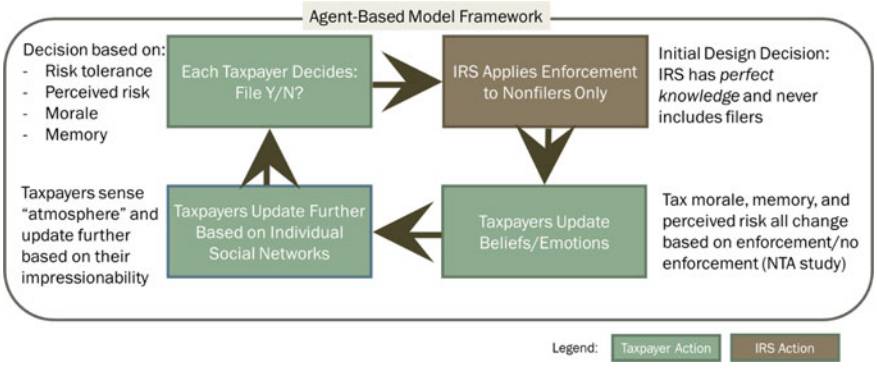


Fig. 1 Agent-based model cycle, representing a single tax cycle

in social networks and share information about their beliefs following enforcement (not their direct experiences) across the network, which in turn influences the decisions made by other taxpayer agents to file or not.

While initial modeling constructs and assumptions are very straightforward, the model is designed so that new modules (for taxpayer behavior, enforcement, etc.) may be substituted at any time; this allows for the exploration of alternative behavioral hypotheses, enforcement methods and policies, and relaxed assumptions about IRS or taxpayer accuracy, as well as increasing fidelity to IRS data as it is gradually incorporated into the model.

Inspirations

Two existing theoretical models of agent decision-making processes were influential in the current and ongoing development of the Nonfiler Decision Model: Agent_Zero [12], and the RAND General ABM of Social Learning [13].

- *Agent_Zero* models an individual software agent as being endowed with distinct affective, deliberative, and social modules, all grounded in contemporary neuroscience. These modules affect agent decision-making in that their intersection determines when an agent reaches an action threshold, generating behavior without the requirement for imitation (hence, the "agent zero" terminology). Agent_Zero also models social network effects not as direct communication, but as a general social "temperature" that affects agents according to the strength of their social bonds [12]. The initial formulation of the Nonfiler Decision Model borrows from Agent_Zero in two ways: it distinguishes among emotional, cognitive, and social aspects of the decision to file or not file, and it models social temperature more than explicit communication about filing/tax decisions. This is an especially significant point because taxpayers do not casually discuss specific tax-related details, especially noncompliance, with their peers or neighbors: the nonfiler focus groups mentioned previously revealed extreme reluctance on the part of participants to be identified as such among their neighbors or friends, perhaps because of personal crises that hinder many participants from filing their tax returns. At the same time, there is clear spatial evidence among these taxpayers, suggesting the presence of social contagion [5].
- *The RAND General Agent-Based Model of Social Learning* also models agent decision-making as a binary choice, but in terms of an agent's individual propensity to act/not act, as informed by the experiences of self and others. An "experience" is defined as a combination of decision and outcome, on the principle that the effect of an outcome on future decisions is conditional on the decision that led to that outcome (for example, reporting taxes accurately and then getting audited by IRS might have a very different effect on future compliance than deliberately underreporting and then getting audited). The experience of each neighbor in the agent's social network will move the agent further toward or away from making the decision being modeled [13]. While the initial version of the Nonfiler Decision Model hews closer to the Agent_Zero than the RAND model, the explicit distinction between experiences and decisions/outcomes, and the ability to model

hidden information (such as discussions of outcomes without the decisions that preceded them) allows for more sophisticated social network communication that may be fruitful in further explorations.

Agent Features and Rules

Agent features and rules are discussed in much greater detail in the ODD + D description of the model in the Appendix.

Taxpayers

Taxpayers have one primary decision: whether to file their taxes during a given tax year (currently modeled as one timestep of the model). Taxpayers may interact within a social network, which can be modeled as a Watts-Strogatz small-world network (a graph with short average path lengths and high clustering, a structure empirically seen in many real-world social networks) [14] or a Barabasi-Albert preferential attachment network (a scale-free network containing several large hubs, a structure that reflects many other real-world networks and especially employment and/or tax preparer networks) [15]. Alternatively, taxpayers may act independently and not communicate with a social network.

Taxpayers are not aware of the IRS's enforcement rate, nor are they directly aware of other taxpayers' decisions or beliefs. However, they are affected by both their own enforcement experiences and (optionally) the general "social temperature" of other taxpayers' beliefs when making the decision to file.

Tax Authority/IRS

The single tax authority/IRS agent decides which nonfiling taxpayers to select for enforcement in a given tax cycle. Each tax cycle, after taxpayers have decided to file/not file, IRS randomly selects a proportion of nonfiling taxpayers determined by the *enforcement rate*. IRS then "applies enforcement" to these taxpayers. The precise definition of enforcement is currently abstract.

Interface

The model interface is depicted in Fig. 2. A visualization of the taxpayers is at the left, surrounded by parameter selectors. Figure 3 depicts the model's plots that track outputs such as the proportion of nonfilers each timestep; filers and nonfilers experiencing or not experiencing enforcement each timestep; the distribution of tax_morale, perceived_risk_of_enforcement, and memory; the overall "tax gap" or cumulative balance due; and the behavior (filing/nonfiling) and state of an individual taxpayer (taxpayer 10) over time.

Segmentation (Model Variant)

The base implementation of the NDM initializes all taxpayers identically, with properties including risk tolerance and impressionability drawn from uniform distributions. A model variant was constructed with two populations of taxpayers, one with a high-risk tolerance (0.8) and one with a low-risk tolerance (0.2). The proportion of

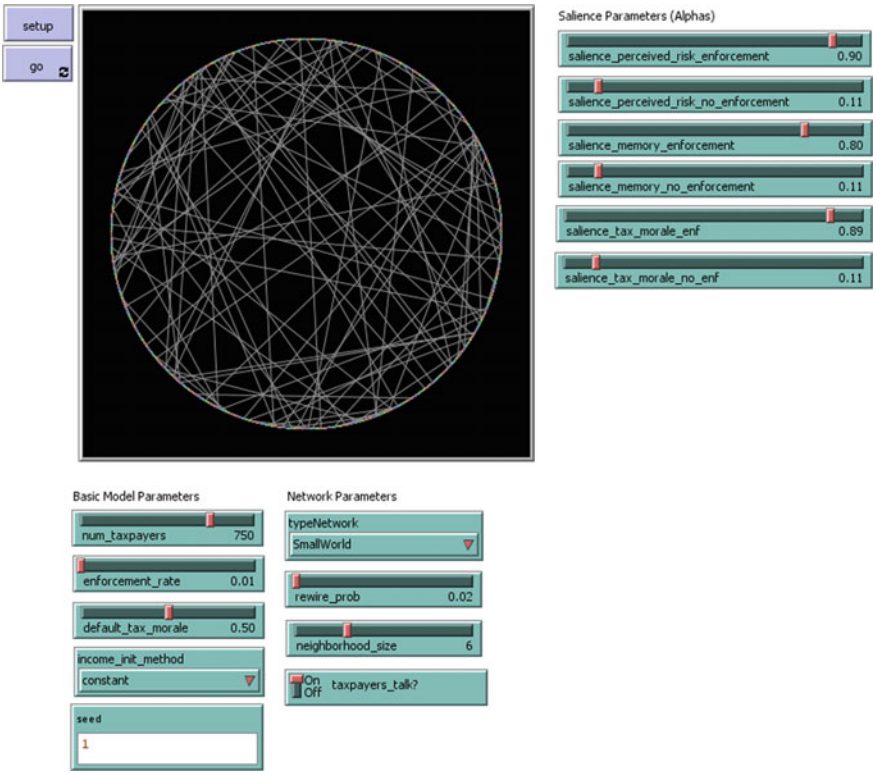


Fig. 2 NetLogo interface for the nonfiler decision model

agents in the high-risk segment was further made configurable via a global parameter. This allowed for the investigation of differences in filing dynamics across population segments.

2.2 Econometric Modeling

An econometric analysis of the taxpayer data set is currently being carried out for eventual calibration/validation of the NDM. The current analysis uses the probability of nonfiling as the dependent variable and applies logistic regression with various explanatory factors and weights. The general form of the econometric model is as follows:

$$p(nonfiling) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n,$$

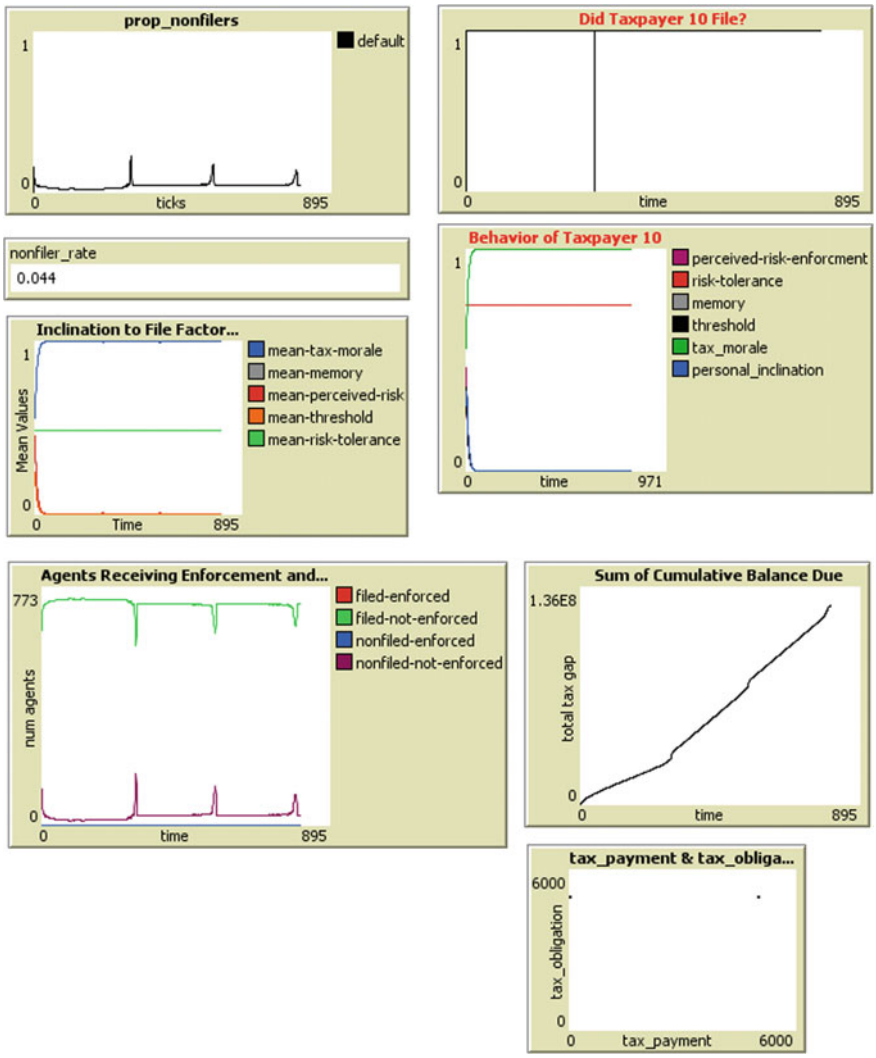


Fig. 3 Model output plots

where β_0 represents the y intercept, β_1 thru β_n are the weights, and x_1 thru x_n are the explanatory variables.

Candidate explanatory variables have been selected from IRS data sources to proxy for taxpayer attributes in the agent-based model: risk tolerance, tax morale, impressionability, and salience/memory. In addition, other proxies for demographics and compliance burden have been identified. The set of potential explanatory variables is as follows:

- *Demographic variables* include age, income, filing status, type, and size of income (wage, nonwage, unemployment, etc.).
- *Risk tolerance variables* may include the DIF (Discriminant Function) score, which rates the potential for change based on past IRS experience with similar returns, and incidence of past nonfiling/late filing/past audit.
- *Impressionability* will make use of tax preparer characteristics. Did the taxpayer use a preparer? If so, was the preparer credentialed? How many returns, on average, are prepared each year by the preparer? How long has the preparer been preparing returns and what is their presence in the taxpayer's community?
- *Tax morale* will use variables that may proxy for civic engagement (charitable contributions—incidence and size, type of investments, etc.) as well as the incidence of past audit, findings, and size of finding; past notices; balance due/refund history; incidence of principal and interest; and taxpayer's neighbors' tax compliance history.
- *Salience and memory variables* are similar to tax morale with respect to the incidence of audits, past notices, and principal and interest. The back history of audit and time since audit are of particular interest [16], as are proxies for life events that may impact salience.
- *Compliance burden* is the time and money expended by the taxpayer to file their tax return. Compliance burden can be proxied by determining if the taxpayer files other returns besides an individual return (e.g., partnership, employment tax).

3 Data

For calibration and validation of the model, a data set was derived from the IRS Compliance Data Warehouse (CDW), containing informational returns data for approximately 600 K taxpayers for the tax year 2011. Information from prior and subsequent tax years was also used to establish filing patterns of behavior. Data regarding taxpayer identities were protected in accordance with applicable agency policies and Federal Law.

The data included features such as biographic information about taxpayers and information about taxpayer earnings, assets, and tax payments. Of course, in the case of nonfilers, the absence of filing information is indicative of noncompliant behavior.

4 Results

4.1 Agent-Based Model

The Nonfiler Decision Model is still undergoing calibration and will be validated against the in-process econometric approach. Furthermore, Design of Experiments (DoE) methods are being applied to understand the sensitivity of the model parameters. Here we discuss several high-level findings regarding model dynamics.

The NDM was run with several distinct input parameter settings or "scenarios", exploring high-level impacts of several model parameters such as enforcement rate, social network communication, and segmentation by risk. This section discusses the results for each of these scenarios. (With the exception of the variant Scenarios 4 and 5, the salience values for the enforcement/nonenforcement curves have been held constant, as sensitivity analysis is still being performed on these parameters.) Table 1 presents model parameters for the scenarios we highlight here.

Scenario 1—Base Scenario

Scenario 1 presents a base scenario in which the enforcement rate is 1% (low). Enforcement has a strong effect/salience on perceived risk, memory, and tax morale (0.1), and non-enforcement has a low effect/salience on these three values. (Receiving enforcement impacts the taxpayer strongly, but not receiving enforcement elicits a much more gradual change/emboldening). Taxpayers are allowed to sense the average perceived risk, tax morale, and memory of their social network neighbors, and adjust their own beliefs according to these averages and their own impressionability—representing general sensing of prevailing opinions about tax compliance (a sort of "taking the tax temperature" of the social network), rather than directly hearing about others' experiences.

Nonfiling behavior under these conditions does not equilibrate: a cycle develops with a periodicity of around 300 timesteps (or "ticks"), as seen in Fig. 4.

Figure 5 depicts how the mean of the taxpayer state variables changes over time. With low enforcement (1%), perceived risk and memory decline quickly, and tax morale rises at an equivalent rate. On the right side of Fig. 5, one can see filers/nonfilers receiving/not receiving enforcement.

The periodicity of the nonfiling rate can be interpreted as taxpayers becoming gradually "emboldened" over time due to repeated lack of receiving enforcement: perceived risk and memory gradually decrease, while tax morale gradually increases. At some point, the taxpayer's inclination to file (determined by perceived risk and memory) dips below their threshold (affected by tax morale and risk tolerance), and they will make the decision to become a nonfiler at that tax cycle. The low enforcement rate (1%) means a slow deterrent response: nonfilers will persist in being nonfilers for several tax cycles. However, once enforcement is received, taxpayers are strongly affected (a great increase in perceived risk and memory, and a sharp decrease in tax morale) and will return to the filing. With social network sensing/communication enabled, this effect of enforcement on one taxpayer will

Table 1 Parameters for sample scenarios of the nonfiler decision model

Parameter	Base (Scenario 1)	No social network sensing (Scenario 2)	Higher enforcement rate (Scenario 3)	Two risk segments (Scenario 4)	Two risk segments, higher enforcement (Scenario 5)
num_taxpayers	750	750	750	750	750
network_type	Small World[a]	None	Small World	Small World	Small World
neighborhood_size	6	N/A	6	6	6
rewiring_prob	0.02	N/A	0.02	0.02	0.02
income_init_method	Constant	Constant	Constant	Constant	Constant
enforcement_rate	0.01	0.01	0.03	0.01	0.03
default_tax_morale	0.5	0.5	0.5	0.5	0.5
salience_perceived_ risk_enforcement	0.9	0.9	0.9	0.3	0.3
salience_perceived_ risk_no_enforcement	0.1	0.1	0.1	0.3	0.3
salience_memory_enforcement	0.9	0.9	0.9	0.7	0.7
salience_memory_ no_enforcement	0.1	0.1	0.1	0.7	0.7
salience_tax_morale_enf	0.9	0.9	0.9	0.4	0.4
salience_tax_morale_no_enf	0.1	0.1	0.1	0.3	0.3
Seed	1	1	1	1	1
taxpayers_talk?	True	False	True	True	True
prop_high_risk_tolerance	N/A	N/A	N/A	0.05	0.05

[a]Initial scenarios presented here use the Small World (Watts-Strogatz) network type exclusively; investigation of the impact of Preferential Attachment (Barabasi-Albert) networks are ongoing, but initial qualitative results do not differ significantly from those described in this section

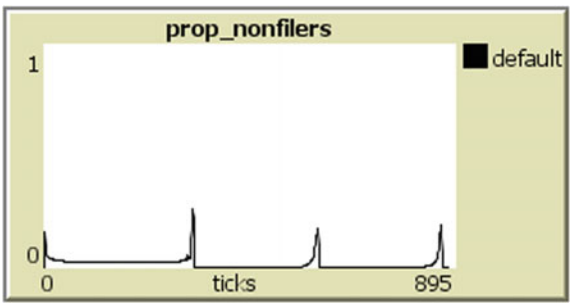


Fig. 4 Proportion of nonfilers over time, scenario 1

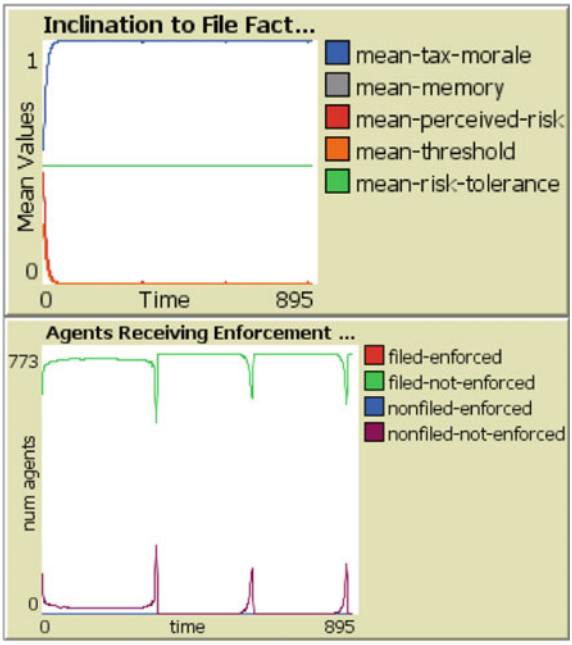


Fig. 5 For scenario 1, (top) mean taxpayer state/belief variables ("Inclination to File" factors), and (bottom) agents filing/nonfiling and receiving/not receiving enforcement over time

spread to its neighbors (six neighbors each in the case of the Small World network) and will quickly dampen the nonfiler rate. Over time, the cycle repeats itself, and the next enforcement received by a taxpayer impacts agents in the social network to a lesser degree. Further investigation is ongoing.

Scenario 2—No Social Network Effects

Scenario 2 uses the same parameter set as Scenario 1, with the exception that social network sensing of "tax temperature" is disabled: taxpayers operate only on their own experience and are not affected by others.

The difference between this and Scenario 1 is striking in that the periodicity is both longer and more pronounced, as seen in Fig. 6. This is understandable in that a) taxpayers are not emboldened by their neighbors' unenforced nonfiling behavior (and therefore high perceived risk/memory), delaying their own experimentation with nonfiling, and b) each taxpayer will persist in nonfiling until it individually experiences enforcement, rather than being deterred by its neighbors' experiences.

Scenario 3—Higher Enforcement Rate

Scenario 3 repeats the parameters from Scenario 1 but increases the enforcement rate to 0.05 (5%). Under these conditions, even a relatively small increase in the likelihood of enforcement is sufficient to drastically suppress the periodic "spikes" in nonfiling behavior—indeed, for most timesteps, the nonfiler rate is zero. Figure 7 displays these results.

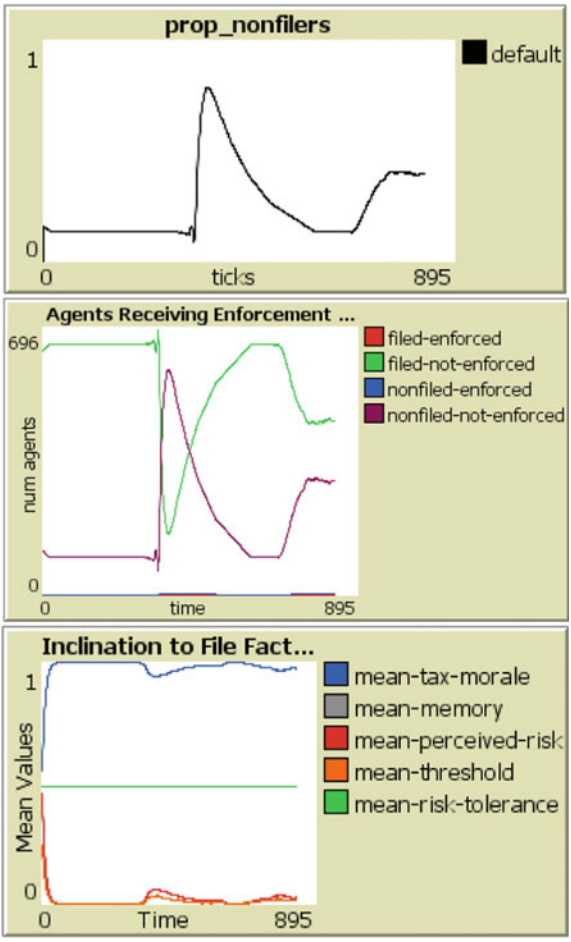


Fig. 6 Behavior of aggregate taxpayers under scenario 2. Top: nonfiler rate or the proportion of nonfilers. Middle: counts of agents filing/nonfiling and receiving/not receiving enforcement over time. Bottom: factors in the inclination to file, averaged across all taxpayers

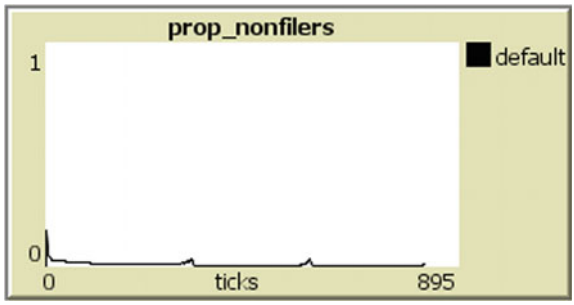


Fig. 7 Proportion of nonfilers, under Scenario 3

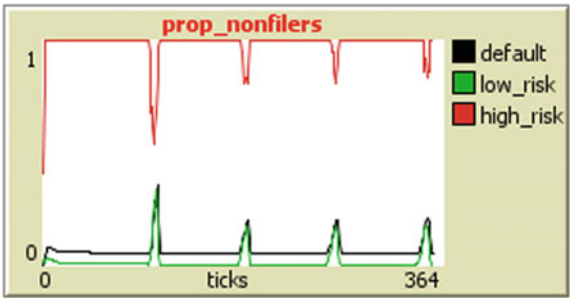


Fig. 8 Two-segment (high-risk and low-risk taxpayers) model, 5% high-risk taxpayers

Scenario 4—Two-Segment, High-Risk and Low Risk

Scenario 4 presents the variant model in which there are two populations of taxpayers, a high-risk-tolerant population (comprising 5% of the whole) and a larger low-risk-tolerant population. The salience/alpha parameters for this run are set differently from Scenarios 1–3 in order to illustrate the dynamics that can emerge from this scenario.

As seen in Fig. 8, this scenario produces the same cyclical/non-equilibrating behavior as seen in the other scenarios, but with a much shorter cycle. High-risk-tolerant taxpayers persist in non-filing behavior, while low-risk-tolerant taxpayers undergo a similar cycle as seen in Scenarios 1 and 2: gradual "emboldening" leads them to experiment with nonfiling. The influence of high-risk taxpayers in their social networks is presumed to be responsible for the shorter cycle. When enforcement is received, low-risk taxpayers are quick to return to compliance—and their experiences affect their high-risk neighbors as well, who (once their neighbors are receiving enforcement) temporarily behave more cautiously, but eventually return to their (default) nonfiling behavior.

Scenario 5—Two-Segment, Higher Enforcement Rate

Scenario 5 duplicates the settings from Scenario 4 but increases the enforcement rate from 1 to 3%. Under these conditions, the cyclical behavior disappears (as seen in Fig. 9): the increased enforcement is enough to keep the low-risk population

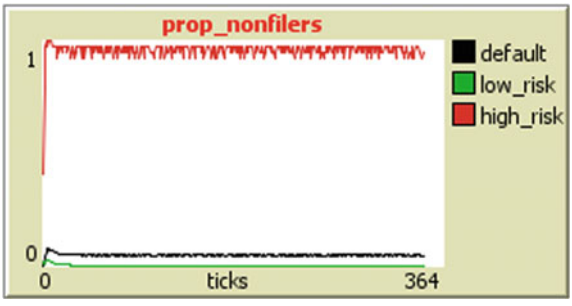


Fig. 9 Two-segment (high-risk and low-risk taxpayers) model, 5% high risk taxpayers, 3% enforcement rate

consistently filing, and all of the fluctuations in nonfiling rate are now to be found within the high-risk population.

5 Discussion

5.1 Agent-Based Model

Preliminary Findings

The scenarios presented in the Results section (0) are preliminary, and much more extensive experimentation and sensitivity/statistical analysis is needed. Still, several observations can be made that align with previous observations and suggest areas for further work:

- Nonfiling behavior is cyclical and non-equilibrating, as seen in all scenarios. Decaying memory and perceived risk lead many taxpayers to experiment with nonfiling behavior, which only ceases when they or their neighbors receive enforcement. The current timestep of the model (one timestep = one tax cycle) is very coarse compared to actual dynamics, so further refinement of the model is needed.
- Social network effects appear to have a strong influence on nonfiling behavior. Taxpayer communication can either sharply dampen out "outbreaks" of nonfiling (as seen by comparing Scenarios 1 and 2) or embolden normally cautious taxpayers to experiment with nonfiling if they are affected by high-risk-tolerant neighbors (as seen by comparing Scenarios 1 and 4). Qualitatively, this aligns with theoretical and econometric observations by Alm et al. [17, 18]. Further work with the sample dataset to impute socioeconomic connections and social (or even preparer) networks will allow for deeper investigation.
- Segmentation by high-level characteristics such as risk tolerance does show sharply differing behavior between taxpayer populations. The conditions investigated allow for the simulation of high-risk taxpayers that are persistent nonfilers until they are influenced by their low-risk neighbors and "tipped" over into filing behavior, a state which is temporary until their perceived risk and memory decrease.
- Overall, a higher enforcement rate does suppress nonfiling behavior. In Scenario 3, higher enforcement reduced the magnitude of the periodic spikes in nonfiling, although it did not remove nonfiling altogether. Scenario 5 is the most interesting in that increasing enforcement for a system in which two groups were segmented by risk kept the low-risk group (95% of the population) consistently filing.

6 Conclusions and Future Work

Agent-based modeling provides a promising direction of research for IRS in understanding the nonfiler problem as well as larger issues of tax compliance; it allows for the modeling of taxpayers as more than purely rational utility-maximizing agents, in ways not easily approached through standard statistical or econometric techniques, or laboratory and field experiments. We present the Nonfiler Decision Model, a general agent-based model for exploring the dynamics and drivers of nonfiler behavior. While the model is in the early stages of development, we hope that when its calibration/validation against taxpayer data is complete, it will aid understanding of this interesting and diverse population and help to identify fruitful intervention efforts.

6.1 Future Work

The Nonfiler Decision Model is in its initial stages. Future planned enhancements include the following:

- *Calibration against taxpayer data* from a representative sample population, to better model income distributions, among other factors.
- *Explicit modeling of further taxpayer segments* such as new to the workforce; new retirees; and habitual nonfilers.
- *Temporal improvements:* A finer temporal resolution (such as weekly), for precise scheduling of treatments (calibration data for this will be derived from ongoing IRS experiments); the ability to track the effects of treatments for longer than a year.
- *Multiple IRS treatments* having different effects on memory/risk/tax morale.
- *Explicit calculation of tax burden*, the monetary cost associated with filing/paying taxes (as affected by tax complexity).
- *Alternative modules/implementations within the NDM for modeling* how memory, perceived risk, and tax morale change in response to experience (for example, different types of decay curves based on other than the Rescorla-Wagner/Agent_Zero paradigm), as informed by ongoing NDM validation and further research. The back audit time period is of particular interest for saliency [16].
- *Multiple decision points for filing* (file on time; file late, after reminder; etc.).

The eventual intent is for the Nonfiler Decision Model to become a modeling framework, with a library of potential alternative modules (different behavioral frameworks for decision-making, different notice streams, etc.) for straightforward exploration of both alternative factors affecting taxpayer filing decision-making, and alternative enforcement/outreach options for the IRS to encourage filing.

Appendix—Overview, Design Concepts and Details, and Decision Making (ODD + D) Protocol Documentation of the Nonfiler Decision Model

Purpose

This model aims to understand the dynamics behind nonfiler behavior and why taxpayers might choose to become nonfilers. Specifically, the model seeks to understand the influence of Internal Revenue Service (IRS) enforcement activities and rates, and taxpayers' response and subsequent (possible) change of behavior due to enforcement. The eventual intent is to develop the model into a framework that can be extended as needed to explore additional enforcement activities, additional behavioral economic models for taxpayers, and additional directions for research. It is implemented in NetLogo.

The primary audience for this model is the Taxpayer Behavior Lab at IRS/RAAS for exploring taxpayer compliance behavior around nonfiling and the potential policy levers available to IRS for affecting compliance behavior, and avoidant economic behavior in general. In its initial form, described here, it is designed for exploratory purposes, with the intent for further validation in the future.

Entities, State Variables, Scales

There are two entities in the model: The tax authority (i.e., the IRS) and the taxpayer.

Internal Revenue Service

IRS is a single agent that (in this current implementation) interacts with taxpayers by receiving filings/payments and leveraging enforcement against taxpayers that do

Table 2 IRS state variables

Variable	Type	Purpose	Value
enforcement_rate	Fixed	The proportion of nonfiling taxpayers against which IRS will issue an enforcement action	(0.0, 1.0]

not file. This is achieved through IRS's main lever, *enforcement_rate*, which could be used as a proxy for something like enforcement staffing (Table 2).

The Taxpayer

Each taxpayer is imbued with a set of attributes as described in Table 3, which affect its subsequent filing behavior and in turn, are affected by its experiences receiving or not receiving enforcement. Taxpayers are organized into a social network, which (in tandem with the taxpayer's *impressionability* factor) affects how much the experiences and attitudes of neighbors affect the taxpayer's decisions regarding filing/nonfiling.

Global Variables

Global variables are used to initialize the simulation and are described in Table 4.

Scale

Currently, a time step corresponds to a single year, but future changes include adjusting the timestep to a week to better understand how the timing of enforcement might affect taxpayer behavior.

Process Overview and Scheduling

At each timestep (representing a yearly tax cycle), taxpayers **decide** if they will file and pay their taxes. If they do pay, an amount is deducted from their cumulative balance due.

Following the taxpayer decision process, IRS identifies nonfilers (with perfect accuracy) and **applies enforcement** to a proportion of nonfilers, depending on the *enforcement_rate.* "Enforcement" in the model is a deliberately abstract concept that covers any number of IRS interventions, including penalties; strong or weak notices/reminders; automated substitute for return (ASFR); or other approaches.

Table 3 Taxpayer state variables

Variable	Type	Purpose	Value
true_income	Fixed	The taxpayer's true income per year	Varies according to the chosen initial distribution
tax_obligation	Fixed	Annual tax obligation for the taxpayer	Currently initialized to true_income * 0.1, or 0, whichever is larger
risk_tolerance	Fixed	A measure of the taxpayer's willingness to take risks. Higher values mean more risk-tolerant, lower mean more risk-averse	[0.0, 1.0]
tax_morale	Dynamic	A measure of the taxpayer's natural attitudes towards/inclination to file and pay taxes. A higher value indicates more willingness to file/pay	[0.0, 1.0]
memory	Dynamic	A measure of the salience of the need to file/pay taxes; higher values indicate higher salience	[0.0, 1.0]
perceived_risk_of_enforcement	Dynamic	How likely the taxpayer believes IRS will apply an enforcement action of some kind on nonfilers. Higher values mean the taxpayer is more confident that enforcement will be applied	[0.0, 1.0]
impressionability	Fixed	How much the taxpayer will be affected by the attitudes of others in its social network. Higher values mean a greater influence from neighbors	[0.0, 1.0]
cumulative_balance_due	Dynamic	A running balance of how much the taxpayer owes IRS. It is affected by payments and increases with earnings	Varies

Future development of the model will explore how the timing and nature of different enforcements affect filing behavior.

Taxpayers then react to enforcement/lack of enforcement by **updating beliefs and attitudes** about filing and IRS's enforcement actions. After individual beliefs are updated, taxpayers optionally are affected by their social network neighbors' beliefs and attitudes (Fig. 10).

Table 4 Global variables for initialization

Variable	Type	Purpose	Value
num_taxpayers	Fixed	The number of taxpayers	[10, 1000]
network_type	Fixed	The type of the social network	"Small World", "Preferential Attachment"
neighborhood_size	Fixed	(For Small World networks) The neighborhood size/number of connections of each taxpayer	[0, 20]
rewire_probability	Fixed	(For Small World networks) The rewiring probability	[0, 1] in steps of 0.01
default_tax_morale	Fixed	The starting value of tax_morale for each taxpayer	[0.0, 1.0]
Income_init_method	Fixed	The method (distribution) for initializing income across taxpayers	"constant", "normal", "uniform", "gamma"
salience_perceived_ risk_enforcement	Fixed	The salience factor (alpha) applied to perceived_risk_of_enforcement when a taxpayer receives an enforcement action	[0.0, 1.0]
salience_perceived_ risk_no_enforcement	Fixed	The salience factor (alpha) applied to perceived_risk_of_enforcement when a taxpayer does not receive an enforcement action	[0.0, 1.0]
salience_memory_ enforcement	Fixed	The salience factor (alpha) applied to memory when a taxpayer receives an enforcement action	[0.0, 1.0]
salience_memory_ no_enforcement	Fixed	The salience factor (alpha) applied to memory when a taxpayer does not receive an enforcement action	[0.0, 1.0]
salience_tax_morale _enforcement	Fixed	The salience factor (alpha) applied to tax_morale when a taxpayer receives an enforcement action	[0.0, 1.0]
salience_tax_morale _no_enforcement	Fixed	The salience factor (alpha) applied to tax_morale when a taxpayer does not receive an enforcement action	[0.0, 1.0]
taxpayers_talk?	Fixed	A Boolean indicating whether taxpayers update beliefs based on social network neighbors	true, false
Seed	Fixed	The random seed used to initialize the model (for reproducibility)	(any integer)

(continued)

Table 4 (continued)

Variable	Type	Purpose	Value
prop_high_risk_tolerance	Fixed	(Model variant only) The proportion of taxpayers that have their risk_tolerance set to the high value of 0.8 (all other taxpayers will fall in the low-risk pool of 0.2)	[0.0, 1.0]

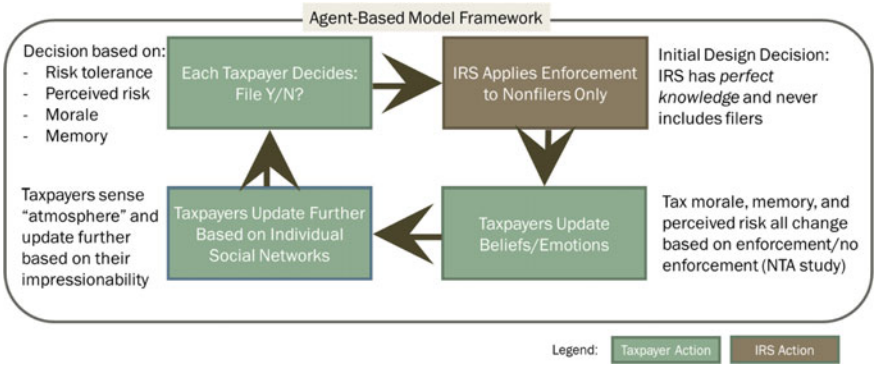


Fig. 10 Agent-based model framework cycle, representing a single tax cycle

Design Concepts

Basic Principles

The current form of the model is grounded in IRS's understanding of the levers driving taxpayer compliance behavior, such as utility; opportunity; prior experience of enforcement, and the salience of those experiences; the awareness of enforcement efforts; social network effects; and social norms [1, 9, 19]. Of particular interest is how spreading awareness of the IRS's relatively low audit rate may affect tax compliance behavior [20, 21]. The overall approach taken for the model borrows heavily from the Agent_Zero modeling framework in two ways. First, it distinguishes between different modules of taxpayer behavior (here perceived risk, tax morale, and memory) with distinct affective/emotional, rational, or social aspects, all grounded in the current understanding of the neuroscience involved. Second, communication among agents is modeled not as direct information-sharing about tax experiences, but a sensing of the "general social temperature" in the form of attitudes and fears about taxes, filing or not filing and the accompanying risk [12].

Emergence

A primary emergent pattern to be considered using this model is how a low enforcement rate propagates throughout the population by affecting the experiences and therefore the *perceived risk of enforcement* of taxpayers over time. We are interested in the effects of this propagation particularly on the tax gap or the sum of the cumulative balance due of taxpayers.

Adaptation

IRS does not adapt—it does not adjust enforcement rate over time, for example, depending on success in applying enforcement against taxpayers.

Taxpayers adapt in a limited way in that they respond at each tax cycle (timestep) to current conditions, namely their perception of enforcement probability, their tax morale, and their memory, as well as the influence of others in the social network. All of these factors contribute to the primary choice of deciding to file or not. Taxpayers do not adjust this strategy on the fly by, for example, changing the weights they apply to memory or perceived risk, or the calculation they use to determine their inclination to file or their threshold.

Objectives

For IRS in a general sense, the objective is to minimize the number of nonfiling taxpayers by controlling the enforcement rate or otherwise maximizing the attributes that contribute to positive filing behavior. Little of this is explicitly modeled in the current form.

For the boundedly rational taxpayer, the objective is to keep as much income as possible while minimizing the chance of receiving enforcement, within the bounds of their respective tax morale and perceived risk of enforcement.

Learning

IRS does not learn. Taxpayers, through interacting with each other and IRS, learn in a limited sense by changing their knowledge of the tax environment as mediated through their and their neighbors' experiences.

Prediction

The chief decision of taxpayers—whether to file or not—is strongly linked to prediction in that the agent (i.e., the taxpayer) makes a calculation as to whether the risk of receiving enforcement is worth taking, conditioned on that agent's personal risk

tolerance, memory, and overall tax morale. Agents who decide not to file are essentially gambling that they will get away with it given low enforcement rates, but their own sense of obligation (in the form of tax morale) affects the predilection to gamble.

In the current implementation of the model, no taxpayer knows the true enforcement rate. Plan for future implementations is to allow some agents to be imbued with knowing exactly what the true audit rate is, which lets them more easily dance around enforcement and affect their network.

Sensing

Taxpayers are only truly aware of their own enforcement experiences. As taxpayers receive enforcement (or not), they update their own beliefs—including *perceived risk of enforcement, memory*, and *tax morale.* While taxpayers are unable to know whether someone in their network received enforcement, they can sense changes to perceived risk from their network and incorporate that sentiment into their own beliefs.

Each taxpayer has their own *impressionability* metric that determines how much they consider the filing compliance sentiment of their network.

Interactions

Taxpayers interact with other taxpayers in their network, represented by bidirectional linkages in their network graph. They do not broadcast their beliefs, but they sense the others as affected by their own impressionability. This could be considered an abstraction of social conversation or "temperature" (on matters related to taxes) of their neighbors.

Stochasticity

Stochasticity is present in several forms, all controlled by the NetLogo random *seed*:

- Many taxpayer state variables such as *true_income, risk_tolerance,* and *perceived_risk_of_enforcement* are randomly initialized at the start. Similarly, taxpayers' social network connections are randomly initialized within the constraints of the network model chosen.
- In the current implementation of the model, IRS only applies enforcement actions to taxpayers who did not file. The set of nonfilers against which to apply enforcement is chosen based on a random draw according to the *enforcement_rate* alone—there is no consideration of income, likely payoff, etc. when selecting taxpayers against which to apply enforcement.

Collectives

There are currently no collectives in the model other than those defined by the modeler (the set of taxpayers, etc.). Future directions may look at self-segregating groups or segments of the taxpayer population.

Observation

Although any aspect of the model state may be observed, the following data are tracked and updated at each time step:

- The number of nonfilers in the tax cycle/year
- The total tax payment against the total tax obligation of all taxpayers
- The mean tax morale, memory, perceived risk, risk threshold, and risk tolerance of all taxpayers
- The number of agents that did file and received enforcement or not, and those that did not file and received enforcement or not
- The distribution of the risk threshold of all taxpayers
- The distribution of the personal inclination to file of all taxpayers
- The sum of the cumulative balances due.

Initialization

All information of previous simulations (global values, timesteps, agents, plots, and all output) are cleared and timesteps reset. A random seed is also selected if the modeler has not specified one.

At initialization, the selected number of taxpayers *num_taxpayers* are created and the Small World or Preferential Attachment network (depending on global settings) is initialized and links created.

When each taxpayer is created, it is initialized as follows:

- The *true_income* is selected from the distribution specified in *income_init_method:*
  - For "fixed", the income is set to 50,000.
  - For "uniform", the incomes are set to a random draw from a uniform distribution with the bounds [0, 50000].
  - For "normal", the income is set to a random draw from a normal distribution with a mean of 50,000 and a standard deviation of 70,000.
  - For "gamma", the income is set to a random draw from a gamma distribution with a mean of 50,000 and a standard deviation of 70,000.
- The tax obligation is set to 10% of the true_income.
- The cumulative balance due is set to 0.

- The tax_morale is set to the default_tax_morale.
- The risk_tolerance is set to a random draw from a uniform distribution between [0.0, 1.0].
  - For the variant model, the number of high-risk-tolerance taxpayers is first calculated using *prop_high_risk_tolerance* and the number of agents. Each taxpayer is then randomly assigned to the high-risk-tolerance or low-risk-tolerance pools using these proportions. Each taxpayer in the high-risk-tolerance pool is given a *risk_tolerance* of 0.8, and the remainders (low-risk-tolerance pool) are given a *risk_tolerance* of 0.2.
- The *memory* is set to a random draw from a uniform distribution between [0.0, 1.0].
- The *perceived_risk_of_enforcement* is set to a random draw from a uniform distribution between [0.0, 1.0].
- The *impressionability* is set to a random draw from a uniform distribution between [0.0, 1.0].

Input Data

At present, the initialization values for the taxpayer agents are notional and have been selected with an eye toward model validation. IRS input data in the future will be used to initialize:

- Income distributions and actual tax liability
- Taxpayer segmentation around tax morale, risk tolerance, and impressionability
- Tax burden.

Submodels

Taxpayers Decide to File

A taxpayer's decision to file is a threshold function. For each taxpayer, a personal inclination to file is first calculated as a weighted sum:

$$personal \cdot inclination \cdot to \cdot file = \frac{(memory \times 2 + perceived\_risk\_enf)}{3}$$

(At present, *memory* is weighted twice as heavily as *perceived risk of enforcement*. This is exploratory.)

The taxpayer's *(risk) threshold* is then defined as a combination of (fixed) *risk_tolerance* and (dynamic) *tax_morale*:

$$threshold = risk\_tolerance \times (1 - tax\_morale)$$

If the taxpayer's personal inclination to file is *less than* their threshold, then they will not file. Otherwise, the taxpayer will file taxes.

Taxpayers Pay Taxes

Only taxpayers who have filed a return will pay taxes. In the current implementation of the model, the tax payment is equal to the tax obligation of the taxpayer (which is 10% of the taxpayer's *true_income*).

At the beginning of each timestep/tax cycle, the *cumulative_balance_due* of the taxpayer is increased by the *tax_obligation* amount. If the taxpayer files (and pays), the *cumulative_balance_due* is then reduced by the tax payment/*tax_obligation* amount, resulting in a balance due of 0 for that given tax cycle.

IRS Applies Enforcements

IRS counts the total number of nonfiling taxpayers and multiples this number by the *enforcement_rate*—this is the *number to enforce.*

These taxpayers then receive enforcement. Currently, this is modeled by informing the taxpayer that enforcement was levied. It (currently) has no financial impact (this is an area for future development).

Taxpayers React to Enforcement/Lack of Enforcement

After IRS has acted, taxpayers react to their filing/nonfiling experiences during this tax cycle. A taxpayer's *perceived_risk_of_enforcement*, *memory*, and *tax_morale* are all affected by its experiences during a particular filing cycle.

Overview

At present, all three of these state variables are affected according to the same general equation, inspired by the use of the Rescorla-Wagner model for associative strength in classical conditioning (this model is used as the emotive component in the initial form of the Agent_Zero framework) [12, 22]. Expressed in general terms, the strength of an experience—in this case, receiving an IRS enforcement will have an impact on the taxpayer proportional to both the current and the maximum possible associative strengths for that experience and affected by the salience (alpha/rate parameter) of the experience itself.

Figure 11, for example, shows how a taxpayer's belief in the model, such as the perceived risk of enforcement by IRS will be reinforced by consistent, repeated enforcement experiences. The less strong the belief (strength of association in the Rescorla-Wagner model) is to start with, the stronger the effect the enforcement will

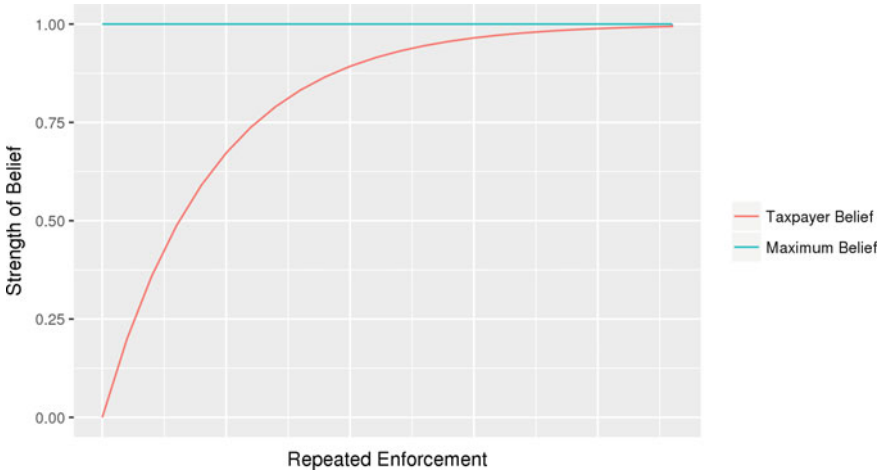


Fig. 11 Rising taxpayer strength of belief (for example, in the perceived risk of enforcement) following repeated enforcement experiences

have. Explaining this an alternate way, the more surprised the taxpayer is to receive enforcement, the more impact that enforcement will have on the taxpayer's beliefs and expectations going into the next tax cycle. There is a maximum strength of belief—here 1.0—where the taxpayer is certain that he/she will receive IRS enforcement. The alpha or salience parameter controls the shape of the curve and how rapidly the strength of belief changes.

Figure 12 shows how the strength of belief decreases in the model following a *lack* of enforcement in a given time step. In this case, the strength of belief decreases

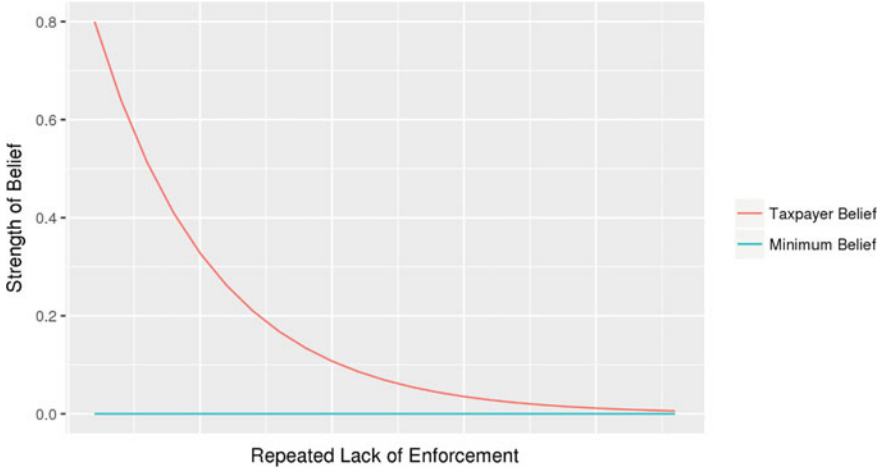


Fig. 12 Falling taxpayer strength of belief following repeated lack of enforcement experiences

following an alpha/salience parameter, which for this example is identical to that in Fig. 11.

In the Nonfiler Decision Model, the salience of enforcement is not assumed to be the same as for lack of enforcement, and the two parameters can be tuned separately for each of the *perceived_risk_of_enforcement, memory, and tax_morale* state variables of the taxpayers. The salience of (negative) enforcement is presumed to be greater than the salience of lack of enforcement. Figure 13 depicts an example salience curve pairing, where the salience of enforcement on taxpayer belief is stronger (alpha = 0.2) than the salience of non-enforcement (alpha = 0.05).

The choice of alpha/salience parameter affects the shape of the curve and the responses of taxpayers vis-à-vis changing strength of beliefs. Figure 14 presents an example of the effects of different alpha/salience values.

Equations

For perceived_risk_of_enforcement, memory, and tax_morale, the minimum is 0.0 and the maximum is 1.0.

If a taxpayer receives enforcement, their *perceived_risk_of_enforcement* will be affected as follows:

$$\begin{aligned} &perceived \cdot risk \cdot of \cdot enforcement \\ &\quad = perceived \cdot risk \cdot of \cdot enforcement \\ &\quad + salience \cdot perceived \cdot risk \cdot of \cdot enforcement \\ &\quad * (1.0 - perceived \cdot risk \cdot of \cdot enforcement) \end{aligned}$$

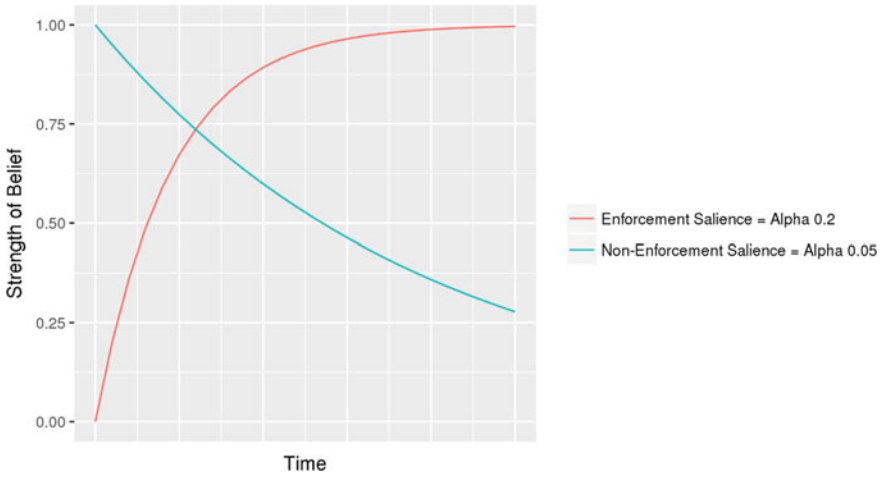


Fig. 13 Example of "non-symmetrical" salience curves: the salience of non-enforcement is less than of enforcement

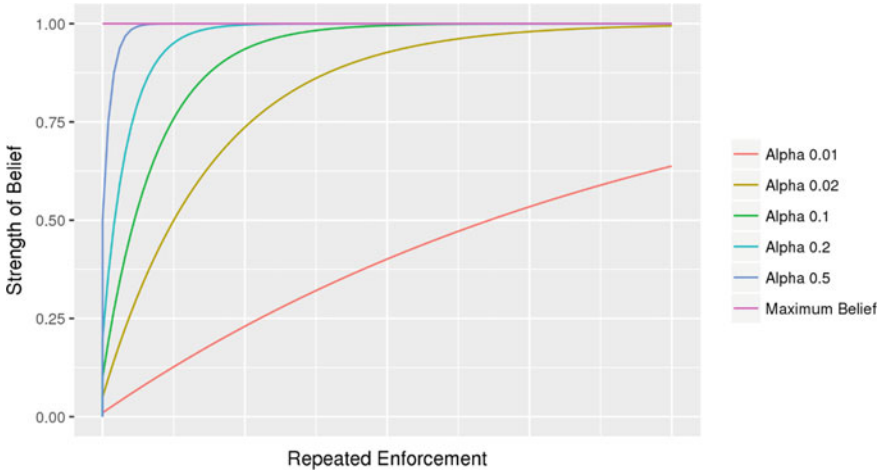


Fig. 14 Influence of alpha on the magnitude of belief strength, hereafter repeated enforcement. A larger alpha can result in a drastic increase in response

That is, the *change in perceived risk* is the difference between the current *perceived_risk* and the maximum *perceived_risk* (1.0), multiplied by the parameter representing the salience of enforcement on perceived risk. The *perceived risk* is then modified by this change. The higher the alpha/rate/salience parameter, the more dramatic the perceived risk will increase as a result of experiencing enforcement.

When the taxpayer does not receive enforcement, the *perceived_risk_of_enforcement* is affected as follows:

$$\begin{aligned} perceived \cdot risk \cdot of \cdot enforcement \\ = perceived \cdot risk \cdot of \cdot enforcement \\ + salience \cdot perceived \cdot risk \cdot no \cdot enforcement \\ * (0.0 - perceived \cdot risk \cdot of \cdot enforcement) \end{aligned}$$

This brings the *perceived_risk* closer to its minimum (0.0), using an alpha/rate different from the above. Lack of enforcement is assumed to have less of a dramatic impact on the taxpayer's perceived risk than enforcement: the taxpayer's heightened sense of risk will diminish slowly in a year in which they do not receive enforcement.

Memory changes currently follow a similar pair of curves to those described above, using the *salience_memory_enforcement* and *salience_memory_no_enforcement* alpha/rate values. The team is exploring alternative approaches to modeling memory.

Tax_morale uses a third set of alpha/rate values, *salience_tax_morale_enforcement,* and *salience_tax_morale_no_enforcement*. The difference here is that the directions of the curves are reversed for *tax_morale*: when enforcement is received, *tax_morale* is assumed to *decrease*, whereas no enforcement will cause it to increase again.

Taxpayers "Talk" (Optional)

After all taxpayers have updated their *perceived_risk_of_enforcement, memory, and tax_morale*, they are (optionally) affected by "talking" with their social network (if the *taxpayers_talk?* global variable is TRUE). The term "talk" is a bit of a misnomer, as this is more of sensing of the general "temperature" regarding risk, memory, and morale.

Using the example of *tax_morale*, taxpayer state variables are affected via a weighted average of their own state and the states of others, as follows:

$$tax \cdot morale = \frac{(tax \cdot morale + impressionability\ (avg\ tax \cdot morale\ of\ neighbors))}{1.0 + impressionability}$$

The same update is performed for *memory* and *perceived_risk*.

References

1. Alm, J.: 12. Expanding the theory of tax compliance from individual to group motivations. In: A Handbook of Alternative Theories of Public Economics, p. 260 (2014)
2. Torgler, B., Schneider, F.: What shapes attitudes towards paying taxes? Evidence from multicultural European countries. Soc. Sci. Q. **88**(2), 443–470 (2007)
3. S. a. R. D. a. A. (. a. P. A. (PPA), Interviewee, Personal conversation [Interview]. 1 August 2019
4. Colon de la Matta, J., Guyton, J., Hodge II, R., Langetieg, P., Orlett, S., Payne, M., Qadri, A., Rupert, L., Schafer, B., Turk, A., Vigil, M.: Understanding the nonfiler/late filer: preliminary findings. In: 2016 IRS Research Bulletin (2016)
5. IRS/Fors Marsh Group: Nonfiler focus groups and cognitive interviews (Unpublished internal report) (2019)
6. Herlache, A., Orlett, S., Javaid, R., Roy, I., Turk, A.: Enforcement vs. outreach—impacts on tax filing compliance. In: 2019 IRS Research Bulletin (2019)
7. Bloomquist, K.: Multi-agent based simulation of the deterrent effects of taxpayer audits. In: Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association, vol. 97 (2004)
8. Bloomquist, K. M., Koehler, M.: A large-scale agent-based model of taxpayer reporting compliance. J. Artif. Soc. Social Simul. **18**(2) (2015)
9. Bloomquist, K.: Tax compliance as an evolutionary coordination game: an agent-based approach. Pub. Fin. Rev. **39**(1), 25–49 (2011)
10. Willensky, U.: NetLogo. Evanston, IL. Wilensky, U.: Netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL (1999)
11. Muller, B., Bohn, F., Dressler, G., Groeneveld, J., Klassert, C., Martin, R., Schluter, M., Schulze, J., Weise, H., Schwarz, N.: Describing human decisions in agent-based models—ODD+D, an extension of the ODD protocol. Environ. Model. Softw. **48**, 37–48 (2013)
12. Epstein, J.M.: Agent_Zero: Toward Neurocognitive Foundations for Generative Social Science. Princeton University Press, Princeton (2014)
13. Nowak, S.A., Matthews, L.J., Parker, A.M.: A General Agent-Based Model of Social Learning. The RAND Corporation, Santa Monica, CA (2017)

14. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440 (1998)
15. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
16. Hokamp, S., Gulyas, L., Koehler, M., Wijesinghe, S.: Agent-based modeling of tax evasion: theory and application. In: Agent-Based Modeling of Tax Evasion: Theoretical Aspects and Computational Simulations, pp. 3–36. Hoboken, NJ, Wiley and Sons (2018)
17. Alm, J., Bloomquist, K.M., McKee, M.: When you know your neighbour pays taxes: information, peer effects and tax compliance. Fisc. Stud. **38**(4), 587–613 (2017)
18. Alm, J., Clark, J., Leibel, K.: Enforcement, socioeconomic diversity, and tax filing compliance in the United States. South. Econ. J. **82**(3), 725–747 (2016)
19. Antunes, L., Balsa, J., Respicio, A., Coelho, H.: Tactical exploration of tax compliance decisions in multi-agent based simulation. In: International Workshop on Multi-Agent Systems and Agent-Based Simulation. Springer, Berlin, Heidelberg (2006)
20. Eisinger, J., Kiel, P.: How the IRS was gutted. ProPublica, 11 Dec 2018 [Online]. http://www.propublica.org/article/how-the-irs-was-gutted. Accessed 23 July 2019
21. Eisinger, J., Kiel, P.: After budget cuts, the IRS' work against tax cheats is facing collapse. ProPublica, 1 October 2018 [Online]. https://www.propublica.org/article/after-budget-cuts-the-irs-work-against-tax-cheats-is-facing-collapse. Accessed 11 July 2019
22. Rescorla, R.A., Wagner, A.R.: A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II: Curr. Res. Theory **2**, 64–99 (1972)

Capturing the Effects of Gentrification on Property Values: An Agent-Based Modeling Approach

Niloofar Bagheri-Jebelli, Andrew Crooks, and William G. Kennedy

Abstract Cities are complex systems that are constantly changing because of the interactions between the people and their environment. Such systems often go through several life cycles which are shaped by various processes. These may include urban growth, sprawl, shrinkage, and gentrification. These processes affect the urban land markets which in turn affect the formation of a city through feedback loops. Through models, we can explore such dynamics, populations, and the environments in which people inhabit. The model proposed in this paper intends to simulate the aforementioned dynamics to capture the effect of agents' choices and actions on the city structure. Specifically, this model explores the effect of gentrification on population density and housing values. The proposed model is significant in its integration of ideas from complex system theory which is operationalized within an agent-based model stylized on urban theories to study gentrification as a cause of an increase in land values. The model is stylized on urban theories and results from the model show that the agents move to and reside in properties within their income range, neighboring agents that have similar economic status. The model also shows the role of gentrification by capturing both the supply and demand aspects of this process in the displacement and immobilization of agents with lower incomes. This is one of the first models that combine several processes to explore the life cycle of a city through agent-based modeling.

N. Bagheri-Jebelli (✉) · W. G. Kennedy
George Mason University, Fairfax, USA
e-mail: nbagher2@gmu.edu

W. G. Kennedy
e-mail: wkennedy2@gmu.edu

A. Crooks
University at Buffalo, Buffalo, USA
e-mail: atcrooks@buffalo.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_15

1 Introduction

All cities change over time, with periods of growth and development followed by shrinkage and demolition. They transform in size [3], form [4], density [20], and land values [1]. Over the last two decades, a modeling approach has emerged to simulate urban environments and their transformations through time focusing on individual interactions from the bottom up. This approach is that of Agent-Based Modeling (ABM) which has been utilized in various studies (e.g., [30, 41, 43, 46]). Such an approach provides a platform for spatial and agent-agent interactions for heterogeneous commodities to be traded and for studying the resulting non-equilibrium dynamics. As such, ABM is an approach suitable for modeling land markets [43]. To this point, agent-based models have often focused on examining one aspect of a city's life cycle; in this paper, we examine multiple aspects of urban dynamics.

To focus on the formation of micro-dynamics between agents, this model is designed in an artificial environment, from stylized facts [36, 44]. While utilizing real data to set parameters such as budget, income, and rent, the model aims to examine the development of urban dynamics using assumptions and behavior rules that simplify and abstract the real world so that they can be implemented in an ABM framework. Agent-based models are structured with different sections, including agents, agent attributes, assumptions, and rules, as the framework to study specific phenomena. To inform each of these sections, we review relevant literature and use the material to create the various pieces (sections) of the model. By using plots and monitors, the model allows us to follow the evolution of the city throughout the simulation process.

The agent-based model studied in this paper focuses on demonstrating the relationship between gentrification and housing value. To examine this relationship, fluctuations in the housing value are studied in every neighborhood in which agents reside and trade properties. Then, patterns of gentrification by demand and supply are searched to understand their effect on fluctuations in population density and housing value in a neighborhood. By observing the relationship between housing value and gentrification, we can gain a deeper understanding of invasion and succession in the environment that leads to the rise and fall of a neighborhood, shaping the city cycle. Geographical factors such as topology are not studied in the model due to their actual and concrete nature which creates a specified rather than general spatial environment. Every method and theory applied to the model is structured to build a hypothetical city. In the remainder of this paper, we discuss the related works in Sect. 2, before introducing our model design in Sect. 3. We present our results in Sect. 4 and discuss them in Sect. 5. Finally, we draw some initial conclusions in Sect. 6.

2 Related Work

To develop a full picture of the life cycle of a city, several models and theories have to be taken into account. The urban dynamics of growth, sprawl, shrinkage, and gentrification have been previously examined by researchers (e.g., [17, 23, 42, 46]). Previous models and theories have examined the various processes and subsystems of an urban environment, but as cities comprise a variety of interacting processes and subsystems, they show novel results when studied as a whole rather than in isolation [28]. As the focus of the model presented in this paper is on the life cycle of a city, its research requires multiple phases. All the phases rely on understanding urban land markets. The urban land market studies that are included in the literature review use existing cities as their case studies while similar to the work of other researchers [35, 43], this model will present a hypothetical city. The model will demonstrate urban growth based on flat land with concentric zones represented by multiple circular rings around the Central Business District (CBD) [9]. Residential land prices are inspired by Alonso [1]. The residential prices vary based on distance from the CBD which causes competition among agents who prefer to be in close proximity to the city center. One could consider the life cycle of a city comprise of processes such as urban growth, sprawl, shrinkage, and gentrification. Urban growth is described as a spatial system growing by expansion and compaction. The expansion results from a geometric extension caused by an increase in space being occupied, while the compaction results from an increase in density or intensity. Urban growth has been a topic of interest among researchers developing descriptive and mathematical models dating back to the early twentieth century and through their most current computer simulations and models (e.g., [5, 9, 12, 29]). Urban growth and urban sprawl are highly interlinked. However, it is important to note that urban growth can occur without leading to sprawl, while urban sprawl must be generated from urban growth [8].

Urban sprawl is based on population mobility and occurs in a process called suburbanization. It results from the demand for greater space and lower density [5] among car-dependent communities [23]. Based on the static model of cities, residents want to be closer to the CBD. This desire, along with the residents' preference to be as far from the congestion of city life as feasible, contributes to the formation of urban sprawl. Therefore, this growth in mobility is observed at the edge of the city. Urban sprawl has been modeled mainly in the past two decades [6, 48]. Urban sprawl happens simultaneously as a city grows or shrinks.

The shrinkage of an urban area is a product of population decline caused by deindustrialization and out-migration from the inner-city. Urban shrinkage can have two reasons: economic (i.e., long-term industrial transformation such as in the USA) or demographic (i.e., falling birth rates such as in Germany) [52]. The process of urban shrinkage results in an oversupply of housing and a decline in housing prices which then accelerates migration in the region and causes an increase in the housing prices of the areas where migration happened. Urban shrinkage has been studied and modeled using case studies of various cities across the globe [26, 52]. Schwarz and

Ernst [46] argued that analyzing the housing price oscillation resulting from urban growth and shrinkage mirrors the dynamics between supply and demand.

Gentrification, also affected by supply and demand, has been described as the "middle-class settlement in renovated or redeveloped properties in older, inner-city districts formerly occupied by a lower-income population" [25]. This urban process has been studied both from the supply and the demand aspects and modeled utilizing each aspect [30, 33, 47]. Gentrification happens either through a bottom-up process of people attracted to the neighborhoods located near the CBD, with low rent prices [17, 30, 33] or by developers who recognize the opportunity in the Potential Rent (PR) of the land [47]. Developers seize these opportunities that arise in the market [13, 21]. The developers' goal is to earn profit and increase their Net Operating Income (NOI) [39].

Smith's [47] Rent Gap Theory presents an understanding of the developers' decision-making process. The rent gap explains the supply side of the rise of the land value and developers' decision to revitalize. The theory suggested by Smith [47] states that every piece of land could have an improved value (capitalized rent) or unimproved potential rent. The capitalized rent (CR) is the actual rent on the land under current use while the potential rent (PR) is the possible rent under the highest and best use. Initially, the capitalized and the potential rent are equal, but in time, the property loses value as it becomes obsolescent which causes the capitalized rent to decrease. At the same time, the potential rent of the property remains the same or increases as the investments within the city increase. This process creates a rent gap which is the opportunity ground for developers to gentrify the land by injecting new capital, rehabilitating buildings, and investing in large-scale residential projects [17]. This chain of events leads to gentrification and changes the land market at the aggregate level by increasing the potential and capitalized rent of the neighborhood [47]. The developers owning higher capital than individual households find the rent gap a profitable opportunity. The process of demand leads to the changing of a neighborhood, resulting in an increase in its density and land value, which may lead to sprawl, in the sense that the supply process increases land value in a previously shrinked neighborhood, causing growth.

Models of the urban structure have also been studied to understand the aforementioned processes. The three classic models of urban growth and structure are the concentric zone model [9], the sector model [29], and the multiple nuclei model [27]. The zonal model considers the city to form in concentric circles with the central business district (CBD) as the core [9]. This model presents the process of urban expansion in terms of invasion and succession of one zone into the next outer zone, resulting in the physical expansion of the city [9, 25]. The source of urban growth and instability among communities is mobility. The general circles following the CBD (also known as the loop) are the factory zone, the immigrant residential zone (low-income blue-collar residents), and the single-family housing zone (middle and high-income white-collar residents). These models are useful in providing a spatial base for modeling urban land markets.

Land markets form where land is traded or purchased monetarily or through services. Social, political, cultural, economic, legal, and environmental factors affect

land markets, and land markets affect them in return [25]. Land use is the management and modification of the natural and artificial environment. The Alonso model [1] also known as the bid rent model is perhaps the most famous model of land use and forms the basis of urban land market theory. Alonso's [1] bid rent model is constructed in zones of different land uses within an urban area. The model considers accessibility, a major parameter in determining variations in land use, land value, and intensity. The residential location pattern is organized based on the trade-off between three main parameters, travel cost, rent, and space needed by the household. Like the Burgess [9] model, this example focuses on the CBD as the agent's work destination. The model, grounded in location choice feedback loops, adaptation, and evolution, demonstrated a bidding and competing process between firms and firms, residents and residents, and firms and residents [15]. Alonso's [1] model forms a distance-decay relationship between location-rent and distance from the center, where residential properties with the lowest bid rent curves are positioned in the outer zone [15]. Regardless of the model implemented, the spatial environment of that model will largely affect the narrative and results.

The neighborhood where an agent resides or is deciding to reside in paints a local picture that creates global patterns. Both abstract and geographically detailed spatial representations have their utilities and purposes. While the specificity in geographically detailed spatial representations allows for an accurate and extensive study of a given case, abstract spatial representations allow for a more general study that aims to exhibit the dynamics and interactions of a system and its subsystems. The traditional cellular spaces define neighborhoods either in the Von Neumann style with an agent in the center cell and four neighboring cells covering each side or the Moore style with the center cell and eight neighboring cells covering each side and corner. To examine the effects of spatial structure on segregation, Flache and Hegselmann [19] applied Schelling's [45] model to irregular grids using a Voronoi tessellation. The neighbors in this model are defined as cells that share common borders with the main cell [19]. The results of the Flache and Hegselmann [19] model showed that the size and structure of a neighborhood does not affect the outcome of segregation. Table 1 demonstrates example applications of models of urban dynamic researched and selected to design the model presented in this paper. This model is inspired by the parameters, environments, and agent behaviors of the models presented in Table 1. The models noted in Table 1 employ mathematical and cognitive approaches for the decision-making of their agents [14, 31]. While the mathematical models use ad hoc direct and custom coding of behaviors, cognitive models implement cognitive frameworks to capture better human decision-making [14].

3 Model Design

The agent-based model demonstrating the life cycle of a city is designed containing various elements based on the research presented above. Building on the theories and models discussed in the previous section, this NetLogo model intends to simulate the

Table 1 Example applications of models of urban dynamics

Author	Application	Entity	Behavior	Spatial scale	Temporal scale
Benenson [7]	City dynamics	Individual	Cognitive and mathematical	Neighborhood and city	Years
Crooks [15]	Residential segregation	Individual	Mathematical	Neighborhood and city	Years
Devisch et al. [16]	Residential choice	Individual	Cognitive and mathematical	Neighborhood	Years
Landis [32]	Urban growth	City	Mathematical	City	Years
Schelling [45]	Segregation	Individual	Mathematical	Neighborhood	Not specified
Haase et al. [26]	Urban shrinkage	Household	Mathematical	Neighborhood and city	Years
Batty and Xie [5]	Urban growth and sprawl and decline	City	Mathematical	Neighborhood and city and regions	Years
Clarke et al. [12]	Urban growth and sprawl	Individual	Mathematical	Neighborhood	Years

effect of gentrification by supply and demand on density fluctuation through urban growth, sprawl, and shrinkage, and its effect on property value. While the model includes calculations for all of the dynamics, in this paper, we explain those relevant to gentrification and land value. For interested readers, the model and data to run the model are available at https://github.com/niloofar-jebelli/UrbanDynamics.

3.1 Data

The data used in the model is stylized on the real-world data of Washington, D.C. Stylized data provides the flexibility to explore a model for the purpose of understanding its dynamics, rather than extracting exact outcomes. The data concerning income, budget, housing, and land rent prices are extracted from the websites of the United States Census Bureau (USCB) [49], Mayor Muriel Bowser Office of Planning (MMBOP) [37], the Economic Policy Institute (EPI) [18] and the Urban Land Institute (ULI) [50]. This data is focused on Washington, D.C. with its 131 neighborhoods for modeling with realistic amounts. According to the Census Bureau [49] between 2010 and 2014, from the 306,184 housing units in the District, 37.6% were single-family units while 62.4% were multifamily units. There were 277,378 occupied housing units or households of which 40.6 percent were owner-occupied and 59.4 percent were renter-occupied. The average household size was 2.2 persons. The median value of an owner-occupied unit was $486,900. The median household income with a mortgage was $125,870. All the data is gathered for the creation of the model. The data is then used as a reference for the input parameters of the model. Adopting real data for simulation input and development has a great effect on the

validity of the process and output. Table 2 demonstrates in detail the input parameters, their range of values, default settings, and references. The majority of the default values of the developers, professionals, non-professionals, and properties is initiated at the beginning of every simulation.

3.2 Environment

The environment is modeled with the bottom layer of patches as houses in a Moore neighborhood definition of 3×3 cell configuration to generate an abstract spatial representation. The middle layer holds Voronoi tessellations representing a region in the city (e.g., a neighborhood). To form the abstract urban structure, the top layer is formed according to the zonal model by Burgess [9]. The zones are designed (from the center out) with the CBD in the core, then the inner-city, and then the suburbia. The CBD and suburbia are considered to be more expensive zones, while the inner-city is less expensive as is often the case in many cities [9, 27, 29].

3.3 Agent Classes

Inspired by various studies, the agents are designed to be in three categories: professionals, non-professionals (e.g., [30, 35]), and developers (e.g., [17, 22]). Agents make decisions and move in a temporal scale expressed as time-steps. Each time-step of this model is notional but could be considered to represent one year because all the rules and assumptions are set to adjust in that manner to accommodate the rates (e.g., shrinkage rate). Similar to Schelling's [45] model, time is purely notional in this abstract model but can be considered in yearly intervals [44]. The model is also run for 300+ time-steps to represent enough time in the life cycle of a city.

3.3.1 Professionals and Non-professionals

Professional and non-professional agent groups have incomes, budgets, and preferences. Professional agents have a higher income than non-professional agents. Professional and non-professional agents are randomly located in the inner-city and suburbia (the two rings around the CBD). The agents that are initialized with a random income and a budget then move around based on their preferences until they are satisfied. The agents' initial state is unhappy to ensure their movement across the landscape to reach a happy state when their preferences are met. Professional and non-professional agents check unoccupied properties based on their available budget for renting the property, preference of being closer to the CBD, inspired by Alonso [1], with at least 20% of its neighbors being of the same class, inspired by Schelling [45], or the same color, and choose a housing type with travel cost

Table 2 Input parameters of the urban life cycle model

Parameter	Value	Default	Reference
Developer	Normal distribution (mean, standard deviation): $N(\mu, \sigma)$		
State	Happy/unhappy	Happy	Benenson [7], Schelling [45]
Income	N(5,000,000, 4,000,000)		Miles et al. [39]
Budget	69% of annual income		Miles et al. [39]
Saving	Income–budget		Miles et al. [39]
NOI	≥ 0	0	Miles et al. [39]
Professional			
State	Happy/unhappy	Unhappy	Benenson [7], Schelling [45]
Income	N(137,814, 20,728)		EPI [18], MMBOP [37], ULI [50], USCB [49]
Budget	69% of annual Income		EPI [18], MMBOP [37], ULI [50], USCB [49]
Saving	Income–budget		EPI [18], MMBOP [37], ULI [50], USCB [49]
Housing	28% of budget		EPI [18], MMBOP [37], ULI [50], USCB [49]
Non-professional			
State	Happy/unhappy	Unhappy	Benenson [7], Schelling [45]
Income	N(42,814, 10,938)		EPI [18], MMBOP [37], ULI [50], USCB [49]
Budget	69% of annual Income		EPI [18], MMBOP [37], ULI [50], USCB [49]
Saving	Income–budget		EPI [18], MMBOP [37], ULI [50], USCB [49]
Housing	28% of budget		EPI [18], MMBOP [37], ULI [50], USCB [49]
Properties			
State	Occupied/vacant	Vacant	Author's estimation
Zone	Inner-city/suburban		Burgess et al. [9]
Type	C: Condo/S: single family house		MRIS [38]
Size	C: N(926.96,31.26), S: N(1650.40,504.14)		MRIS [38]
Price	C: N(492,867,14,715.43), S: N(769,387,201,379.80)		MRIS [38]

(continued)

Table 2 (continued)

Parameter	Value	Default	Reference
Age	[0–100]		Author's estimation
Potential rent	>0		Diappi et al. [17], Smith [47], Author's estimation
Capitalized rent	>0		Diappi et al. [17], Smith [47], Author's estimation
Environment			
Cap rate	[4.75–7.75]	7.50	CBRE [34]
Urban growth rate	[0–1]	0.17	CBRE [34]
Initial population	[100–1000]	1000	CBRE [34]
Gentrification rate	[0–1]	0.26	CBRE [34]
Sprawl density threshold	[0–1]	0.2	CBRE [34], Author's estimation
Sprawl moving rate	[0–1]	0.1	CBRE [34]
Shrinkage rate	[0–1]	0.0050	CBRE [34]

in relation to their budget. The parameters of travel cost, rent, and space that play central roles in Alonso's [1] bid rent theory are used as inspirations for the preference setting of the professional and non-professional agents. The developers are driven by the profitability of the land. They check their available savings, the vacancy of their interested residential type (single-family house or multi-family), assess market demand, get an appraisal of the property value using a neighborhood index and using the methods explained in the previous section, and search for the rent gap using a rent gap threshold [17, 47]. The movements arising from the interaction of the agents with each other and their environment shape the state (growing, sprawling, shrinking, or gentrifying) of the city modeled.

3.3.2 Developers

The developers are fixed agents because their movement is not essential to the model, and it is only their decision-making and developments that affect the model. They depend on the property value and the capitalization rate (cap rate) to earn their NOIs. The cap rate determines the rate of return on a real estate investment based on the income generated by the property. Based on the 2014 Coldwell Banker Richard Ellis (CBRE) cap rate survey report [34], Washington, D.C. cap rates for multi-family housing market are between 4.75 and 7.75%. This range is used in the model as a slider which can affect the developers' NOI and the market condition. Developers are in the CBD symbolizing working in the city center. The developers also have randomly assigned budgets and preferences. Although the agents are limited by their incomes and budgets, their decision-making process determines a large portion of their movements. In the model, the developers make their decisions based on available

budget, vacancy of their preferred residential type, appraisal of value, rent gap, and market demand. Developers buy vacant properties if the rent gap [47] is high and the property age is 60 or higher. By doing so, inspired by Gilbert et al. [22] model of the English housing market, they change the properties' ages to zero which makes them great options for agents to buy. The developers also acquire properties when unhappy professionals or non-professionals sell their property to the developer with the most development in their neighborhood if the source of their unhappiness is an increase in population that does not match their preferences. The properties that are developed take the color of their developer. When an agent purchases a property from a developer, that agent is paying the price of the property from its savings. This transfer places the purchased money in the developer's savings and adjusts the NOI. Continuing with Diappi and Bolchi's [17] method, neighborhood IDs were used to count the patches within them and apply the summation of their capitalized rents to arrive at the neighborhood rent parameter. This method helps the developers make better assessments of the market value of their properties.

3.3.3 Properties

According to the Metropolitan Regional Information System (MRIS), the average sold price of a two-or-less bedroom detached property in 2015 was $553,782, while an attached property was $512,290 [38]. The same market analysis showed that the 2015 average sold price per square feet for all property types averaged $491 and ranged from $453 to $509. The properties in this model are divided into single-family houses developing in the outer circle and condominiums developing throughout the three zones. Each property has a spatial location, occupancy, a CR, and PR and a state of decay. Properties have ages, types based on their zones, sizes based on their types, and renting and buying prices based on their sizes. Property age is randomly assigned between zero to one hundred. Properties in the inner-city zone are condos and coops, while properties in the suburban zone are attached and detached single-family houses.

3.4 Gentrification Calculation

Focusing on the central question of this paper, while there are underlying mechanisms for all the urban dynamics introduced, we will present those causing and affecting gentrification by supply and demand. From the prices determining total value, rental prices were achieved using United States Census Data on the 32.02 price-to-rent ratio [51] in Washington, D.C. As mentioned by Smith [47], the property capitalized and potential rents are equal. In time, the capitalized rent declines and creates a rent gap which explains the supply side of gentrification. This decline is demonstrated by applying a decay rate to the age of the property using Diappi and Bolchi's [17] method. Governing analysis from the American Community Survey and Longitudinal Tract Database [24] shows a 51.9% gentrification occurring in Washington, D.C. between

2000 and 2015. Gentrification by supply occurs in the model as follows: for a patch that is unoccupied, the potential rent is calculated as presented in Eq. 1:

$$PR = \frac{p}{r \times 12} \tag{1}$$

where p is the property price and r is the price-to-rent ratio. As presented in Eq. 2, we estimate the *PR* using the *CR* and a decaying function over the age of the property:

$$CR = PR \times e^{-\lambda \times a} \tag{2}$$

where λ is the decaying parameter and a is the age of property. We use $\lambda = 0.04$ throughout the experiment since it provides a more realistic value for the *CR*. As demonstrated in Eq. 3, we use the *PR* and the *CR* to calculate the rent gap:

$$\text{Re}ntGap = \frac{PR - CR}{PR} \tag{3}$$

Due to its effect on land value and urban dynamics, an important element in the model is neighborhood density. Let the number of agents in an area be n and the number of possible agents in an area as np, neighborhood density is calculated in Eq. 4 as follows:

$$\text{density} = \frac{n}{np} \tag{4}$$

Density is a significant parameter for gentrification and urban sprawl [10, 33], and therefore needs to be accounted for as we do in Eq. 5; as the density increases, the property prices increase as follows:

$$p = e^{1+\text{density}} \times (CR \times r \times 12) \tag{5}$$

The direct relationship between density and housing price is demonstrated in their corresponding plots in Figs. 3 and 4. While gentrification by supply is modeled by setting the preferences of the developers, gentrification by demand is modeled by setting the preferences of the professional and non-professional agents. Then, 30% of the agent population is randomly selected to have an interest to pay low rent, while everyone has the interest to be near CBD. Gentrification by demand is represented by calculating the number of non-professional and professional agents who prefer to be near CBD and pay low rents. Gentrification by demand is considered to have taken place if the professional agents who can afford to live in suburbia decide to live in the inner-city due to the aforementioned preference, and by doing so cause displacement to the non-professionals. Putting all the agent behaviors and urban dynamics in the model together, the interconnectivity and interactions of the subsystems within themselves and with each other become clear. Figure 1 demonstrates

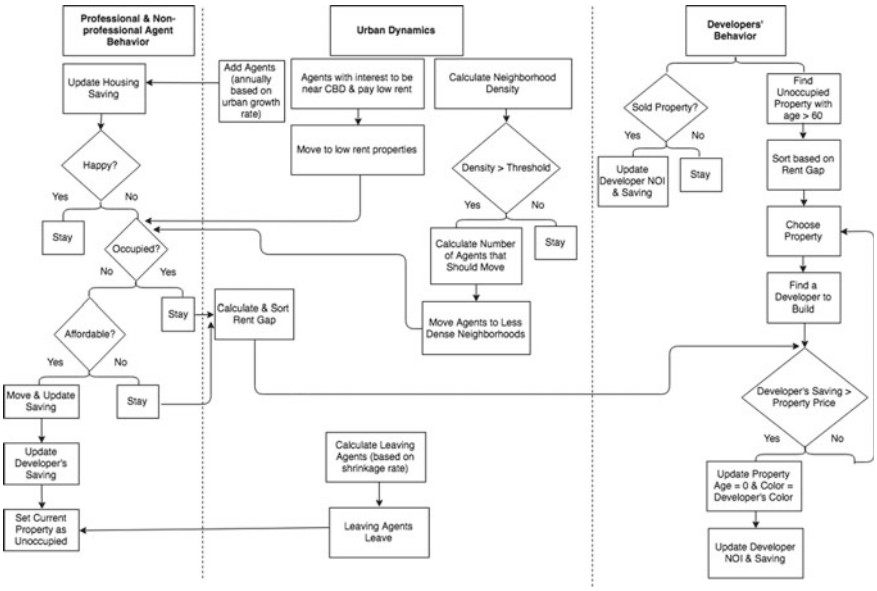


Fig. 1 Model logic and interactions between entities during a simulation

this interconnectivity from the initiation of the model for the course of a simulation.

3.5 Verification

As the purpose of this model is to capture the qualitative agreement with emerging patterns of macro-structures forming a city cycle, it is validated as a level one type according to Axtell and Epstein's [2] validation levels. Therefore, the mode was examined in accordance with its level type. An essential part of building an agent-based model is verifying the model for correct performance [11]. To implement the model, various verification procedures have taken place. We walked through the code to ensure the matching of the model inputs with the background data. Then, we performed testing measures such as printing the outputs of each section of the code for debugging. Once we gathered the output data and results, we visually inspected the tables, figures, and plots to track the behavior of the variables and verify their intended performance. Finally, by observing the behavior change demonstrated through the interface of the model, we traced the model's dynamics to detect emergent behavior. Since this model is demonstrating an abstract city to focus on the urban processes, its inner validity has been examined by comparing the implemented model with its design [40, 44]. Our examination of the model's inner validity showed that it was behaving as expected.

4 Results

To test the methodology and data, an environment with agents was simulated in NetLogo. To examine various aspects and dynamics, the model was tested for 300+ time-steps. The environment of the model is demonstrated as zones that represent the CBD (central circle with the zID = 30), inner-city (middle ring with the zID = 60), and suburbia (outer ring with the zID = 90), and the polygons represent regions that form the neighborhoods (referred to by nIDs). As explained in the Model Design section, professional and non-professional agents of all colors are randomly placed in the inner-city and suburbia (gray background), while the developers are placed fixed in the CBD.

The cap rate, urban growth rate, gentrification rate, shrinkage rate, sprawl moving rate, sprawl density threshold, and initial population ranges appear as sliders to support sensitivity analysis. The developed properties are demonstrated on a monitor. There are six plots on the geographical user interface that display various information such as developers' average savings, developers' NOI, and professional and non-professional agents in the inner-city and in suburbia and their savings. These details are illustrated in Fig. 2.

The model was tested by multiple runs and with various settings for sensitivity analysis. Due to the focus of the paper on the effect of gentrification on density and property value, relevant parameters such as initial population were changed to find the most interesting cases worth describing in the paper. The model demonstrated the most dynamic results when tested with a population of 1000 agents and 20% similarity in the happiness preference setting of professional and non-professional agents which created a segregated environment based on income and color. Setting 30% of the population to randomly prefer to be near CBD increased the number of professionals in the inner-city while displacing the non-professionals and causing a

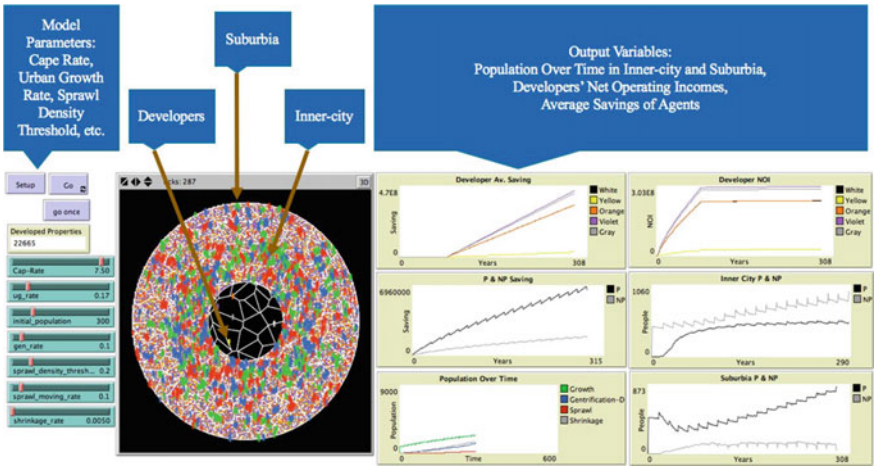


Fig. 2 Model graphical user interface at default settings

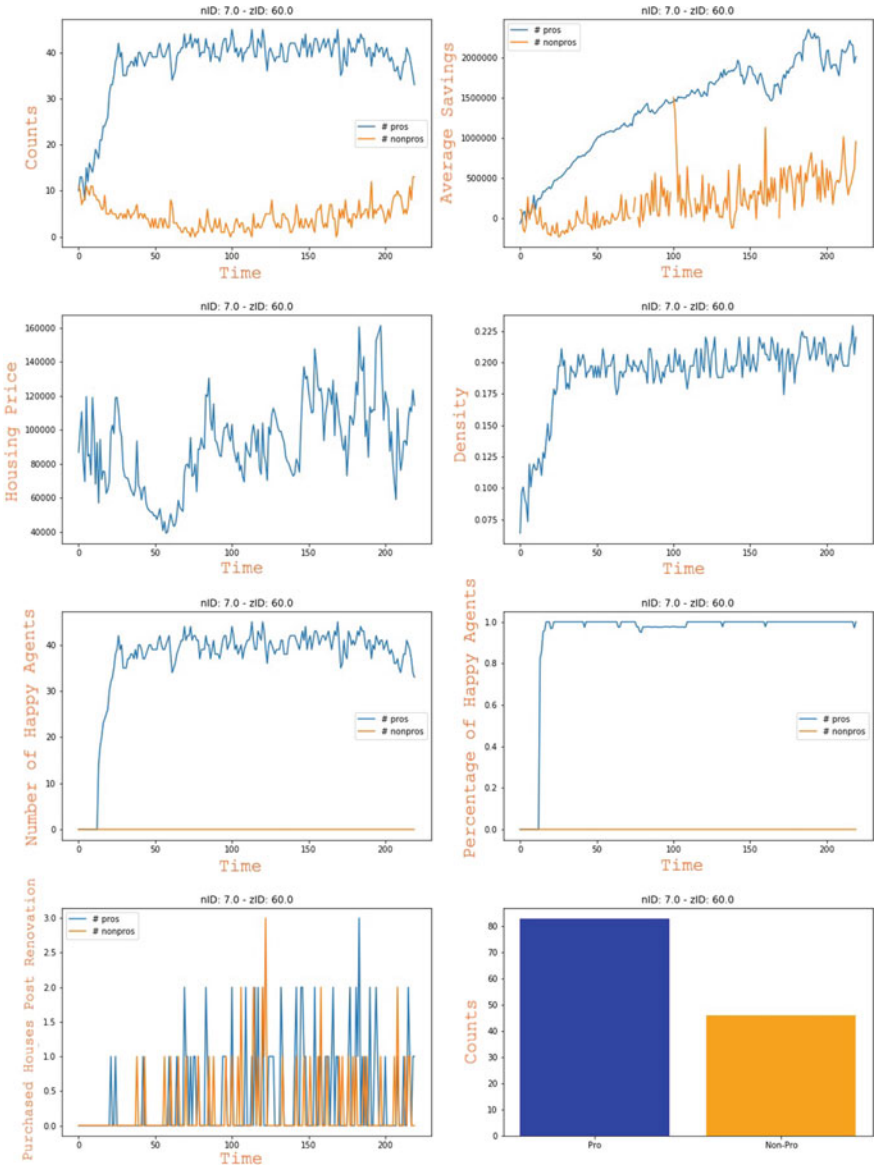


Fig. 3 Gentrification by supply in the 7th neighborhood of the inner-city

longer and more fluctuated movement for them across the landscape. This caused a fluctuation in the population density which affected the property values. The parameters affecting the rent gap along with the developers, attempt to buy, renovate, and sell the properties, creating a fluctuation in the property values and thus the population density. Figures 3 and 4 are representative of the plots gathered to analyze the effect of gentrification on density and property value.

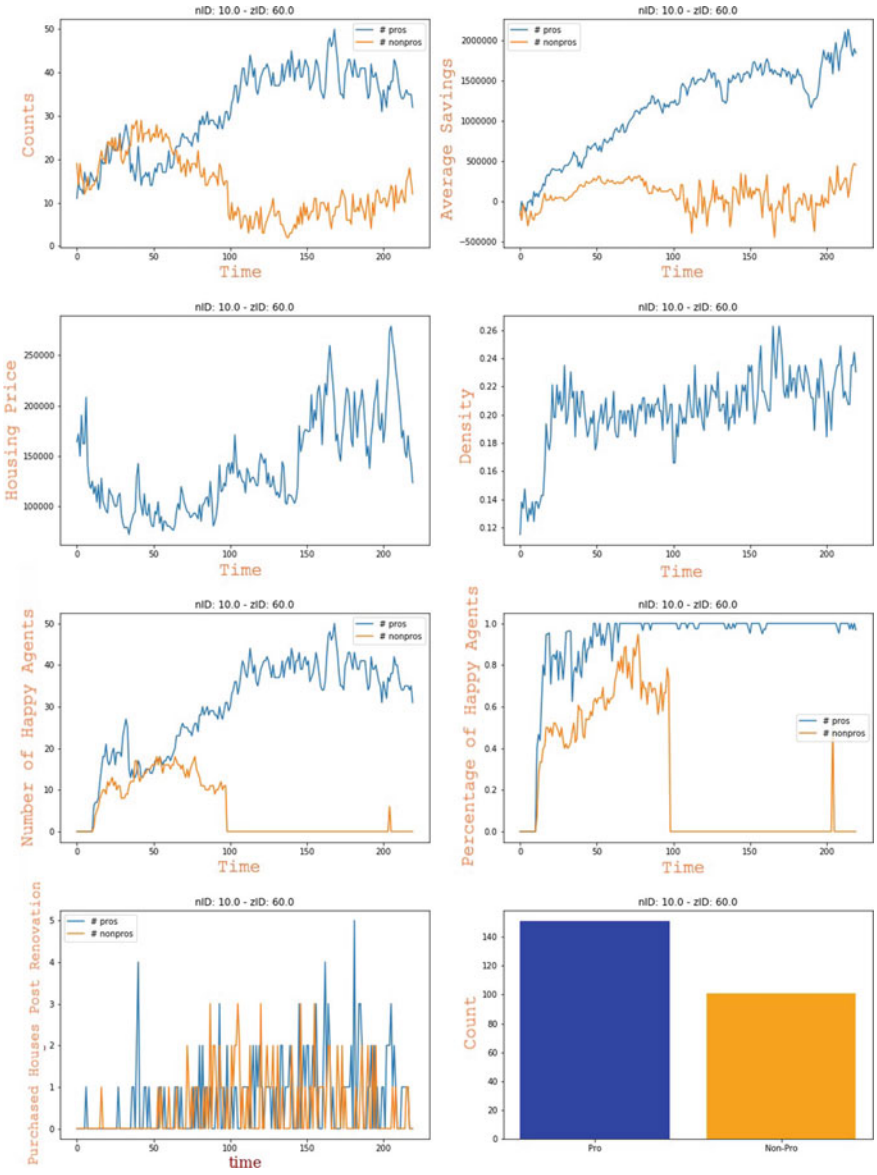


Fig. 4 Gentrification by demand in the 10th neighborhood of the inner-city

4.1 Gentrification by Supply

Figure 3 represents an example of gentrification by supply. The plots in this figure show the results of neighborhood number 7 in the inner-city. The plots demonstrate

the continuous and dominant presence of the professionals with rising savings in this neighborhood. There is a considerable happiness gap among the professional and non-professional agents. Most important, it seems that more professionals bought properties after renovation and remained in the neighborhood. The presence of the professionals is correlated to the population density and the housing price.

4.2 Gentrification by Demand

An example of gentrification by demand is demonstrated in Fig. 4. The plots in this figure show results of neighborhood number 10 in the inner-city. These plots show a fluctuation in dominance between the agent classes, with the professionals succeeding in dominance over the neighborhood. Population density and housing prices correlate with their rising presence. The fluctuation in dominance seems to positively correlate with the happiness of the agents. Professionals also show a higher number of purchases of renovated properties.

5 Discussion and Further Works

5.1 Main Results

The overall findings of the model indicate the dominance of the professional agents in the majority of neighborhoods. While for the purpose of the study, the neighborhood is defined as the tessellation in which an agent resides, similar to Flache et al. [19], the structure and size of the neighborhood does not affect the urban dynamics occurring within it. After 300+ time-steps, the segregation caused by the happiness preference of agents is quite visible. The pattern of segregation by income and color follows notions from Schelling's [45] model.

5.1.1 Gentrification by Supply

The results of Fig. 3 prove the top-down nature of gentrification by supply explained in the Background section. The developers find opportunities during times that the neighborhood is undergoing shrinkage and is left with vacant properties to buy. They renovate and sell these houses, increasing the property value to its potential rent and its age to zero. This act affects the population density depending on the agent class most dominant in the area. Neighborhoods with high housing prices and a large population of professionals tend to thrive in such dynamics with more professionals moving to them. While this increases the density and may result in urban sprawl, it also causes a sense of unhappiness among non-professionals, leaving them

displaced and unable to purchase properties in the neighborhoods to which they formerly belonged. Neighborhoods with high housing prices and a large population of non-professionals tend to cause a locked-in state for this agent class due to having inadequate savings to move. Other non-professionals who are attracted to the neighborhood are mostly unable to purchase a property there. This state results in more professional agents moving into the neighborhoods and buying the recently renovated properties. Renovations in neighborhoods follow a trend of displacement or immobilization for non-professionals while providing a high-quality residential opportunity for professionals.

5.1.2 Gentrification by Demand

Results in Fig. 4 demonstrate an example of the bottom-up structure of gentrification by demand. The preference of the 30% of the agents to not only live near CBD, but pay lower rents, is at the core of this dynamic. Figure 4 shows a neighborhood that is considered affordable to non-professionals but preferred to professionals since the non-professionals residing there mostly have no other option but the professionals who chose to live there can most likely afford a higher-priced property. It is this choice that ultimately results in the increase in population density and thus property value, causing a sense of unhappiness among non-professionals and their displacement to other neighborhoods. While their displacement may result in a temporary shrinkage, the neighborhood will eventually be populated by professionals with the savings to acquire the recently renovated and higher-priced properties. The choice of agents to reside in neighborhoods that are the only option of some others is followed by the displacement of the non-professionals.

One of the utilities of agent-based models is to test existing theories; however, when it comes to urban processes, there are very few theories that encapsulate a large number of such processes. Therefore, we have chosen those which are prototypical to the model. The purpose of this explanatory model is to combine various urban processes and demonstrate the life cycle of a city while focusing on the effect of gentrification on property value. To this point, the model is not constructed to recommend policy solutions, but rather to utilize empirical data in developing urban processes and demonstrating the outcome of their interactions. However, the model can be extended to possibly prevent gentrification or decrease the high rate of unhappiness among agents by performing sensitivity analysis to find the right threshold for various population levels. Further extension of the model focused on other theories and urban processes is left to future researchers to explore.

6 Conclusion

While many models have been developed to explore a single urban process, our model examined multiple processes, namely, urban growth, sprawl, shrinkage, and gentrification to portray their effect on the life cycle of a city. The model focused on the agent–agent and agent–environment interactions that play a role in population density and property value. For the purpose of this paper, we narrowed the analysis down to observing the role of gentrification by supply and demand. We observed in the results that the immobilization and displacement of the lower-income agents caused by developers revitalizing a neighborhood for profit and also by higher-income agents choosing to live near CBD and pay lower rents. The actions of the developers and professionals in the model directly affected the options that non-professionals are left with through the rise of population density and housing prices. Our model demonstrates patterns of increased segregation as a result of the preferences, rules, and interactions occurring between agents. Finally, the model supports the previous research revealing the disabling nature of gentrification for the unprivileged population residing in or native to a neighborhood undergoing development.

References

1. Alonso, W.: Location and Land Use. Toward a General Theory of Land Rent. Harvard University Press, Cambridge (1964)
2. Axtell, R., Epstein, J.: Agent-based modeling: understanding our creations. Bull. Santa Fe Inst. **9**(4), 28–32 (1994)
3. Batty, M.: Rank clocks. Nature **444**(7119), 592–596 (2006)
4. Batty, M., Longley, P.A.: Fractal cities: a geometry of form and function. Academic Press (1994)
5. Batty, M., Xie, Y.: Urban growth using cellular automata models. In: GIS, Spatial Analysis, and Modelling, pp. 151–172. ESRI Press (2005)
6. Batty, M., Xie, Y., Sun, Z.: Dynamics of urban sprawl. In: Working Paper Series, Paper 15. University College London (1999)
7. Benenson, I.: Multi-agent simulations of residential dynamics in the city. Comput. Environ. Urban Syst. **22**(1), 25–42 (1998)
8. Bhatta, B.: Causes and consequences of urban growth and sprawl. In: Analysis of Urban Growth and Sprawl from Remote Sensing Data, pp. 17–36. Springer (2010)
9. Burgess, E.W.: Can neighborhood work have a scientific basis. In: The City, pp. 142–155 (1925)
10. Chin, N.: Unearthing the roots of urban sprawl: a critical analysis of form, function and methodology. In: Centre for Advanced Spatial Analysis University College London. Citeseer (2002)
11. Cioffi-Revilla, C.: Introduction to Computational Social Science: Principles and Applications. Springer (2017)
12. Clarke, K.C., Gazulis, N., Dietzel, C., Goldstein, N.C.: A decade of sleuthing: lessons learned from applications of a cellular automaton land use change model. In: Classics in IJGIS: Twenty Years of the International Journal of Geographical Information Science and Systems, pp. 413–427 (2007)

13. Coiacetto, E.J.: Places shape place shapers? Real estate developers' outlooks concerning community, planning and development differ between places. Plan. Pract. Res. **15**(4), 353–374 (2000)
14. Crooks, A. Agent-based modeling and geographical information systems. In: Geocomputation: A Practical Primer, pp. 63–77. Sage Publications (2015)
15. Crooks, A.T.: Exploring cities using agent-based models and gis. In: Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects. University of Chicago and Argonne National Laboratory (2006)
16. Devisch, O.T.J., Timmermans, H.J.P., Arentze, T.A., Borger, A.W.J.: An agent-based model of residential choice dynamics in nonstationary housing markets. Environ. Plan. A **41**(8), 1997–2013 (2009)
17. Diappi, L., Bolchi, P.: Smith's rent gap theory and local real estate dynamics: a multi-agent model. Comput. Environ. Urban Syst. **32**(1), 6–18 (2008)
18. Economic Policy Institute. Family budget. http://www.epi.org/resources/budget/ (2015). Accessed 2017
19. Flache, A., Hegselmann, R.: Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. J. Artif. Soc. Soc. Simul. **4**(4) (2001)
20. Fujita, M. Urban Economic Theory: Land Use and City Size. Cambridge University Press (1989)
21. Geltner, D., Miller, N.G., Clayton, J., Eichholtz, P.: Commercial Real Estate Analysis and Investments, vol. 1. Thompson South-Western (2007)
22. Gilbert, N., Hawksworth, J.C., Swinney, P.A.: An agent-based model of the english housing market. In AAAI Spring Symposium: Technosocial Predictive Analytics, pp. 30–35 (2009)
23. Glaeser, E.L., Kahn, M.E.: Sprawl and urban growth. In: Handbook of Regional and Urban Economics, pp. 2481–2527. Elsevier (2004)
24. Governing.com. Washington D.C. gentrification maps and data. http://www.governing.com/gov-data/washington-dcgentrification-maps-demographic-data.html (2016). Accessed 2017
25. Gregory, D., Johnston, R., Pratt, G., Watts, M., Whatmore, S. The Dictionary of Human Geography. John Wiley & Sons (2011)
26. Haase, D., Haase, A., Kabisch, N., Kabisch, S., Rink, D.: Actors and factors in land-use simulation: the challenge of urban shrinkage. Environ. Model. Softw. **35**, 92–103 (2012)
27. Harris, C.D., Ullman, E.L.: The nature of cities. Ann. Am. Acad. Polit. Soc. Sci. **242**(1), 7–17 (1945)
28. Heppenstall, A., Malleson, N., Crooks, A.: Space, the final frontier: how good are agent-based models at simulating individuals and space in cities? Systems **4**(1), 9 (2016)
29. Hoyt, H.: The structure and growth of residential neighborhoods in American cities. Government Printing Office, U.S. (1939)
30. Jackson, J., Forest, B., Sengupta, R.: Agent-based simulation of urban residential dynamics and land rent change in a gentrifying area of boston. Trans. GIS **12**(4), 475–491 (2008)
31. Kennedy, W.G.: Modelling human behaviour in agent-based models. In: Agent-Based Models of Geographical Systems, pp. 167–179. Springer (2012)
32. Landis, J.D.: The California urban futures model: a new generation of metropolitan simulation models. Environ. Plan. Plan. Des. **21**(4), 399–420 (1994)
33. Lees, L., Slater, T., Wyly, E.: Gentrification. Routledge (2008)
34. Ludeman, C., Stoffers, T., Mcdonnell, B., Levy, S., Havsy, J., Rice, J.: Cbre cap rate survey. https://www.cbre.us/Research-and-Reports/US-Research (2014). Accessed 2017
35. Magliocca, N.R.: Exploring coupled housing and land market interactions through an economic agent-based model (chalms). In Agent-Based Models of Geographical Systems, pp. 543–568. Springer (2012)
36. Malik, A., Crooks, A., Root, H., Swartz, M.: Exploring creativity and urban development with agent-based modeling. J. Artif. Soc. Soc. Simul. **18**(2), 12 (2015)
37. Mayor Muriel Bowser Office of Planning. D.C. quickfacts. https://planning.dc.gov/publication/dc-quick-facts (2015). Accessed 2017

38. Metropolitan Regional Information System. Housing market data for Washington D.C. http://www.getsmartcharts.com/mls/signup (2016). Accessed 2017
39. Miles, M.E., Notherton, L.M., Schmitz, A.: Real Estate Development: Principles and Process. Urban Land Institute (2015)
40. North, M.J., Macal, C.M.: Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation. Oxford University Press (2007)
41. Otter, H.S., van der Veen, A., de Vriend, H.J.: Abloom: location behaviour, spatial patterns, and agent-based modelling. J. Artif. Soc. Soc. Simul. **4**(4) (2001)
42. Park, R.E., Burgess, E.W., McKenzie, R.D.: The growth of the city: an introduction to a research project. In: The City, pp. 85–97. (1925)
43. Parker, D.C., Filatova, T.: A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. Comput. Environ. Urban Syst. **32**(6), 454–463 (2008)
44. Patel, A., Crooks, A., Koizumi, N.: Slumulation: an agent-based modeling approach to slum formations. J. Artif. Soc. Soc. Simul. **15**(4), 2 (2012)
45. Schelling, T.C.: Dynamic models of segregation. J. Math. Sociol. **1**(2), 143–186 (1971)
46. Schwarz, N., Ernst, A.: Agent-based modeling of the diffusion of environmental innovations—an empirical approach. Technol. Forecast. Soc. Chang. **76**(4), 497–511 (2009)
47. Smith, N.: Toward a theory of gentrification a back to the city movement by capital, not people. J. Am. Plan. Assoc. **45**(4), 538–548 (1979)
48. Torrens, P.M.: Simulating sprawl. Ann. Assoc. Am. Geogr. **96**(2), 248–275 (2006)
49. United States Census Bureau.: District of columbia quickfacts. https://www.census.gov/quickfacts/fact/table/DC/PST045216 (2016). Accessed 2017
50. Urban Land Institute. Professional real estate development: developer's tool kit. https://americas.uli.org/book/professional-real-estate-development-developerstool-kit/ (2012). Accessed 2017.
51. Wallace, N.: Where to buy: price-to-rent ratio in 76 US cities. https://smartasset.com/mortgage/price-torent-ratio-in-us-cities (2016). Accessed 2017
52. Wiechmann, T., Pallagst, K.M.: Urban shrinkage in Germany and the USA: a comparison of transformation patterns and local strategies. Int. J. Urban Reg. Res. **36**(2), 261–280 (2012)

Negative Influence Gradients Lead to Lowered Information Processing Capacity on Social Networks

Nisha Baral, Chathika Gunaratne, Chathura Jayalath, William Rand, Chathurani Senevirathna, and Ivan Garibay

Abstract Communication networks are known to exhibit asymmetric influence structures, constructed of a spectrum from highly influential individuals to highly influenced individuals. Information Processing Capacity (IPC) determines the level of responsiveness expressed by individuals when communicating with others in such networks. In this study, we explore the asymmetric influence structure of GitHub's cryptocurrency developer community and show how it affects the IPC of the users in such networks. We use an agent-based model of information diffusion and conversation based on dynamic individual-level probabilities extracted from data on activity from cryptocurrency-related GitHub repositories. In this model, users that receive notifications from their neighbors at a rate above their IPC enter an overloaded state. We show that users who are influenced substantially more than they influence other users are typically expected to be overloaded and constantly experience lower IPC. In other words, these users are influenced more than they are able to express this magnitude of influence toward their neighbors. These results have potential implications in the design of viral marketing and reducing the harm of misinformation campaigns.

N. Baral · C. Gunaratne · C. Jayalath · C. Senevirathna · I. Garibay (✉)
University of Central Florida, Orlando, FL, USA
e-mail: igaribay@ucf.edu

N. Baral
e-mail: nishabaral@knights.ucf.edu

C. Gunaratne
e-mail: chathika@knights.ucf.edu

C. Jayalath
e-mail: acj@knights.ucf.edu

C. Senevirathna
e-mail: csenevirathna@knights.ucf.edu

W. Rand
North Carolina State University, Raleigh, NC, USA
e-mail: wmrand@ncsu.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_16

Keywords Influence · Information processing capacity · Information overload · GitHub · Cryptocurrency · Information diffusion

1 Introduction

Over the past two decades, individuals have begun to rely on Online Social Media (OSMs) as their primary mode of communication at an exponential rate. Typically, users follow large numbers of profiles and are constantly bombarded with information at a rapid rate. As a result, OSM users become overloaded with information, which will affect their responsiveness to future incoming information, depending on the source, and the actions they may take on the OSM. Classically, information overload can be explained as when: information processing requirements > Information Processing Capacity(IPC) [5]. If there is a rise in incoming information, users become overloaded and cannot respond to all messages received. Recent studies have shown that the responsiveness of OSM users may be temporarily suppressed under an excess of information inflow [7]. It was further shown that this decline in response capability of a cognitively overloaded user obeyed a power-law relationship with the magnitude of the experienced overload. Cognition is a driving force behind human interaction, yet, humans have biological limits to their cognitive abilities [19, 21]. Kang et al. show that social media users have finite attention, hence cannot always respond to all the messages received [15]. For instance, the rate of social media notifications received by an OSM user exceeding the limit of their IPC, could cause the suppression of user responsiveness and cause their IPC to change dynamically.

There is evidence for the "social-brain hypothesis," which claims that cognition determines the social network size of an individual [3, 4, 11, 13, 22]. However, these studies do not suggest detail explanation on how cognition affects their IPC. Yet, IPC is an important factor for information diffusion over such networks. Moreover, we are more interested to explore on social influence network structure for this work. We believe that the user influence structure within these networks plays a crucial role in determining the responsiveness and thereby, the limits to user IPCs. Furthermore, the influence experienced by an individual is not homogeneous. People are influenced by other users at different intensities. The total influence experienced by a user is different from the influence they are able to exert on their neighbors, causing them to experience an influence gradient. In this study, we explore the relationship between the influence gradient experienced by an OSM user and their respective IPC dynamics.

OSMs, similar to any other context of human communication, bring together individuals for the dissemination and discussion of information. Also similar to other human communication structures, individuals influence each other through online communication. This influence, however, is heterogeneous and asymmetric, forming a network of influence with some users being highly influential hubs, while others are highly influenced by the actions of other users. Transfer entropy [20] has been shown to be an effective estimator of influence in human behavior [2, 25]. We

measure marginal transfer entropy (TE) between users of Git Hub's cryptocurrency community, and extract their underlying influence network.

We hypothesize that the influence structure in the extracted network is a determining factor of the information-overloaded state of users. In particular, we test the effect of the influence gradient, the influence exerted minus influence experienced, on IPC. We perform these tests by simulating 2 weeks of activity of the users of the influence network of Git Hub's cryptocurrency community as described above. For this purpose, we develop an agent-based model of information diffusion and conversation, the Multi-Action Cascade Model (MACM) [10], where users take actions based on dynamic endogenous and exogenous probabilities extracted from the user activity data. This influence was modeled from the uncertainty estimated by the transfer entropy values of the network. Users in the MACM maintain a memory of the activity of their neighbors to which they may respond to, the actionable information stack. The length of this stack is adjusted dynamically, according to the information overload experienced by the individual users, simulating IPC.

Our results confirm that the influence gradient experienced by users has a significant effect on their IPC. In particular, users who experience substantially higher influence from their neighbors than they exert on their neighbors are more likely to experience information overload and constantly experience a low IPC. We see from the extracted influence network, that many such users exist, who are influenced more than they influence others, while the users who exert more influence than they experience from their neighbors are rare.

2 Related Work

We use the MACM model [10] which is an agent-based model of information diffusion and conversation based on dynamic individual-level probabilities of activity. Agents incorporated with cognition maintain their memory of received messages and have a stack of actionable information. Our work relates to some of the studies on social media information diffusion that focus on users' limited responsiveness and their effects [1, 6, 8, 12, 14, 15, 17].

Gonçlaves et al. performed a theoretical study on cognitive limits of users using data from Twitter conversations [8] and tested the validity of Dunbar's claim [4] that due to cognition limits, people can maintain a maximum of 100–200 stable relationships and they found agreeable results. Backstrom et al. observed how Facebook users prioritize their attention across the focused group and dispersed group of friends [1]. Another similar study by [12] discovered that, when a tweet's visibility decays over time, users need more cognitive effort to re-tweet that particular tweet. Kang et al. studied social media users' limited attention to incoming information and its effects on the diffusion process through the network [15]. Their model shows that users have non-uniform and finite attention to their friends in the network, and they tend to respond only to some friends, whom they are more attentive. A similar study

on Facebook by [16] made an assumption that the IPC of users is homogeneous, and observed the information diffusion process.

However, we believe that IPC is an important factor to be considered for information diffusion over OSMs and the differences in users' influence gradient within these networks play a crucial role in determining their responsiveness since people experience influence at different levels. The total influence experienced by users varies with the influence they can exert on their neighbors. Hence, the influence experienced by an individual is not homogeneous and therefore the IPC is heterogeneous and not the same for all individuals. In this study, we explore the relationship between the influence gradient experienced by an OSM user and their respective IPC dynamics.

3 Method

We simulate the diffusion of information over a social network by developing the MACM, an agent-based model of endogenous and exogenous influence probabilities extracted from data on activity from cryptocurrency-related GitHub repositories of around 2419 users (139 unique repositories) [10]. The users' probabilities of acting in response to other users are modeled as an adaptive probability, adjusted by an uncertainty estimated by transfer entropy between user activity time series data. We model actions such as initiating repos, contributing to repos, and sharing content actions as conditional probabilities of the rate of these actions as performed by their neighbors. Neighbors were determined through a two-step process as follows.

First, snowball sampling was performed, starting from a seed sample according to the ratios of the prevalence of archetypes previously extracted on this data [18]. Since our available data was too huge and computing was too slow, we used users archetypes to gather the fair sample of seed users to the model by scaling down the users from existing clusters. Second, transfer entropy [20] was used to quantify the influence from each user to the other, and any network edges with a zero (or less than zero) transfer entropy value were dropped. Hence, we got 2419 users and 139 unique repositories in our existing network. We have a total of 9734 edges. In-degrees range from 0 to 573 nodes, while out-degrees range from 0 to 489 nodes. Accordingly, a user V responded to notifications of any neighbor U's activity at a probability $q_{V,U}$ at every simulation tick, where ticks were modeled as minutes. Notifications of neighbor activity were propagated to neighbors in the form of messages. Typically, the rate at which messages were received by a user would exceed the rate of response, therefore, a memory of recently received messages was stored in a last-in-first-out stack, known as the actionable information stack, A. For any given time step, t, $|A_t| <= M_t <= M_{max}$, where M_t is the user's IPC at t, and M_{max}, a parameter of the simulation, represents the information overload threshold [9] of any user in the simulation. It is the size of this memory (M_t) in our model that maps to the notion of *IPC*, or the number of messages that a user can take action on. In the case that, $R_t + |A_t| > M_{max}$, where R_t is the number of messages received in t, a user is

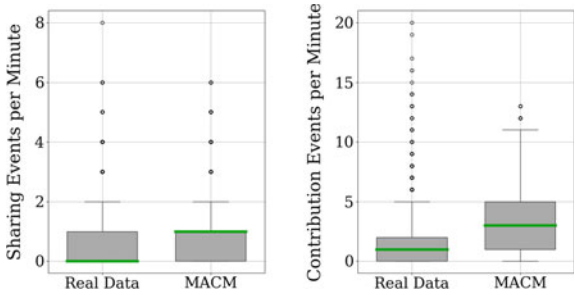


Fig. 1 Comparison between the minute-wise frequency of sharing and contribution GitHub events predicted by MACM simulations against real data

considered to be overloaded. In other words, this user has more messages to respond to than it is currently capable of processing.

An overloaded user adjusts its M_t according to the magnitude of the overload experienced. According to [7], this reduction of cognitive capacity is related to the overload experienced by a power-law relationship. The magnitude of a user's overload O_t is found as

$$O_t = (|A| + R_t) - M_{\max}$$

Taking this into consideration, we calculate the new information capacity (i.e., the IPC), of an overloaded user as

$$M_{t+1} = M_t - O_t^{\alpha} = M_t - ((|A| + R_t) - M_{\max})^{\alpha}$$

where α, a parameter of the model, establishes the power-law relationship between cognitive overload and IPC in [7]. Finally, the oldest messages of A are discarded, such that $|A_{t+1}| = M_{t+1}$.

We calibrated the MACM parameters, α and $M_{\max}$, for the 2-week period from August 1 to August 15, 2017, to minimize the real data to simulation output errors for users' sharing [1] and contribution [2] action frequencies. The model was implemented in NetLogo [26] and calibrated using BehaviorSearch [24], via a genetic algorithm with a population size of 72, mutation rate of 0.03, and cross-over rate of 0.7. The best parameters found were $\alpha = 0.65$ and $M_{\max} = 5$ for GitHub. The IPC-related experiments performed in this study were performed over the prediction period August 15 to September 1, 2017, and use the best parameter choices found in the calibration above. Figure 1 compares the sharing and contribution event frequency distributions of the MACM against that of the real data, which confirms that there is no significant difference between the minute-wise event frequencies between the MACM and the Real Data.

[1] Fork and Watch actions.

[2] Issues, IssueComment, Push, PullRequest, PullRequestReviewComment, CommitComment,Gollum actions and comment.

4 Results

The histogram plots in Fig. 2 compare the distributions of IPC of the users of different in-degrees and out-degrees of influence in GitHub. In this figure, the x-axis represents M_t, or the user's IPC, while the y-axis is the frequency. We categorize users based on their corresponding in-degree and out-degree bins. In-degree determines the influence experienced by the user while their out-degree determines the influence exerted to other users in the network which is obtained from calculating peer influence using transfer entropy. We see that a majority of individuals do not have an influence gradient; show no substantial difference in their in-degree and out-degree of influence. However, there exists a large portion of users with a negative influence gradient, with in-degree of influence substantially larger than their out-degree

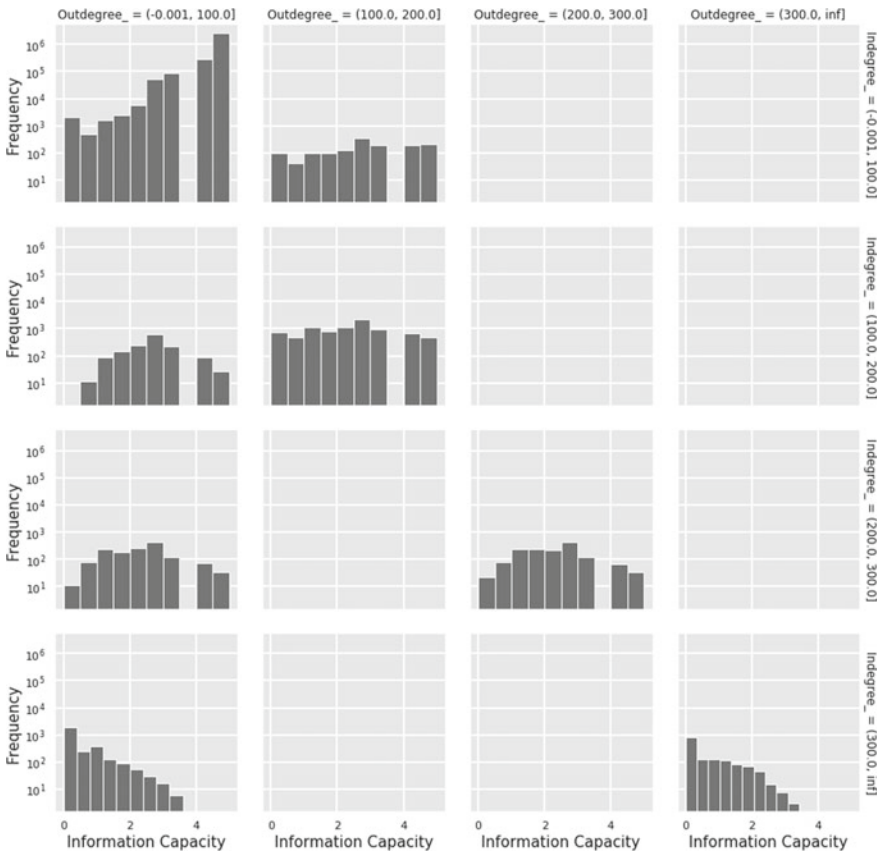


Fig. 2 Distributions of the IPC M_t of users of different in-degrees and out-degrees of influence on GitHub Users. In-degree indicates the influence experienced by the user while out-degree determines the influence exerted to other users

Table 1 p-values non-parametric statistical tests for the alternate hypothesis that there is a significant difference between IPC distributions under varying in-degree and out-degree of influence, as shown in Fig. 2

Tests	In-degree	Out-degree
Two ways KS test	$2.2e^{-}16$	$2.2e^{-}16$
One way welch test	$2.2e^{-}16$	$2.2e^{-}16$
Kruskals wallis test	$2.2e^{-}16$	$2.2e^{-}16$

of influence. Such users demonstrate a strong tendency to function under very low IPC. Only a few users, existed with positive influence gradients and only experienced moderate amounts of overload. Users that have both extremely high in-degree and out-degree of influence values also demonstrate a similar tendency of functioning with very low IPC. Meanwhile, users who have both mid-ranged in-degree and out-degree of influence values have a near-normally distributed IPC and typically function with a mid-range IPC. In contrast, users with both very low in-degree and very low out-degree of influence typically function at full capacity.

We performed non-parametric statistical tests to test the alternate hypothesis that there is a difference in the IPC over varying in-degree and out-degree of influence as shown in Fig. 2. In other words, we can say that OSM users experience the effects on their IPC with variations on the amount of influence they experience and the amount of influence they can exert on others. Non-parametric tests were chosen as the IPC proved to be heteroskedastic over increasing in-degree and out-degree values. In particular, we performed 2-way KS Tests, One-Way Welch Tests, and Kruskal–Wallis Tests to test the validity of the alternative hypotheses that in-degree of influence has an effect on IPC and that out-degree of influence has an effect on IPC. The results for these tests are summarized in Tab. 1. We reject the corresponding null hypotheses, that there is no difference in IPC over varying in-degree and out-degree of influence, with a 0.05 significance, as the p-values for both in-degree and out-degree are less than 0.05.

Next, we performed a two-way KS test, to validate our alternate hypothesis that both in-degree and out-degree of influence have effects on IPC. In Table 2, we see that across different in-degree and out-degree groups, there are higher D values for most of the in-degree and out-degree bins. Observing these results, we can say that there is significant variation in IPC between the people across the different network sizes. Hence, we can reject the null hypothesis that there is no effect on IPC due to the varying size of in-degree and out-degree of influence.

We test our hypothesis that the influence gradient, out-degree of influence minus in-degree of influence (i.e., influence exerted minus influence experienced), has a determining effect on user IPC, in Fig. 3, where the x-axis represents the bins for difference between out-degree to in-degree values and the y-axis represents the number of messages a user can process which is referred as the IPC. The maximum IPC for GitHub simulations was set to 5 as we get this value after calibrating the model using behavior search. There is no intersection between the interquartile ranges of

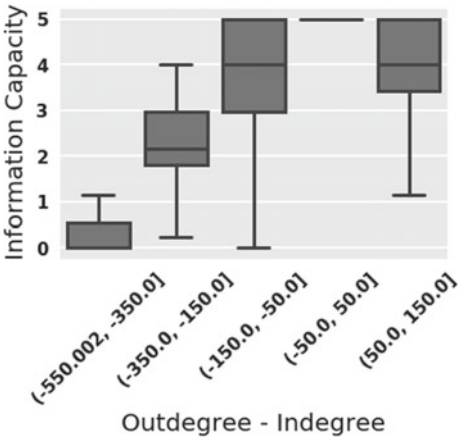


Fig. 3 IPC of users relative to the influence exerted to influence experienced grouped by influence network structure in GitHub

Table 2 Two sample KS test In-degree and Out-degree of influence versus IPC

Network Type	Bin(100–200)	Bin(200–300)	Bin(300-inf)
Bin(200–300)	*	*	$D_{\text{in}} = 0.73565$ $D_{\text{out}} = 0.70139$ $P_{\text{value}} = 2.2e^{-}16$
Bin(100–200)	*	$D_{\text{in}} = 0.091518$ $D_{\text{out}} = 0.11339$ $P_{\text{value}} = 2.2e^{-}16$	$D_{\text{in}} = 0.69812$ $D_{\text{out}} = 0.65565$ $P_{\text{value}} = 2.2e^{-}16$
Bin(0.001–100)	$D_{\text{in}} = 0.79866$ $D_{\text{out}} = 0.76846$ $P_{\text{value}} = 2.2e^{-}16$	$D_{\text{in}} = 0.8509$ $D_{\text{out}} = 0.85061$ $P_{\text{value}} = 2.2e^{-}16$	$D_{\text{in}} = 0.97748$ $D_{\text{out}} = 0.97602$ $P_{\text{value}} = 2.2e^{-}16$

*Data is symmetric

IPCs of users with very large, negative influence gradients, compared to the distribution of IPC of users with smaller, negative influence gradients. Users with positive influence gradients, in-degrees of influence equal to or lower than their out-degrees of influence, typically showed no signs of overloading. Overall, for users with negative influence gradients, IPC was shown to be correlated to influence gradient. We validate this result by performing a two-sample KS test between the IPC of the bins in Fig. 3. The results of these KS tests are shown in Table 3, and fail to falsify our claim that the influence gradient is a determining factor of IPC.

Table 3 Two Sample KS Test on relative influence exerted to influence experienced Vs IPC

Network Type	Bin(-150.0,-50.0)	Bin(-350.0,-250)	Bin(-550.001,-450.0)
Bin(-350.0,-250)	*	*	$D_{out-in} = 0.76111$ $P_{value} = 2.2e^{-16}$
Bin(-150.0,-50.0)	*	$D_{out-in} = 0.52708$ $P_{value} = 2.2e^{-16}$	$D_{out-in} = 0.90683$ $P_{value} = 2.2e^{-16}$
Bin(-50.0, 50.0)	$D_{out-in} = 0.53843$ $P_{value} = 2.2e^{-16}$	$D_{out-in} = 0.87205$ $P_{value} = 2.2e^{-16}$	$D_{out-in} = 0.97988$ $P_{value} = 2.2e^{-16}$

* Data is symmetric

5 Discussion

We show the relationship between the difference of influence exerted and influence experienced and the IPC of users in OSM. We expect to see similar effects on other OSM networks like Twitter and Reddit too. We will test this in our next experiment. Reference [4] showed there is a limit on a maximum number of social connections people can handle, discovering a limit to around 100–200 connections, approximately. The Dunbar limit has been confirmed on OSMs as well [8]. Our results show the Dunbar limit is an average and instead there is heterogeneity in these limits, as we find some individuals are able to influence more individuals than others may. This diversity of influence causes users to experience influence gradients. We examine the effects of these influence gradients on the IPC of users. The maximum number of connections has no relationship to the ability to process the information on OSM.

From our study, we see that people are overloaded when they are influenced more than they can influence others. This means that information can not be propagated further from such users who are already overloaded since they have low IPC. This leads them to be less influential since they are unable to process information. Our results shed light on the most important and least important person to process information in the network. It means that given the influence gradient of OSM, it is possible to have a good understanding of a person's ability to process messages by observing their information-overloaded state which we can determine from their IPC. This study also hints that we can identify the most influenced and influential users in the network and their information overload state as well. This can be useful in influence campaigns since determining the influential users and their capacity to disseminate the information is very important in such cases. We are able to show that IPC is heterogeneous over the population of users and changes dynamically with neighbor activity.

6 Conclusion

We build a model that matches real-world social media diffusion and show that there is an effect on IPC due to varying influence gradients. We show that users with higher in-degree and lower out-degree of influence are probable to be overloaded since they are more influenced by other people than they are capable to influence other people in the network. This information has several implications.

For instance, it means that researchers who are studying these networks can use network structure as a proxy for users' IPC. These findings can inform viral marketing strategies [23] when attempting to determine optimal seeding and other methods to support information diffusion. We see that it is insufficient to consider highly influential users for vial marketing seeding. Instead, it is essential to consider the influence gradient experienced by the seed users. According to our results, users who are currently exposed to less influential sources and exert more influence on their followers would be more effective at spreading marketing messages, without being at risk of slowing down the propagation process after experiencing information overload. From these results, we can say that broadcasting information from multiple sources does not mean that information will process faster. In contrast, multiple sources of influence would lead to negative influence gradients and leave users more susceptible to information overload and lower IPC, hampering information propagation.

In addition to promoting the spread of information, these findings can also be used to inhibit the spread of information. For instance, if an intelligence operator was interested in preventing certain information from spreading online, key users could be flooded with alternative information, essentially overloading these key users and preventing the spread of misinformation and propaganda. According to our results, users with large negative influence gradients are more viable targets for this process, as they are more easily overloaded.

Acknowledgements We thank Leidos for providing data and DARPA SocialSim grant (FA8650-18-C-7823) for funding us to perform this study.

References

1. Backstrom, L., Bakshy, E., Kleinberg, J.M., Lento, T.M., Rosenn, I.: Center of attention: how facebook users allocate attention across friends. ICWSM **11**, 23 (2011)
2. Bauer, T.L., Colbaugh, R., Glass, K., Schnizlein, D.: Use of transfer entropy to infer relationships from behavior. In: Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop, p. 35 (2013)
3. Bickart, K.C., Hollenbeck, M.C., Barrett, L.F., Dickerson, B.C.: Intrinsic amygdala-cortical functional connectivity predicts social network size in humans. J. Neurosci. **32**(42), 14729–14741 (2012)
4. Dunbar, R.I.: The social brain hypothesis. Evol. Anthropol.: Issues, News, Rev.: Issues, News, Rev. **6**(5), 178–190 (1998)

5. Eppler, M.J., Mengis, J.: The concept of information overload: a review of literature from organization science, accounting, marketing, mis, and related disciplines. Inf. Soc. **20**(5), 325–344 (2004)
6. Feng, L., Hu, Y., Li, B., Stanley, H.E., Havlin, S., Braunstein, L.A.: Competing for attention in social media under information overload conditions. PloS One **10**(7), e0126090 (2015)
7. Gomez-Rodriguez, M., Gummadi, K.P., Schoelkopf, B.: Quantifying information overload in social media and its impact on social contagions. In: ICWSM, pp. 170–179 (2014)
8. Gonçalves, B., Perra, N., Vespignani, A.: Modeling users' activity on twitter networks: validation of dunbar's number. PloS One **6**(8), e22656 (2011)
9. Gunaratne, C., Baral, N., Rand, W., Garibay, I., Jayalath, C., Senevirathna, C.: The effects of information overload on online conversation dynamics. Comput. Math. Organ. Theory **26**(2), 255–276 (2020). https://doi.org/10.1007/s10588-020-09314-9
10. Gunaratne, C., Senevirathna, C., Jayalath, C., Baral, N., Rand, W., Garibay, I.: A multi-action cascade model of conversation. In: 5th International Conference on Computational Social Science (2019)
11. Hill, R.A., Dunbar, R.I.: Social network size in humans. Hum. Nat. **14**(1), 53–72 (2003)
12. Hodas, N.O., Lerman, K.: How visibility and divided attention constrain social contagion. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), pp. 249–257 (2012)
13. Kanai, R., Bahrami, B., Roylance, R., Rees, G.: Online social network size is reflected in human brain structure. Proc. R. Soc. B: Biol. Sci. **279**(1732), 1327–1334 (2011)
14. Kang, J.-H., Lerman, K.: La-CTR: a limited attention collaborative topic regression for social media. In: Twenty-Seventh AAAI Conference on Artificial Intelligence (2013)
15. Kang, J.-H., Lerman, K., Getoor, L.: La-IDA: a limited attention topic model for social recommendation. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 211–220 (2013)
16. Li, P., Sun, Y.: Modeling and performance analysis of information diffusion under information overload in facebook-like social networks. Int. J. Commun. Syst. **27**(9), 1268–1288 (2014)
17. Miritello, G., Lara, R., Cebrian, M., Moro, E.: Limited communication capacity unveils strategies for human interaction. Sci. Rep. **3**, 1950 (2013)
18. Saadat, S., Gunaratne, C., Baral, N., Sukthankar, G., Garibay, I.: Initializing agent-based models with clustering archetypes. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp. 233–239 (2018)
19. Sarter, M., Gehring, W.J., Kozak, R.: More attention must be paid: the neurobiology of attentional effort. Brain Res. Rev. **51**(2), 145–160 (2006)
20. Schreiber, T.: Measuring information transfer. Phys. Rev. Lett. **85**(2), 461 (2000)
21. Smit, A.S., Eling, P.A., Coenen, A.M.: Mental effort causes vigilance decrease due to resource depletion. Acta Psychol. **115**(1), 35–42 (2004)
22. Stiller, J., Dunbar, R.I.: Perspective-taking and memory capacity predict social network size. Soc. Netw. **29**(1), 93–104 (2007)
23. Stonedahl, F., Rand, W., Wilensky, U.: Evolving viral marketing strategies. In: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, pp. 1195–1202 (2010)
24. Stonedahl, F., Wilensky, U.: Behaviorsearch [computer software]. In: Northwestern University, Evanston, IL, Center for Connected Learning and Computer Based Modeling (2010). http://www.behaviorsearch.org
25. Ver Steeg, G., Galstyan, A.: Information transfer in social media. In: Proceedings of the 21st International Conference on World Wide Web, pp. 509–518 (2012)
26. Wilensky, U., et al.: Center for connected learning and computer-based modeling. In: Northwestern University, Netlogo (1999)

A Complex-Systems Agenda for Influencing Policy Studies

Paul K. Davis, Tim McDonald, Ann Pendleton-Jullian, Angela O'Mahony, and Osonde Osoba

Abstract This paper describes a possible agenda for changing the teaching and conduct of policy analysis to better reflect the centrality of complex adaptive systems and the potential role of new analytic methods, including computational social science (CSS). The agenda identifies desirable changes in four categories: (1) world view when conceiving and posing problems, (2) the basis for reasoning and inference, (3) analytic style, and (4) the character of models and model-based analysis. The paper's intent is to obtain reactions and suggestions from others in the CSS community to help refine the ideas.

1 Introduction

1.1 Purpose

It has been 30-some years since awareness of complex adaptive systems (CAS) burgeoned as the Sante Fe Institute [1]. The realm of policy studies, however, has not yet reflected adequately the new insights. We are rethinking such matters at the Pardee RAND Graduate School, which is transforming its approach to teaching PhD students of policy analysis [2]. This paper, a shorter version of an earlier RAND

P. K. Davis (✉) · T. McDonald · A. Pendleton-Jullian · A. O'Mahony · O. Osoba
Pardee RAND Graduate School, 1776 Main Street, Santa Monica, CA 9047-2138, USA
e-mail: pdavis@rand.org

T. McDonald
e-mail: tmcdonal@rand.org

A. Pendleton-Jullian
e-mail: apendlet@rand.org

A. O'Mahony
e-mail: Aomahon1@rand.org

O. Osoba
e-mail: oosoba@rand.org

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_17

Working Paper, and a related paper to appear in the Journal of Policy and Complex Systems, presents the CSS community with a tentative set of complexity-related items on which changes in teaching and practice seem important.

Many aspects of policy studies need not change: tried-and-true methods and mindsets will often be apt, but we focus on where change is warranted, to include making better use of CSS. We also have suggestions for CSS practitioners. We hope that responses to this paper will assist in the next iteration of our work.

We define out terms for discussing CAS in Sect. 1.2; in Sects. 2–5 we then consider issues in four categories (1) worldview, (2) basis for reasoning and inference, (3) analytic style, and (4) character of models and model-based analysis.

1.2 Our Definitions

We use the term complex adaptive system (CAS) to refer to systems with significant nonlinear interactions among adaptive components that lead to emergent macroscopic properties that are uniquely characteristic of macroscopic levels (e.g., the wetness of water or a sense of community) and that are not describable in the terms associated with more microscopic levels. Familiar examples include the human body and social systems. We see CAS as a subset of complex systems more generally, i.e., systems comprised of parts with nonlinear interactions and corresponding higher level properties. Table 1 summarizes the way we distinguish among simple, complicated, complex, and complex adaptive systems.

1.3 Additional Discussion of Definitions

Our distinctions differ from others in the literature. For example, we argue that most systems (not just CAS) are "more than their parts," which is why they are called systems rather than collections, and that the higher level properties of many such systems (not just CAS) are nonlinear functions of component characteristics and interactions (i.e., they are not sums of parts). An automobile is not an automobile without its structural, propulsion, and steering subsystems: the automobile's functionality is zero if any of these are absent, a nonlinear relationship. Further, we see various phenomena in inanimate systems as emergent. For example, we see the Bénard cells of non-equilibrium chemistry [4] and the superfluidity of liquid Helium as emergent phenomena. Complexity-related effects, including chaos, occur in some relatively small systems (e.g., 4-species Lotka-Volterra systems [5] and in Chua's electrical circuit [6]). As for complicated versus complex, we reserve "complex" for use in systems with subtle nonlinear interactions (typically dynamic).

The definition of "complicated" is fuzzy because of matters of degree and because natural-language use of "complicated" depends on expertise and perceptions: using vector algebra in physics is complicated or simplifying depending on familiarity with

Table 1 Distinctions among types of system

Characterization	Numbers	Nonlinearities	Adaptive agents	Emergence	Examples
Simple system	Small	No	No	No	A printer-computer system with plug-and-play
Complicated system	Larger	Yes	No	No	Troublesome furniture, to be assembled at home; intricate assemblages
Complex system	Usually large	Yes	No	Yes (due to inanimate features)	Modern automobile, aircraft; dissipative systems in chemistry; army divisions used rigidly
Complex adaptive system (CAS)	Usually large (but with some exceptions)	Yes	Yes	Yes (due to animate agents)	Human body, social organizations, nations

vectors. As another example, parents may be aghast at how complicated it is to put together a gift on Christmas Eve (many parts, many shapes and connections; poor directions). So also, the remote control of television, speaker, and DVD player is "complicated" (even if explained in manuals).

Finally, the term "emergence." Significant disagreements exist on both the semantics of emergence and the phenomena itself, as demonstrated in a panel discussion [7]). As merely one part of this, some researchers distinguish between weak emergence (the vast bulk of examples) and a postulated strong emergence for which the whole has properties inexplicable from even precise knowledge of the parts and their interactions (consciousness is a proposed example). Some interesting reviews exist in the philosophical literature on the history, issues, and controversies about emergence [8] and the related topic of mereology, the study of the relation of parts to whole [9]. A recent book discusses challenges of emergence for simulation [10] (see particularly Chap. 5 [11]).

2 Worldview

The worldview under which inquiry occurs is crucial. We touch on issues related to system thinking, confronting complexity in CAS, seeing decision-making as a

continuing process, values and objectives, knowledge and uncertainty, and the control possible over systems.

2.1 System Thinking

Our first admonition is not really tentative: it is about seeing the world in "system" terms. That idea was stressed with the advent of systems analysis and policy analysis [12–14]. Oddly, the field of System Dynamics developed in parallel [15–17] with these, with surprisingly limited interactions. In any case, the profound significance of system thinking is often not visible in modern-day policy studies. It should be—to better understand the issues and also to understand the mechanisms of how to bring about real and sustainable change. System thinking includes taking care to include everything that needs to be included. These may be parts, relationships, or processes. It also includes paying close attention to the system's environment and other aspects of context. Doing so is aided by consciously trying different frames for viewing the problem, as in using material, social, and mental lenses as discussed in Chap. 2 of [18].

Transformational change often is the emergent result of changes initiated at microscopic levels where human beings have agency—i.e., have the capacity to influence the system. Those changes interact with other aspects of the system, including top-down and mid-level policies and constraints. Interestingly, transformational changes in systems often occur as the result of sequential increments—although change can sometimes be rapid—and are as much or more the result of changed agent behaviors within the system as of changes in system structure. Encouraging useful transformations requires fearlessness along the way—designing, doing, learning, adapting, and often trying new ideas. One aspect of this is designing for emergence as described in Chap. 11 of [18]. As an example, General Stanley McChrystal's transformation of the Joint Special Operations Center (JSOC) during the war in Afghanistan is described in detail in [19] (pp. 290–331). The changes reframed the challenge as being no longer a fight for body count (a focus in Vietnam) but a fight for information. The resulting change in frame led to a strategic principle of "It takes a network to defeat a network" [20].

2.2 Confronting Complexity and Wicked Problems

More than restoring the system view is necessary, however, because policy analysis is mostly about social systems, which are usually complex adaptive systems (CAS). Seldom does current policy analysis represent well the emergent behaviors important in CAS thinking. Econometric analysis, for example, highlights elasticities in a stable system. Even System Dynamics with its feedback loops does not ordinarily represent

agent behaviors directly. Classically, it is top-down in nature. With exceptions, agent-based modeling is usually more microscopic and does not connect well with the type of system thinking needed by policy-makers. Relatively few papers connect the domains [21].

Recognizing systems and adaptation is still not enough. The problems in policy analysis are not like those of physics and engineering: instead, they are often wicked problems (defined as difficult to define and inherently unsolvable [22]). When wicked problems are successfully addressed, stakeholders often agree on a course of action rather than a broader philosophy—i.e., on compromise actions that will be good enough even though inconsistent with earlier intentions. This is the stuff of soft system methodology (SSM) or soft OR, the classic papers of which remain essential reading [3, 22, 23], as are later works [24, 25]. Interesting comparisons have been made across methods [26], but—in our view—practice in the United States has lagged in dealing with soft system issues. Closely related matters, however, have been championed influentially in the language of learning organizations [27], system thinking for social change [28], and "dancing with systems" [29] (Chap. 7, from 2001).

2.3 Decision-Making as a Continuing and Messy Process

Policy analysis has often proceeded as though a single decision is to be made, perhaps with relatively minor adjustments from time to time. Many social systems, however, change over time—perhaps with abrupt and unexpected shifts. Major course corrections may be necessary and management of CAS should be conceived accordingly. Some might relate this to the incrementalist approach of "muddling through" [30, 31], but we relate it to planning for adaptiveness. This can include scheduling branch-point decisions and building capabilities that allow responses to unpredictable developments [32–34]. It can include comparing adaptive strategies as a function of possible developments and a measure of regret, which can be defined in significantly different ways [35] (pp. 33–34).

2.4 The Values and Objectives that Drive Decisions

Utilities are core concepts in traditional policy analysis, part of a unifying theme [36]. Stakeholders, however, often do not know their utility functions and no such utility function may even exist. After all, the concept of utility function depends on the stability of values, but values change as the result of experience, events, and challenges. Negotiations can cause evolutions in each side's appreciation of other perspectives. A corporation may come to value the quality of life of its employees and union members may come to value the commercial success of their company.

Nations may come to value strategies that are at once "tough" for deterrent purposes but also not threatening to other countries.

Values inform the goals set when making policy choices. Some goals are broad (e.g., better health for Americans), whereas others are more specific (e.g., making prescription drugs more affordable). Conceiving goals in multiple levels is necessary—to see both forest and trees and to establish concrete objectives at different levels of detail, as in DoD's strategies-to-task approach [37, 38], and to understand the origin of emergent properties. A difficulty is that collections of clearly stated, seemingly unobjectionable value-laden goals can often harbor deep conflicts or unresolvable trade-offs (e.g., data privacy vs. data utility, procedural fairness vs. equality of outcomes). Where such value conflicts occur, they often raise the meta-issue of adjudicating among conflicting accepted goals.

Values are reflected in objectives. In classic policy analysis, utilities and utility functions are core concepts and the objective may be seen as a composite utility to be optimized. In a world of complex problems, objectives are multiple, often in tension and often controversial; compromise leads to goal-setting that attempts to balance (not "optimize") across considerations.

2.5 Knowledge, Uncertainty, and Disagreement

Policy analysis often proceeds as though much is known and only a few items are uncertain enough to worry about explicitly. This is simply wrong. The uncertainties in policy problems are often ubiquitous, large, and not amenable to best-estimate analysis. This was first noted in defense planning [32, 39], in a technologist's paper offering new ideas about reforming policy analysis [40], and in climate-change studies that led to powerful new methods [41, 42]. A core concept from the latter is work is that of deep uncertainty, the condition in which analysts do not know or the parties to a decision cannot agree upon (1) the appropriate models to describe interactions among a system's variables, (2) the probability distributions to represent uncertainty about key parameters in the models, or (3) how to value the desirability of alternative outcomes. Confronting deep uncertainty drastically alters one's world view and the way one goes about policy analysis, in both defense work [39] and social-policy studies [43, 44].

2.6 Moderation in the Search for Control

Past generations of policy analysts have been trained to think in terms of top-down interventions, such as by national governments. For those sensitive to the character of CAS, the way ahead will often be seen as requiring a balance among top-down, bottom-up, and other influences [18]. For example, to improve the US. education system, some initiatives make sense at the federal level, but most should focus on the

state and local levels where most decisions are made and where local contexts are understood.[1] Perhaps most important, the concept of control should be approached with humility because consequences of interventions will often be surprising.[2]

Although the need for monitoring and feedback has long been recognized [45], implementation is often seen as a separate and perhaps less prestigious subject by policy analysts. If, however, we expect to be surprised and dismayed by some of the consequences from interventions, we should be vigilant in watching for troublesome indicators and prompt in adapting efforts appropriately. Implementation then becomes part of sequential decision-making. National governments are notoriously ill-suited for this, so it will probably require more decentralization to states, communities, non-government organizations (NGOs), and commercial companies.

3 Basis for Reasoning and Inference

3.1 Theory, Data, Association, and Causation

Discussing the basis for reasoning and inference highlights a long-standing schism between those who study policy problems with the statistical tools of social and health sciences and those who are more dependent on theory-building, as in assessing what military capabilities will be needed for future wars or how to change incentive structures to bring about changes in the health system. The schism can be described, roughly, as data-driven versus theory-driven inquiry and is described elsewhere (e.g., p. 23 of [46]). Alternatively, it can be seen as focused on associations rather than causal reasoning.[3]

Much current policy analysis for social problems is based on data-driven econometrics and other forms of statistical analysis. An example is evidence-based decision-making, as when the efficacy of health treatments is judged by data from randomized control trials (RCTs).

This is in contrast with policy analysis that considers alternative future scenarios when evaluating options. Future military force structures are evaluated for simulated conflicts thought to measure what is needed for deterrence. Similarly, corporations use scenario planning in strategic planning [27, 50]. Empirical data plays only a

[1] This discussion and other ideas in the paper have been much influenced by colleague Gery Ryan, an anthropologist who observes how changes actually occur in social systems.

[2] A tension exists between attempting to use knowledge when influencing a system and instead responding heuristically to information (e.g., price signals) in ways that have proven effective without pretense of fully understanding the system's mechanisms. This tension is illustrated by disputes between Keynes and Hayek in economics.

[3] Data-driven researchers value causality, but they assess it in terms of what explains past data accurately, whereas the causal modeling we have in mind applies also to circumstances not yet observed and futures not yet experienced [47]. The subject of causality is deep and subtle, with many connotations [48, 49].

limited role in such work because empirical data on future wars or the consumer-product battlefield of 2040 is, let us say, sparse.

Both data-driven and theory-driven work have their place, but the ideal is a hybrid that includes theory-informed data analysis and theoretical work that more effectively draws on empirical information. Making that kind of hybrid activity easier is a significant challenge for social-science modelers, technologists, and scientists [46]. An important part of meeting that challenge is CAS-sensitive teaching of causal theory, albeit uncertainty-sensitive theory as discussed later.

The need for causal theory transcends our concern about CAS. Correlational data has distinct limitations and many questions cannot even be asked without causal models of the phenomena in question. These include questions about the effects of alternative interventions or other policy actions [47, 51]. So also, two problems exist with the evidence-based focus in policy analysis, even without considering complexity. First, the effect may be to stifle innovation that involves trying new things or common-sense cause-effect reasoning for which rigorous data does not exist. Second, empirical data from the past is sometimes misleading as we are all painfully aware as "once-in-a-century" storms occur annually. A better conception for the use of evidence is evidence-informed work that draws also on other research, appreciation of local context, and a deep understanding of the actual problem at hand as discussed in a critique of past efforts to promote evidence-based practices [52].

Returning now to the theme of doing better in dealing with CAS, the need for increased emphasis on causal system modeling should be evident because we seldom have adequate data to understand fully the effects of interventions in social systems or systems of systems in which humans and organizations have big roles. We need theory to anticipate possible instabilities, possible "bifurcations," and so on. Theory is also necessary to sort out the complications of ambiguous cause-effect directionality, which is common in complex systems.

3.2 Types of Theory

Theory is important, but another striking reasoning-and-inference difference across disciplines is the connotation of "theory." Especially in the social sciences myriad definitions exist and are intermixed, causing severe confusion [53]. One social-science definition is that theory is "a systematic set of interrelated statements intended to explain some aspect of social life" [54]. Other times, the word refers merely to discrete hypotheses or speculations. In everyday usage, the word is often used disparagingly, as in "Well, in theory such and such should be true, but—of course—we all know that….". This is in contrast with such disciplines as physics in which "theory" usually refers to an integrative and coherent set of principles by which

to understand knowledge in a broad domain.[4] Interestingly, context is as crucial in physics as in social science, but in physics the variables defining context are specified and the cases distinguished sharply. Doing so is more difficult (but not impossible) when context includes effects of culture, history, and personal experience.

The varied usages will continue to serve different but worthy purposes. However, we see special need for increasingly good theories of the integrated and coherent variety. CSS researchers should aspire to contribute to their development. This will mean constructing increasingly integrative theories to test in simulation for their relative ability to explain data, rather than comparing results across simulations based on simplistic theories.

A tradeoff exists between maximizing a kind of rigor when testing a narrow theory in a narrow context and instead of seeking or testing broader theory for which some data is softer and less controlled. In our view, policy analysis needs relatively more emphasis on the broader constructs and system thinking. This, we believe, will increase the accuracy and relevance of conclusions to inform policy decisions, although sacrificing the rigorous statistical precision demanded in some disciplinary work.[5]

4 Analytic Style

4.1 Analysis and Reductionism, Both Good and Bad

We use the term "analysis" to mean "a detailed examination of anything complex in order to understand its nature or to determine its essential features: a thorough study" (Merriam-Webster online dictionary). In this meaning, analysis is not just about decomposition, despite the Latin roots of the word.

We see "reductionism" as important in studying complex systems, as when Herbert Simon described the human body as a nearly decomposable system [59] with different levels (cellular, tissue, organ) and subsystems (circulatory, digestive,…). Similarly, the system engineer must understand the components of the system being designed, particularly their functions and relationships to other components, to create a working

[4] Abraham Kaplan saw theory more expansively as "a device for interpreting, criticizing, and unifying established laws, modifying them to fit data unanticipated in their formulation, and guiding the enterprise of discovering new and more powerful generalizations." [55] (p. 295).

[5] As an example, research has shown no correlation between states' increases in minimum wage and the loss of low-paying jobs [56, 57]. That, however, does not justify raising minimum wages. What policymakers need is a theory that describes the circumstances under which raising the minimum wage would have positive or negative consequences. Such theory might predict good consequences when profits are good and wage growth has lagged productivity growth, whereas it might predict loss of low-paying jobs in other circumstances, perhaps with a time lag [58]. The data may not exist to validate such a better theory empirically, but it is difficult to have reasoned discussions and debate without one.

system. Reductionism, then, can be good—even when working with CAS. In saying that, however, we use the terms in a broader sense than do others.

This said, it is sometimes better to focus on the whole and interactions at the level of the whole, rather than seeking explanation from below. Some social phenomena (e.g., how to diminish schoolyard bullying) are perhaps best understood without going into individual characteristics and behaviors. Similarly, some scientists work at the thermodynamic level without discussing molecular phenomena, even though thermodynamics can in principle be understood in terms of more microscopic phenomena. As yet another example, in the world of investing and perhaps even in the realm of national economic policy-making, it is sometimes more pragmatic to react to system-level signals (e.g., trends in prices, wages, and employment) than to attempt to make decisions based on analysis of poorly understood microscopic causal relationships.

It follows that some good analysis may be about going broader, deeper, or both. As a related matter, good analysis—even for CAS—will sometimes be more reductionist and sometimes more holistic. Ideally, reductionist and holistic thinking are complementary.

4.2 Character of Analysis: Quantitative Versus Qualitative, and Matters of Rigor

Much was made in early years about the importance of quantitative analysis. This was an antidote to emotional and sometimes irrational or parochial arguments. It was thought to be necessary for rigor. The emphasis on quantitative work, however, had negative side effects because many important variables are inherently imprecise and difficult to measure (e.g., love, hate, loyalty). As noted early [7], ignoring soft variables is to assert that they have no effect.

A related point is that qualitative considerations often are profoundly important in defining the context in which issues arise and policies are considered. What solutions are acceptable, for example, depends on the local culture and on history. Does the proposed solution fit a narrative that resonates in the community?

The larger point, for which we do not yet have adequate language, is to reinterpret the concept of rigor to mean something more like logical, coherent, and the result of considering all relevant information. Given the negative connotations of rigor (formulaic, austere, strict), perhaps we would be better off referring to sound reasoning, but the word rigor will not go away. The rigor we seek is not necessarily quantitative, much less precise. Interestingly, rigor in our sense is demanded in philosophy, which is hardly known for its numerical calculations or data. Well-structured qualitative treatment will sometimes be appropriate, as when dealing with CAS when behavior can be difficult or impossible to predict. Identifying approximate circumstances of instability is more rigorous, i.e., more sound than purporting to predict precisely when instabilities will manifest or, what forms the instability will take [60].

4.3 Confronting Uncertainty and Disagreement

System analysts and policy analysts have long noted the need for uncertainty analysis [61], but in reality it has been given short shrift. The methods available have now improved dramatically, in part due to work that introduced using probabilistic methods with subjective probabilities [62]. Nonetheless, policy analysis often ignores important uncertainties. For those of us who see CAS everywhere, that is a critical shortcoming. We should encourage analysis that addresses the many kinds of uncertainty from the very outset and as a matter of first-order attention. The result of policy analysis should often be identifying strategies that are flexible, adaptive, and robust (FARness)—with no pretense of optimization. It has been argued that assisting policy-makers in finding FARness should be an ethical responsibility of analysts [39]. In different language, we should urge robust decision-making—i.e., seeking strategies that are robust in the broadest sense of that over-loaded term [44].

This search for robust strategies may seem obvious but consider how CSS practitioners often think about their simulation experiments. Do they develop their inputs and models so that, from the outset, all the important inputs can be changed readily, and so that model uncertainty can also be addressed [60, 63]? Or, instead, do they hold nearly everything constant and focus only on variation of a few parameters (as often works well in engineering)? Doing better in this regard will be challenging with current CSS infrastructure. Technological innovation will be necessary. Part of this may involve adapting the benefits of certain high-level languages or tools that make uncertainty analysis routine (Analytica® and Crystal Ball® are examples). We observe that CSS modelers often have remarkably little knowledge of or interest in the higher level languages that have proven their mettle in decision-aiding. The other natural approach involves scenario-generation apparatuses for ensemble approaches. The basic ideas were discussed in a paper on computer-assisted reasoning nearly 20 years ago [64]. The technology has advanced since then, but the ideas are still apt. With supercomputing, such use of ensemble methods is possible in near-real time [65].

Another concept is important in meeting the challenges. This is the concept of multi-resolution modeling (MRM). Given determination to do so, it is often possible for a modeling group to have an imperfect family of models at different levels of resolution (whether as a single program with switches or a family of related models) [66]. It is then possible to do initial exploration at a high level (low resolution) (with perhaps 3–10 variables) and to then zoom into detail only where it is worthwhile to do so. This approach has been used successfully [39], but great opportunities exist for doing it better. For an organization, it requires going beyond the mindset of believing that its "big" model is sufficient.

Interestingly, and despite common impressions to the contrary, many policy-makers are comfortable with planning under uncertainty. They value simulation-based analysis (especially in interactive settings) for helping them to understand connections and possibilities, and to view their problem area coherently. They have no illusions about the simulations being reliably predictive [67, 68].

4.4 Comparing Options

Early systems analysis highlighted cost–benefit and cost-effectiveness analysis [61]. Later, some policy analysis came to emphasize multicriteria policy scorecards [69] and to disparage "adding things up" because an option's utility is often not a linear sum. Regrettably, all too many studies (and government-mandated approaches) employ more formulaic cost-effectiveness methods.

For those sensitive to CAS issues, it seems fundamental to distinguish among qualitatively different criteria and to recognize that the system behaviors of interest are related in nonlinear and sometimes incommensurate ways. It should be routine to emphasize these matters and to show conclusions as a function of what are elsewhere called "perspectives" [39]. Doing so may require effort. In the words of Adam Smith in 1759, still relevant centuries later [70] (part III, 2nd paragraph).

> We can never survey our own sentiments and motives, we can never form any judgment concerning them, unless we remove ourselves, as it were, from our own natural station, and endeavor to view them as a certain distance from us. But we can do this in no other way than by endeavoring to view them with the eyes of other people, or as other people are likely to view them… We endeavor to examine our own conduct as we imagine any other fair and impartial spectator would examine for it.

Setting the goals for a given approach is, of course, crucial but not straightforward. Some goals may be dictated by scientific or engineering considerations, but other goals will necessarily require compromise and partial arbitrariness.

4.5 Changing the Questions Asked of Simulation

The standard approach to simulation-based analysis is to observe simulation results and to then to pose "What if?" questions. With each "What if?" a new simulation can be run, although that may require new data, changes in the simulation program, and other time-consuming activities. In that case, policy-makers become impatient, annoyed, and disgusted. Such dissatisfaction led in recent years to the Department of Defense disbanding a large analytic group that used complex campaign models routinely [71]. The modelers of the CSS community should regard that development as an ominous example of how policy-makers sometimes view analysis based on big, complex computer models.

The way ahead, when dealing with complex systems is to think from the outset in terms of addressing broader, "Beyond-what-if" questions. Years ago, that phrase was introduced in the context of an approach to simulation that exploited simulations with PROLOG-style inference engines that would answer questions (subject to the validity of the simulation) such as "Under what circumstances will we achieve our objective adequately?" [72].

In today's world, we can answer such questions with massive computational experiments and subsequent search for patterns. This is referred to as "scenario discovery"

in the toolkit associated with robust decision-making (RDM) [42, 44]. In any case, we can now aspire toward analysis generating "region plots" and other mechanisms for preemptively addressing policymaker questions [60].

5 Character of Models and Model-Based Analysis

5.1 Different Classes of Model

Social science and much policy analysis have been dominated by statistical models and methods, but we see the need for much more extensive use of causal models of various types. One important type is agent-based computational simulation, some of which has been used in dissertation work in policy studies. More is needed as policy analysis better confronts CAS issues. Many different modeling approaches, however, are needed [73] as discussed in a recent book [74]. Also, research often benefits from using a combination of simple, mid-level, and detailed causal models; human gaming; and different approaches to all of these [39] (e.g., pp. 22–25). This might seem to be covered already by what is called mixed-methods research, but much such research makes little or no use of causal modeling.

5.2 Purpose of Models and Related Issues of Validity

Although the point may seem obvious, and has been made many times by many authors, models have many different purposes—in research generally, and in policy analysis specifically (the focus of this paper). It follows that model validity should be considered separately for those different purposes. This is well discussed in the system dynamics literature [17] but is often not appreciated. Recent work [46, 75] urges that validity of models be viewed along dimensions of (1) description, (2) explanation, (3) postdiction, (4) exploratory analysis and coarse prediction, and (5) prediction. Many models are good for some but not others of these.

What do explain and predict mean in the present context? In the machine-learning community and much of the world of statistical analysis, "explanation" refers to the ability of a model to generate estimates that accord well with data, as measured by something like a correlation coefficient R^2. Good explanation in this sense may refer only to the quality of data fitting. If the model can then produce results close to that of new data, it is said to be predictive. Neither of these meanings are suitable for policy analysts when talking with a policy-maker. "Explanatory power" should then refer to the degree to which the model provides the cause-effect relationships that make the results understandable and actionable. Further, saying to a policymaker that a model is predictive should refer to future situations, not just past situations on which empirical data exists. Again, this often requires a causal model.

We should also comment briefly on internal, external, and measurement validity, core concepts in social science and its scholarly journals. Addressing CAS issues of with cause-effect models and major uncertainties will require different attitudes on these. It will be more important to get major phenomena roughly right (e.g., anticipating possible negative consequences of intervention) than to fit empirical data precisely. Measurement error will also be a bigger problem because some of the important variables are hard to define and even harder to measure. But omitting them, or pre-tending that they are represented adequately by more conveniently measured proxies, will be counterproductive. As merely one among myriad examples, GDP per capita is a poor proxy for understanding the economic well-being of the diverse people in a country. Measurement error is even more troublesome when outcomes important to stakeholders depend on context, history, and tacit cultural norms.

5.3 Measures of Outcome

Typically, the measure of outcome used in decision theory and policy analysis is the alleged "expected value" of some measure of interest. The word "expected" is referring to the mean result of a calculation, which may or may not accord with what is likely to happen. In any case, the larger problem here is that policy analysis has not paid adequate attention to distributional results. For example, we should care not only whether the economy improves on average (e.g., as measured by gross national product per capita), but also on how the benefits of that growth are distributed across segments of the population. Arguably, the focus on expected value has contributed to policies that have so disadvantaged the middle class as to be among the reasons for the political rebellions being felt in the United States, Great Britain, and elsewhere. As a related matter, the belief that "a rising tide lifts all boats" ignores the impact of the increased income disparity that often accompanies economic development.

To those thinking about the implication of CAS for policy analysis, the need to go beyond expected value is probably even more evident. It is the nature of CAS to exhibit distributions at a given time and fluctuations, even sizable ones, over time. Furthermore, the very act of designating outcome metrics will motivate the system's self-interested adaptive agents to "game the metrics" by modifying their behaviors, sometimes in undesirable ways. This has been called Goodhart's Law, although it has many roots. Finding metrics resistant to such behavior becomes crucial. Such metrics can often be natural consequences of multi-level causal modeling [39].

5.4 *Predict and Act Versus Anticipate Possibilities and Act Adaptively*

To end our main discussion, we note that when intervening in a CAS it is nearly certain that actions taken will have both good and bad side effects, some of them unanticipated. It will often be necessary for policy to adapt substantially as real developments play out. If so, models should be constructed and data monitored accordingly. For analysts, this implies the need to monitor and update model structures, as well as parameter values. Sometimes it will mean more emphasis on data-driven modeling. We will need to monitor model correctness with signposts to indicate when relatively drastic changes are needed, something akin to ideas proposed long ago with respect to uncertainty-sensitive planning and assumptions-based planning [34, 76]. It is also consistent with some of the newer approaches to analysis associated with terms like "dynamic adaptive pathways" [77].

6 Conclusions

6.1 *Implications for Teaching Policy Studies*

The concepts, theories, and methods close to the heart of those engaged in CSS have many implications for policy studies. Table 2 summarizes the paper compactly. For each of the items discussed above, Table 2 suggests that an agenda for policy studies should favor moving the relative emphasis from where it is now (the 2d column) toward something more CAS-informed (the 3d column). The idea is indicated schematically in Fig. 1 by depicting current and future schoolhouse coverage of the spectrum from simple problems to CAS-sensitive thinking, using a 0 to 10 scale. Topics covered in the past remain important, but we see the need to extend coverage to the right—not to the extent of dwelling on complete chaos and related mathematics, but enough to appreciate the special issues of dealing with and managing complex adaptive systems (which includes keeping the systems away from chaotic regions).

6.2 *Implications for Computational Social Science*

We have presented ideas about changing the study of policy analysis for reactions from the CSS community. However, we also have suggestions for the practice of CSS. These relate to constructing models, infrastructure for exploration under uncertainty, extracting insights, and proving more informative to policy-makers, scientists, and other experts. The opportunity exists for many forward-leaning developments. A number of related efforts have recently appeared (e.g., [78–80]).

Table 2 Shifting the Balance in Policy Studies (a Straw Man for Discussion)

Issue	As seen now	CAS-informed perspective
Worldview		
Nature of system	Simple or complicated, perhaps complex but well behaved	Complex, adaptive systems, often not well behaved or predictable
Problems	Well posed, solvable	Wicked; solutions, if possible, to be emergent
Stability of system and decisions	Stable system, once-and-for-all decisions	Evolving system with continuing need for decisions to deal with developments
Values	Known, fixed, simple	To be discovered, complex, conflicting, changing
Objectives	Utility-based, as in a social welfare function	Multiple, often in tension; the best option balances considerations (including robustness of outcomes), but does not purport to optimize
Knowledge and uncertainty	Good, with some uncertainties	Deeply uncertain
Control	Top-down; direct; confident	Top-down, bottom-up, sideways; humble; indirect; iterative
Basis for reasoning and inference		
Basis for inference	Data, correlation	Causal theory
Types of theory	Simple concepts motivating discrete hypotheses predicting phenomena in narrow contexts	Integrative and coherent set of principles describing classes of phenomena across many contexts (with explicit contextual distinctions)
Analytic style		
Character of "good" theory	Parsimonious in fitting empirical data accurately; dependent only on measurable variables	Rich enough to include important factors, even if uncertain and hard to measure
Uncertainty analysis	On margin as add-on	From outset, broad exploration
Basis for comparing options	Cost-effectiveness, cost–benefit	Multicriteria scoreboards; net effectiveness by perspective
Character of models and model-based analysis		
Type	Statistical	Diverse causal models, games, and other methods (system dynamics, network, agent-based,…; as well as empirical methods)
Purpose	Explain and predict data in statistical sense (for stable systems)	Describe, explain (in causal terms), retrosopectively predict, explore, and predict

(continued)

Table 2 (continued)

Issue	As seen now	CAS-informed perspective
Explanation and prediction	Meaningful correlations (R^2) with old and new data	Causal explanation and prediction even as system and circumstances change

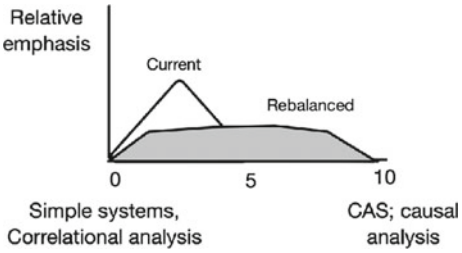


Fig. 1 Shifting the balance of what is taught

Acknowledgements A thank colleagues contributing insights to this: Gery Ryan, Chris Pernin, Chris Nelson, Ben Nyblade, and Rob Lempert. We also thank the anonymous reviewers.

References

1. Waldrop, M.M.: Complexity: The Emerging Science at the Edge of Order and Chaos. Simon & Schuster, New York (1992)
2. Marquis, S.: The Pardee RAND Redesign: Reinventing the Policy. Ph.D. (2019). https://www.prgs.edu/degree-program/redesign.html
3. Checkland, P.: Systems Thinking, Systems Practice: Includes a 30-Year Retrospective. John Wiley & Sons, Chichester, England (1999)
4. Nicolis, G., Prigogine, I.: Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations. John Wiley & Sons (1977)
5. Vano, J.A.: Chaos in low-dimensional Lotka–Volterra models of competition. Nonlinearity **19**, (2006)
6. Matsumoto, T.: A chaotic attractor from chua's circuit. IEEE Trans. Circ. Syst. **31**(12), 1055–1058 (1984)
7. O'Mahony, A., Davis, P.K.: Panel discussion: moving social-behavioral modeling forward. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems. John Wiley & Sons, Hoboken, N.J. (2019)
8. O'Connor, T., Wong, H.Y.: "Emergent properties." In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University (2015)
9. Varzi, A.: "Mereology" In: Calta, E.N. (ed.) The Stanford Encylopedia of Philosophy. The Metaphysical Research Lab. Stanford University, Stanford, California (2019)
10. Mittal, S., Diallo, S., Tolk, A.: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach. John Wiley & Sons (2018).
11. Tolk, A., Diallo, S., Mittal, S.: Complex systems engineering and the challenge of emergence. In: Mittal, S., Diallo, S., Tolk, A. (eds.) Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 78–97. John Wiley & Sons, Hoboken, N.J. (2018)
12. Systems Analysis and Policy Planning: Applications for Defense. Elseveier Science Publishers, New York (1968)

13. Quade, E.S., Carter, G.M. (eds.): Analysis for Public Decisions, 3d edn. North Holland Publishing Company, New York (1989)
14. Walker, W.E.: Policy analysis: a systematic approach to supporting policymaking in the public sector. J. Multi-Criteria Decis. Anal. **9**, 11–27 (2000)
15. Forrester, J.W.: Principles of Systems. MIT Press, Cambridge, MA (1968)
16. Forrester, J.W.: World Dynamics. Wright-Allen Press, Cambridge, MA (1971)
17. Sterman, J.D.: Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill, Boston (2000)
18. Pendleton-Jullian, A., Brown, J.S.: Design Unbound: Designing for Emergence in a White Water World, vol. 1. MIT Press, Cambridge, Mass (2018)
19. Pendleton-Jullian, A., Brown, J.S.: Design Unbound, Volume 2: Ecologies of Change. MIT Press, Cambridge, MA (2018)
20. Arquilla, J., Ronfeldt, D.: The advent of netwar (revisited). In: Arquilla, J., Ronfeldt, D. (eds.) Networks and Netwars. RAND, Santa Monica, CA (1996)
21. Rahmandad, H., Sterman, J.D.: Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. Manage. Sci. **54**(5), 998–1014 (2008)
22. Rittell, H.W.J., Webber, M.M.: Dilemmas in a general theory of planning. Policy Sci. **4**(3), 155–169 (1973)
23. Churchman, C.W.: Prediction and Optimal Decision: Philosophical Issues of a Science of Values. Prentice Hall, Englewood Cliffs, NJ (1961)
24. Ackoff, R.L.: Ackoff's Best: His Classic Writings on Management. John Wiley & Sons, New York (2008)
25. Rosenhead, J., Mingers, J. (eds.): Rational analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity, Uncertainty and Conflict, 2nd edn. John Wiley & Sons, New York (2002)
26. Forrester, J.W.: System dynamics, systems thinking and soft operations research. Syst. Dyn. Rev. **10**(2) (1994)
27. Schoemaker, P.: Scenario planning: a tool for strategic thinking. Sloan Manage. Rev. **36**(2), 25–40 (1995)
28. Stroh, D.P.: Systems thinking for social change. Chelsea Green, White River Junction, Vermont (2015)
29. Meadows, D.H., Wright, D.: Thinking in Systems: A Primer. Earthscan, London, Sterling Va (2008)
30. Lindblom, C.E.: The science of "muddling through." Public Adm. Rev. **19**(Spring), 79–88 (1959)
31. Lindblom, C.E.: Still muddling, not yet through. Public Adm. Rev. 222–233 (1979)
32. Davis, P.K.: Institutionalizing planning for adaptiveness. In: Davis, P.K. (ed.) New Challenges in Defense Planning: Rethinking How Much is Enough, pp. 73–100. RAND Corporation, Santa Monica, Calif. (1994)
33. Davis, P.K.: Strategic planning amidst massive uncertainty in complex adaptive systems: the case of defense planning. In: Minai, A.A., Bar-Yam, Y. (eds.) Unifying Themes in Complex Systems. Springer, Berlin (2006)
34. Davis, P.K.: Uncertainty sensitive planning. In: Johnson, S., Libicki, M., Treverton, G. (eds.) New Challenges, New Tools for Defense Decision Making, pp. 131–155. RAND Corporation, Santa Monica, Calif. (2003)
35. Groves, D.G.: Robust Decision Making (RDM): Application to water planning and climate policy. In: Marchau, V.A.W.J. (eds.) Decision Making Under Deep Uncertainty. Springer, Cham, Switzerland, pp. 135–163 (2019)
36. Nyblade, B., O'Mahony, A., Sieck, K.: State of social and behavioral science theories. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems, pp. 63–99. John Wiley & Sons, Hoboken, N.J. (2019)
37. Kent, G.A., Simons, W.E.: Objective-based planning. In: Davis, P.K. (ed.) New Challenges in Defense Planning: Rethinking How Much is Enough, pp. 59–72. RAND Corporation, Santa Monica, California (1994)

38. Rhodes, C., Hagen, J., Westergren, M.A.: Strategies-to-tasks Framework for Planning and Executing Intelligence, Surveillance, and Reconnaissance (ISR) Operations. RAND Corporation, Santa Monica California (2007)
39. Davis, P.K.: Analysis to Inform Defense Planning Despite Austerity. RAND Corporation, Santa Monica, California (2014)
40. Bankes, S.C.: Exploratory modeling for policy analysis. Oper. Res. **41**(3), 435–449 (1993)
41. Lempert, R.J., Popper, S.W., Bankes, S.C.: Shaping the Next One Hundred Years: New Methods For Quantitative Long-Term Policy Analysis. RAND Corporation, Santa Monica, Calif (2003)
42. Groves, D.G., Lempert, R.J.: A new analytic method for finding policy-relevant scenarios. Glob. Environ. Chang. **17**(1), 78–85 (2007)
43. Lempert, R.J.: A general analytic method for generating robust strategies and narrative scenarios. Manage. Sci. **4**(April), 514–528 (2006)
44. Decision Making Under Deep Uncertainty: From Theory to Practice. Springer, Cham, Switzerland (2019).
45. Cyert, R.M., March, J.G.A.: Behavioral Theory of the Firm, 2nd edn. (1st ed, 1963). Blackwell Business, Cambridge, Mass (1992)
46. Davis, P.K., O'Mahony, A.: Improving social-behavioral modeling. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems, pp. 15–48. John Wiley & Sons, Hoboken, N.J. (2019)
47. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge, Mass (2009)
48. Cartwright, N.: Causation: one word, many things. Philos. Sci. **71**(5) (2004)
49. Halpern, J.Y.: Actual Causality. The MIT Press, Cambridge, Mass (2016)
50. Schwartz, P.: The art of the long view: Planning for the future in an uncertain world. Doubleday, New York (1995)
51. Pearl, J., Mackenzie, D.: The book of why: The new science of cause and effect. Basic Books, New York (2018)
52. Brooks, J.: Making the case for evidence-based decision-making. Stanford Soc. Innov. Rev. **December 6** (2016)
53. Abend, G.: The meaning of "theory." Sociol Theory **26**(2), 173–199 (2008)
54. Rubin, A., Babbie, E.: Research Methods for Social Work, 7th edn. Broooks-Cole, New York (2009)
55. Kaplan, A.: The Conduct of Inquiey. Chandler Publishing Company, San Francisco, CA (1964)
56. Card, D., Kreuger, A.B.: Minimum wages and employment: a case study of the fast food industry in New Jersey and Pennsylvania (1993)
57. Cengiz, D.: The effect of minimum wages on low-wage jobs: Evidence from the United States using a bunching estimator. NBER Working Paper **25434** (2019)
58. Ritholtz, B.: Labor market is doing fine with higher minimum wages. Bloomberg Opin. **January 24** (2019)
59. Simon, H.A.: The Sciences of the Artificial. The MIT Press, Cambridge, Mass (1996)
60. Davis, P.K.: Lessons on decision aiding for social-behavioral modeling. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems. John Wiley & Sons, Hoboken, N.J. (2019)
61. Miser, H.J., Quade, E.S. (eds.): Handbook of Systems Analysis. North Holland Publishing Company, New York (1988)
62. Morgan, M.G., Henrion, M.: Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York (1992)
63. Davis, P.K., Popper, S.W.: Confronting model uncertainty in policy analysis for complex systems: What policy makers should demand. J. Policy Complex Syst. (forthcoming)
64. Bankes, S.C., Lempert, R., Popper, S.W.: Computer-assisted reasoning. Comput. Sci. Eng. **3**(2), 71–77 (2001)
65. Groves, D.G.: Using high-performance computing to support water resource planning, a workshop demonstration of real-time analytic facilitation for the Colorado river basin. Santa Monica. RAND Corporation and Livermore National Laboratore, Santa Monica California (2016)

66. Davis, P.K., Bigelow, J.H.: Experiments in multiresolution modeling (MRM). RAND Corp., Santa Monica, California (1998)
67. Rouse, W.B.: Modeling and Visualization of Complex Systems and Enterprises. John Wiley & Sons, Hoboken, N.J (2015)
68. Rouse, W.B.: Human-centered design of model-based decision support for policy and investment decisions. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems. Wiley, Hoboken, NJ, US, pp. 798–808 (2019)
69. Goeller, B.F., et al.: Policy analysis of water management for the Netherlands, summary report. RAND Corp., Santa Monica, California (1983)
70. Smith, A.: The Theory of Moral Sentiments, 6th edn. Dover, Mineola, NY (1790)
71. Davis, P.K.: Capabilities for joint analysis in the department of defense: rethinking support for strategic analysis. RAND Corp., Santa Monica, California (2016)
72. Rothenberg, J., Narain, S.: The RAND advanced simulation language project's declarative modeling formalism (DMOD). RAND Corp., Santa Monica, Calif (1994)
73. Cioffi-Revilla, C.: Introduction to Computational Social Science: Principles and Applications. Springer, London (2014)
74. Page, S. E. "The model thinker: what you need to know to make data work for you." (2018)
75. Davis, P.K.: Priority Challenges for Social-Behavioral Research and its Modeling. RAND, Santa Monica, California (2018)
76. Dewar, J: Assumption Based Planning. Cambridge, London (2003)
77. Haasnoot, M.: Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. Glob. Environ. Change **23**(2), 485–498 (2013)
78. Yilmaz, L., Ören, T.: Agent-Directed Simulation and Systems Engineering. Wiley-VCH (2009).
79. Yilmaz, L.: Toward self-aware models as cognitive adaptive instruments for social and behavioral modeling. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems, pp. 569–586. Wiley & Sons, Hoboken, NJ (2019)
80. Garibay, I.: The agent-based modeling canvas: a modeling lingua franca for computational social science. In: Davis, P.K., O'Mahony, A., Pfautz, J. (eds.) Social-Behavioral Modeling for Complex Systems. Wiley and Sons, Hoboken, NJ, pp. 521–544 (2019)

Wealth Dynamics in the Presence of Network Structure and Primitive Cooperation

Rajesh Venkatachalapathy, Stephen Davies, and William Nehrboss

Abstract We study wealth accumulation dynamics in a population of heterogeneously mixed agents with a capacity for a certain primitive form of cooperation enabled by static network structures. Despite their simplicity, the stochastic dynamics generate inequalities in wealth reminiscent of real-world social systems even in a fully mixed population. A simple form of cooperation is introduced and is shown to enhance the viability of agents by embedding such dynamics in a network; the impact of social structures on the origins and persistence of inequality can be teased out easily. The models developed here complement traditional modeling approaches based on grid worlds.

1 Introduction

In recent years, the concerns and debates regarding wealth inequality and socioeconomic mobility have been one of the few unifying issues dominating the extremely polarized public spheres of the Global North. While economic and political inequality used to be discussed in heterodox economics circles, the contentious discussions on the topic within mainstream economics since the publication of Piketty's book [23] suggest a lack of consensus about basic foundational questions like the origins and persistence of economic inequality. Not surprisingly, traditional theories and tools of macro and micro-economics are now being diagnosed for their limitations. Simultaneously, insights from related disciplines, along with novel models for scientific inquiry not usually associated with traditional econometrics, are being taken more

R. Venkatachalapathy
Portland State University, Portland, USA
e-mail: venkatr@pdx.edu

S. Davies (✉)
University of Mary Washington, Fredericksburg, USA
e-mail: stephen@umw.edu

W. Nehrboss
Lake Anna Homeschool, Bumpass, USA

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_18

seriously. The work presented here shares this spirit by integrating substantive ideas from anthropology, economic sociology [4] and urban sociology [27] and using modeling approaches from computational social science [13], computational economics [30] and analytical sociology [7] to explore the origins of inequality in a simple model of wealth dynamics in the presence of social structures.

The current work originated in our attempts to incorporate and tease out the effects of social structures in simple models of wealth dynamics in the presence of environmental stochasticity, and a simple form of resource pooling. Specifically, we are interested in studying how structure and dynamics interact to produce and sustain inequality and to influence robustness to scarcity shocks. We answer the former question using Gini coefficients and the latter using survival time type analysis. As we discuss in Sect. 3, while Gini was found to be of limited value, survival time analysis produced more insights on both questions of inequality and robustness.

The model presented here was inspired by our search for analogs of the grid world agent-based model (ABM) of Friesen and Mudigonda [20] where foraging agents that pool their resources were shown to on-average outperform non-pooling agents. This model used foraging to mix the population and create opportunities for interaction, and, when certain conditions were met, resource sharing. Our model achieves agent interaction by postulating a static network structure which partially mixes the agents. Of primary interest is whether network effects alone can generate and sustain differences in economic outcomes of otherwise homogeneous agents.

The Friesen and Mudigonda model drew its inspiration from historical sociology, in Katz's influential study of middle of nineteenth century Hamilton, Canada [12]. Retaining this original motivation, we draw additional inspiration from economic sociology in the work of Granovetter [5]; urban sociology, in the work of Sampson [28] and others; and in anthropology, in the work on cooperation in small to medium scale societies [6, 33], and others. The geographic and economic scale of the systems and the nature of social actors and time scale of interest in these different disciplinary approaches are all very different from the ones used to develop models of representative agents in macroeconomics [2], making the similarity between macroeconomic wealth dynamics models and our models not comparable without further justification. Elaborating on the interplay of conceptual and methodological ideas among these disciplines is beyond the scope of this article. Instead, we anchor this work in economic sociology and revisit the insights from the above disciplines in the article's concluding section.

The important role played by social structures in determining economic outcomes of individuals in a society is not in doubt [5, 10]. Still, in the absence of a unifying foundation for sociology and economics, the full impact of this two-way interpenetration of economic and social structures is demonstrated only on a case-by-case basis. The language of social and economic networks affords a first principle integration by simplifying the non-trivial concept of social and economic structure [19] to only dyadic (binary) relations among actors.[1]

[1] The concept social structures [19] is more general than the widely held belief in computational social science that social structure *is* just a social network.

Models of wealth accumulation dynamics in both macro and micro-economics typically assume the presence of markets and equilibria. Hence, such models are ill-suited to the study of collective phenomena at an intermediate level. Hence, alternative explanations of aggregate phenomena that match the expressiveness of economic models are required. Analytical sociology [8], with its emphasis on explanation of collective emergent phenomena using mathematically formulated social mechanisms [9, 17] anchored at the individual level, is an ideal candidate for this purpose. It is particularly powerful when combined with computational simulation, because when mathematically formulated models reach even a modest level of complexity, they often become analytically intractable. Simulation *in silico* can yield approximate results for these more complex scenarios, which supplement the exact results reached by analytically tractable models.

Models constructed here, like those used in scientific inquiry in general, serve a specific purpose. In this work, we construct simple toy models to reproduce certain aspects of non-trivial wealth inequality distributions in the presence and absence of primitive forms of cooperation, clearly delineating the role of network structure in generating wealth inequality. We make no suggestions that these models *explain* the phenomena of interest; we are only interested in constructing the simplest possible models, with no detailed empirical grounding, but with the potential to generate heterogeneous wealth outcomes reminiscent of ones obtained using more sophisticated models. In doing so, we aspire to shed light on the true social and economic mechanisms underlying the genesis and persistence of economic inequality, and network structure alone can produce large differences in economic outcomes.

The phenomenon under scrutiny is the emergence and evolution of economic inequality in societies with non-trivial social structures that enable economic interactions and cooperation mechanisms. The model we use to answer questions surrounding this phenomenon must possess a dynamic model of wealth accumulation, a model of social structure, and a suitably useful measure of inequality. The dynamics are modeled as Brownian-noise-driven linear dynamical systems, here with constant growth rate; the structure is modeled by networks, here with random graphs; the measures of inequality used are Gini coefficients, and survival time distributions of agents in response to lack of resources. Just as with macroeconomic models, measures of wealth inequality like Gini coefficients are not so well suited for the non-market-based wealth dynamics. We discuss this more in Sect. 4.

Although both the dynamical system and the network model are quite well understood, the precise interplay of structure and dynamics produces interesting emergent wealth distributions and robustness to scarcity. To the best of our knowledge, this specific combination of network structure and social dynamics has not been discussed in the computational social science (CSS) or mathematical sociology literature, and we consider this the primary contribution of this work.

Apart from enabling interactions between social actors, social structures like institutions also shape the form of cooperation and coordination mechanisms. The institutions can take the form of economic institutions, like banks and cooperatives; or the form of norms, like resource-sharing practices in societies. The Friesen and Mudigonda model [20] consists of a simple resource pooling arrangement where

aggregates of agents pool their excess wealth in a common institution called "proto-institutions," agreeing to provide this saved resource to individual agents in times of need. The model of resource pooling used in this work is identical to this model. The analysis to be presented in later sections focuses only on homogeneous agents with simple drift-diffusion dynamics on Erdős-Rényi network models (ER); space constraints unfortunately prevent us from repeating the analysis on other standard *textbook* networks like scale-free and small-world networks. We discuss empirical evidence for the role of social network structures in the concluding section of this paper, motivating the need for more expressive social network models.

In the next section, we discuss the mathematical formulation of the model. In Sect. 3, we present the analysis of our simulation experiments, summarize key findings, and discuss why we chose the `julia` programming language [3] for our implementation. In Sect. 5, we discuss the limitations of our simple models, extensions to dynamics and networks more expressive than the ones presented here and planned future work.

Before discussing our model in greater mathematical detail, we discuss the model qualitatively, contrasting it with more familiar modeling approaches. The model presented here has much in common with models used in social-reality-inspired models in statistical physics [26], dynamic process models in network science [21], computational social science, and agent-based models [18]; however, our modeling philosophy is somewhere at the interface of agent-based computational economics (ACE) [30] and analytical sociology (AS) [8, 9]. In ACE, we acknowledge its aspiration to develop bottom-up models of economic systems at all scales but restrain from its enthusiastic use of complex but well-calibrated detailed models of markets and agents [31]. In AS, we assent its focus on social mechanisms in explaining social phenomena, but instead rely on simplified mechanisms with few parameters [17] with the specific goal of extracting insights from *stylized* models.

More specifically, from statistical physics, we borrow the dynamics: diffusion models and associated first-passage time techniques; from network science, we borrow the structural aspects: Erdős-Rényi network (ER) models; and from non-equilibrium statistical physics, CSS and ACE, we borrow a form of cooperation: the concept of coalescence, institution, and coordination.

2 Model

The model of cooperative wealth accumulation constructed here is best thought of as a stochastic interacting particle system infused with economic sociological semantics. The particle evolves according to a one-dimensional diffusion process with constant drift and is driven by a Brownian noise with a boundary condition at the origin corresponding to particle absorption. After crossing a pre-determined threshold in state space, particles above the threshold follow a protocol and coalesce together. In an ensemble of otherwise identical particles, a given particle may coalesce with a subset of other particles. In grid world ABMs, agents interact with other agents

by moving around this world. The precise movement protocol encodes how the agents interact (mix) among themselves. In the social network setting, this mixing characteristic is encoded via a graph. Both the conjoined particles and individual particles die upon crossing the origin. This model of interacting diffusing particles can be provided with substantive semantics as follows.

The one-dimensional state space of the particle is identified with the wealth of a social actor (agent). We consider a homogeneous population where all agents are required to burn their wealth at a constant specified rate in order to survive. In addition, the agents all gain wealth at a constant rate. The resource draining rate and the resource gaining rate are additive and constitute the drift of the diffusion process. The environmental contingency is modeled by a Brownian noise of a specified intensity. Agents can coalesce to form a cooperative unit deciding to pool their resources and their environmental contingencies into a single unit which we call a *proto-institution* (proto), if they cross a specific wealth threshold. The mixing characteristic of this ensemble is the social network (here ER model).

The questions of interest to us, emergence and persistence of inequality and robustness to scarcity, in the presence of social structure and uncertain environmental conditions, map onto questions about the stochastic dynamical system (SDS). Inequality can be quantified using Gini coefficients of the particle ensemble's state space. Robustness to scarcity can be measured using survival time analysis for the particles to reach the absorbing boundary at the origin by turning off the appropriate drift terms in the dynamics.

2.1 Mathematical Formulation

The discussion of SDS closely follows [26].[2]

The SDS can be defined through a stochastic differential equation of the form

$$dx(t) = vdt + \sqrt{2D}dw \tag{1}$$

where v is the wealth growth rate (the difference of the income and metabolic rate of the agent) and D is the intensity of the Brownian process (white noise process) w. $x(t)$ is the state of the particle (wealth) at time t. The dynamics can be started at any initial point $x_0 > 0$. Since simulations make use of discrete versions of these equations, a slightly different notation is used.[3]

[2] The use of such analysis for studying wealth dynamics was presented by one of the authors at CSSA18 [32]. Unlike the finite interval dynamics used there, the dynamics here take place on a semi-infinite line.

[3] The discrete-time step $\Delta t = 1$. w in the discrete setting is a Gaussian distributed random variable $\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, D\Delta t)$.

While Eq. (1), in its discretized form, is the basis for simulations, other formulations are used [26, 32].[4]

For a single particle, the probability of survival $S(t)$ up to time t and the probability of reaching the origin can be calculated and are functions of x_0, v and D. For example, the expected probability of a particle reaching the origin ($\mathcal{E}(x_0)$), when starting at x_0 is given by

$$\mathcal{E}(x_0) = \begin{cases} e^{-vx_0/D}, & \text{for } v > 0 \\ 1, & \text{for } v \leq 0 \end{cases} \tag{2}$$

That is, there is always a non-zero probability of reaching the origin, even when there is a constant positive drift away from the origin; and, the return to origin is certain for negative or zero drift. Similarly, survival probability, $\mathcal{S}(t)$ is given by

$$\mathcal{S}(t) = \begin{cases} 1 - e^{-vx_0/D}, & v > 0 \\ \sqrt{\frac{4D}{\pi v^2 t}} e^{-v^2 t/4D}, & v \leq 0 \end{cases} \tag{3}$$

Since the particles are independent, they are independent dynamical systems; proto formation is a higher-level construct imposed on this system and is a constraint on what states are considered viable and which are not. Consider two particles x_1 and x_2 that have crossed the threshold for coalescence x_{thresh} (t_p in the next section). After the coalescence event, the state (wealth) of individual particles $x_1(t)$ and $x_2(t)$ is not relevant for survivability; only the aggregate wealth of the coalescent (proto) determines whether the particles in the coalescent survive. As long as $x_1(t) + x_2(t) > 0$, both particles survive. Effectively, the proto is a two-dimensional SDS. Since the wealth of the proto (p_{12}) is additive, the aggregate wealth $p_{12}(t) = x_1(t) + x_2(t)$. The aggregate dynamical variable satisfies a SDE.[5,6]

$$dx(t) = 2vdt + \sqrt{2(2D)}dw \tag{4}$$

As one can see, only the coefficient of drift and diffusion in Eqs. (1) and (4) are different. So, the corresponding expressions for $\mathcal{E}(x_0)$ and $\mathcal{S}(t)$ are suitably scaled. This has implications for both Gini coefficient calculations and survival analysis calculations.

For such ensemble of particles, the primary driver of difference in paths (life histories) and wealth is the Brownian noise intensity D; greater the D leads greater

[4] These formulations make use measure theoretical probability to convert SDEs to partial differential equations known as Fokker-Planck equations. They provide numerical and closed-form estimates of probability density, survival time probabilities, and other quantities of importance.

[5] The result follows from the additivity properties of white noise.

[6] The two pictures: the particle perspective and the proto perspective, are equivalent. While it is easier to mathematically analyze the system in the proto perspective, the individual wealth of the particles carries meaning; it is just not useful for studying survival of the proto or the particles within it.

the diversity (and hence Gini). However, Gini is a measure that is dependent on absolute magnitude. So, if the drift (v) is large, then the variation generated by D gets washed out by v.[7] On the other hand, the mathematical form of expressions for $\mathcal{S}(t)$ suggests a clear dependence on D which separates the population of particles that are not in a coalescent and the population of particles that are in one. Particles reache the absorbing state (death of the agent) without being part of a coalescent are called isolates; the non-isolates, the particles that are in a coalescent when they reach the absorbing state are called protos.

As the Eqs. (2) and (3) show, all agents, irrespective of their starting initial condition and luck, have a non-zero probability of dying. The paper that inspired this work [20] studied inequality dynamics in the context of extreme scarcity in their foraging world. One can mimic this scarcity by simply turning off the income, starving the agents, and leaving the agents to all inevitably die. The differential rate of death between the isolates and the non-isolates then becomes an important quantity. In the absence of tractable mathematical solutions, simulations experiments which probe this expected differential survival rates help illuminating the role of protos.

We first evolve the system under favorable environmental conditions with steady salary, and identify phases before and after proto formation take place (Stage 1 and 2). After making sure that all surviving agents possess wealth above the proto formation threshold x_{thresh}, we cut off the salary for all agents, starving the agents, leaving them susceptible to the white noise process (Stage 3). In each of the stages, the effective dynamics are different from each other, but the form of the equations remains the same.

To break the homogeneity of this particle ensemble, we propose to let the particles interact with only a random subset of other populations. In other words, we embed the particles on a network (graph); particles interact (form protos) with only their neighbors. In this work, we focus on the simplest *textbook* example of network: Erdős-Rényi network model [21]. ER models have only one parameter λ, a parameter that is the average expected degree of any given node that determines the density of connections in the network, and which in this setting determines the expected number of particles available for a particle to coalesce with and form a proto. Commonsense reasoning suggests that a larger λ parameter may help increase the average lifetimes of individual agents by enabling greater likelihood of contact with other particles.

Just as Eqs. (2) and (3) offer insight into the role of dynamics, the key feature of ER models is high probability of formation of giant component at $\lambda >= 1$, in which a finite fraction of the population to be connected, have cliques of very high order, and other useful properties guarantee proto formation in large fraction of the population.

As its network properties are well-understood, the network's wealth dynamics allow us to understand more thoroughly the interaction between dynamics and struc-

[7] Preliminary investigations suggest that for simple non-network ensembles, Gini either stays close to 0 or 1. We suspect that this is because of the constant wealth growth rate used in our models, Gini is a partially useful measure. This is unlike in macroeconomic models where exponential growth rate gives rise an appearance of non-trivial Gini coefficients.

ture, and its role in global behavior. Still, a full analysis of this model remains to be completed and is part of a forthcoming work involving other more expressive network models. As we note subsequently, even this simple setting offers interesting insights regarding the role of network structure in determining important outcomes for the agent.

2.2 Implementation

The simulation proceeds as follows. For configurable parameters $N > 0$, `init_max` > 0, `salary` > 0, $m > 0$ (metabolic rate), $t_p > 0$ (the "proto threshold"), $\sigma^2 \geq 0$ and $0 \leq \lambda \leq 1$:

(1) Create N agents, each with a random initial wealth (uniformly distributed from 0 to `init_max`), and related to one another as per a random ER network with parameter λ.
(2) **Stages 1 and 2**[8]: Repeat until all non-isolate agents[9] are members of a proto:
   (a) Each agent A whose current wealth $\geq t_p$, and who is not currently a member of a proto, chooses at random one of its graph neighbors (call it B) whose wealth *also* exceeds t_p. (If there are no such neighbors, proceed to the next agent.) If B is already in a proto, have A join B's proto. If not, have A and B form a new proto.
   (b) Each agent gains an amount of wealth equal to $(\texttt{salary} - m + \epsilon)$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ (white noise).[10]
   (c) Each agent *that is in a proto* donates all its wealth in excess of t_p to that proto's balance. (Agents not in a proto maintain their current wealth.)
(3) **Stage 3 (starvation)**: Repeat until all agents are dead:
   (a) Each agent *loses* an amount of wealth equal to $(m + \epsilon)$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
   (b) If an agent's wealth would drop below zero as a result of this loss, and if it is not a member of a proto, it dies and is removed from the simulation. If it *is* a member of a proto, it withdraws the necessary amount from its proto's balance to remain at zero wealth. If the proto does not have sufficient funds to cover the loss, both the agent and the proto die and are removed from the simulation.

Various statistical counters are updated as the program executes so that its behavior can be analyzed postmortem. The main simulation loop can also be invoked from

[8] We define "Stage 1" as the period *before* the first proto is formed, "Stage 2" as the period *during* which protos are being formed.

[9] An **isolate** agent is one with no graph neighbors.

[10] Note that this "gain" could be negative, in which case the agent, and possibly its proto, may be subject to death exactly as in step (3).

a "parameter sweep" program which executes it multiple times over a range of parameter values, in order to determine how the model's behavior changes in response to key parameters.

3 Verification

We first verify that the simulation's output matches obviously expected results. Then, in the following section, we investigate aspects of its behavior which are not computable analytically in order to discover important consequences of the model.

3.1 Agent and Proto Life History

For a sensible range of parameter settings, the life history of a single simulation run follows an expected pattern. Assuming a positive growth rate (*i.e.*, salary > metabolic rate), each agent's wealth rises unsteadily during Stage 1, until eventually the first pair of neighboring agents who each reach the threshold form a proto-institution. Throughout Stage 2, these agents contribute all wealth in excess of the threshold to that proto, and so their proto's balance rises unsteadily while their personal wealth remains at the constant threshold. Meanwhile, the other agents also reach the threshold at various points in time, and also form or join protos, until every non-isolate (that is, every node with at least one neighbor) is a member of a proto. Stage 3 (the starvation period) then commences, with isolates drawing on a greater personal wealth than the non-isolates. When an isolate reaches zero wealth, it dies; when a non-isolate reaches zero, it draws from its (shared) proto balance until that too, reaches zero, and it dies.

This history can be seen in Fig. 1. The top plot depicts agent wealth at every iteration, and the bottom plot shows the balances of the protos at the same points in time. (No protos exist until Stage 2, by definition.) Note that the isolates (orange lines) never form protos, and therefore begin the starvation stage with a higher personal wealth to draw from.

Figure 2 shows the same information in another way: instead of plotting each agent's *personal* wealth (as in the top plot of Fig. 1), we show its **effective wealth**, defined as the sum of its personal wealth and its "share" of its proto's wealth (if any). After all, contributions that a proto's members make to its balance are available to those members in times of need; therefore, a fair comparison between isolates and non-isolates should take this into account. From the figure, it can be seen that isolates no longer have a systematic advantage (as they appeared to in Fig. 1).

We verified that changes to basic parameters all have the expected effect: lowering the initial wealth delays the onset of Stage 2; a higher σ^2 for the income distribution makes the lines less jagged; a higher metabolic rate hastens extinction; *etc.*

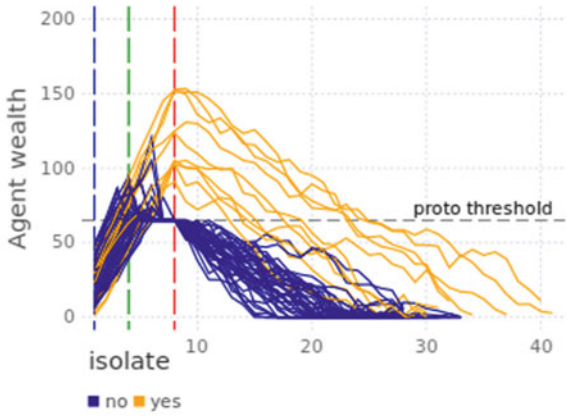
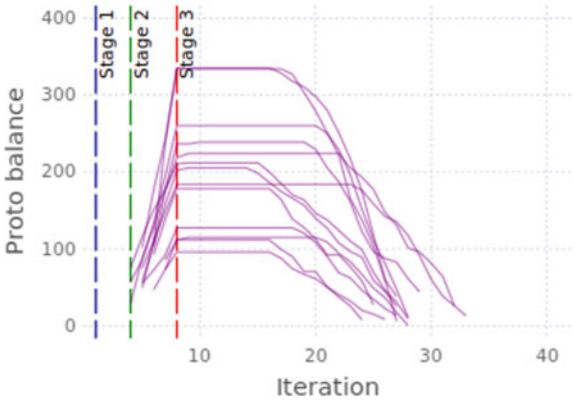


Fig. 1 A single run of the simulation, with $\lambda = 2$. Each of 50 nodes is given an initial wealth of $\sim \mathcal{U}(0, 50)$ units, a regular income distributed as $\sim \mathcal{N}(20, 5)$, a metabolic rate of 5, and a proto threshold of 65



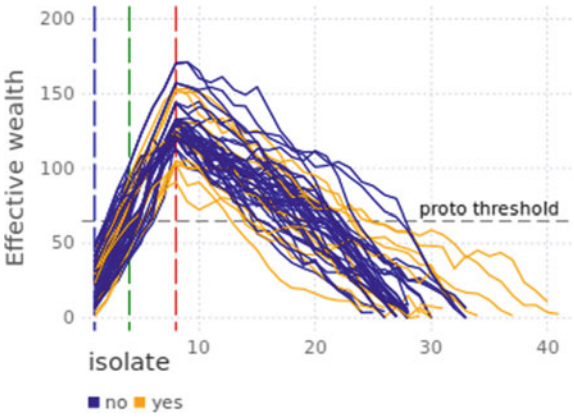


Fig. 2 The same simulation run as Fig. 1, but this time depicting each agent's *effective* wealth (its personal wealth plus its share of its proto's wealth, if any)

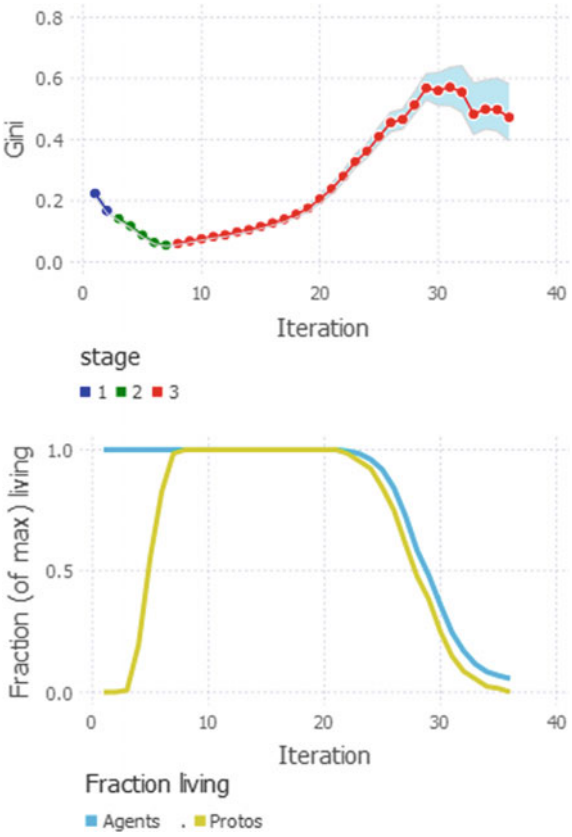


Fig. 3 The simulated society's wealth inequality over time. (The same simulation parameters were used as in Figs. 1 and 2, but this time with 500 agents.) The light blue band represents a bootstrapped 95% confidence interval

3.2 Gini Coefficient History

The Gini coefficient of the agent population as seen in the top plot of Fig. 3 is influenced by two distinct dynamics: loss/accumulation of wealth and the formation/death of proto-institutions. Over the course of Stage 1 and the beginning of Stage 3, this change in agent wealth is exclusively responsible for changes in the Gini coefficient. As agents accumulate wealth over Stage 1 and 2, the size of wealth differentials shrinks relative to absolute agent wealth, leading to the declining Gini coefficient. The opposite effect occurs during Stage 3 as agent starvation leads to the relative growth of these wealth differentials. The variability in starvation rates further stimulates the increasing Gini coefficient.

In addition, as indicated in the bottom plot, the proto formation and proto death also influence the Gini coefficient during Stage 2 and the end of Stage 3, respectively. As expected, the formation of protos during Stage 2 contributes to declining Gini coefficient as the constituent agents of each proto have equivalent wealth values and represent coalitions of perfect economic equality. Accordingly, the death of

protos, beginning around Iteration 20, contributes to increase the Gini coefficient by removing the protos' effect on the system's inequality.

(Note: As the population size decreases over the starvation period, the Gini coefficient becomes increasingly unstable and susceptible to small fluctuations in agent wealth; hence the erratic nature of the red line at the extreme right of Fig. 3.)

4 Analysis

4.1 Gini Coefficient

As mentioned in Sect. 1, the Gini coefficient is not the ideal measure of inequality for our apocalyptic model. Nonetheless, it is illustrative to see how it varies with respect to the ER λ parameter. Figure 4 depicts the Gini computed *at the onset of Stage 3* (before starvation) versus λ, and confirms that increasing λ leads to a decreasing Gini coefficient. Increasing λ fosters wealth uniformity through the increased formation of and growth in size of protos. Firstly, a higher percentage of agents join a proto as greater λ values lead to fewer isolates in the ER network. As more agents join protos, differences in agent wealth are eliminated as each proto establishes perfect equality among its constituent agents, thereby lowering system's overall inequality. Secondly, higher λ values lead to larger average proto sizes, as more densely connected networks increase the likelihood that an agent will join an existing proto rather than form a new one. In much the same way, as smaller protos coalesce into larger ones, the standard inequality between the fragmented protos is eliminated in favor of perfect equality across the larger proto, resulting in a corresponding decrease in the Gini coefficient.

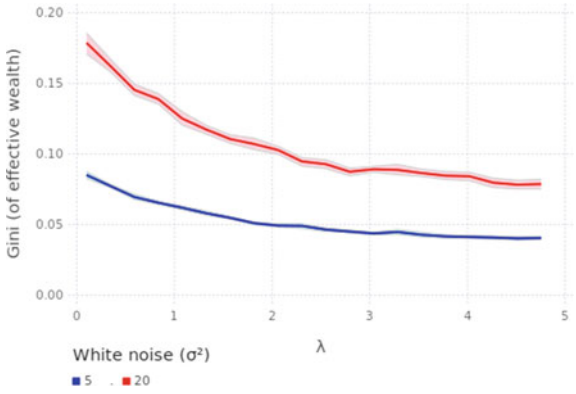


Fig. 4 The average Gini coefficient of effective wealth (computed pre-Stage 3) for various values of the ER λ connectivity parameter, and with both low-noise and high-noise income. 500 agents were used in each simulation. The color band represents a bootstrapped 95% confidence interval

4.2 Life Expectancy

Rather than absolute wealth, which the Gini coefficient measures, an alternate measure of well-being is the ability to survive an economic downturn. This, after all, is the chief benefit an agent should be able to expect from joining a proto: it serves as a kind of insurance policy against future poverty. It is therefore interesting to compare the life expectancy of agents who join protos with those who do not.

There are many factors at play here, one of which is the level of white noise (σ^2) in the agents' income. Figure 5 depicts how the life expectancy of isolates and non-isolates depends on σ^2 for two different values of λ. The top plot shows that for relatively stable agent income levels, there is not much difference between the two lines—and hence, not much advantage (or disadvantage) to an agent's joining a proto. Interestingly, however, the more volatile the income stream becomes, the more benefit there is to pooling resources. The effect is even more pronounced with more densely connected graphs, as in the bottom plot: here, when income is more noisy, agents who join protos live nearly twice as long as those who don't.

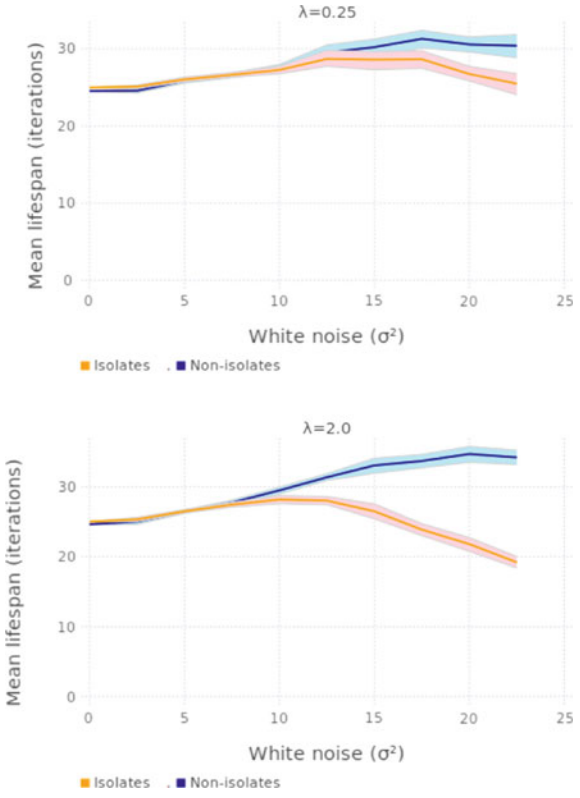


Fig. 5 Life expectancy comparison between isolates (non-proto members) and non-isolates (proto members) for different values of λ and σ^2. The `salary` parameter was set to 20, so the x-axis ranges from a nearly constant agent income to a scenario when the noise is as high as the average

4.3 Interpretations, Conjectures, and Next Steps

Unraveling the interplay of structure and dynamics is a major objective of this offshoot of the Milton and Mudigonda model. The necessarily preliminary analysis reported here shows interesting results in this direction. As the two sources of heterogeneity, both the ambient stochasticity and interaction probability seem to influence the two kinds of inequality indicators.

As expected, Fig. 4 shows the role of environmental noise: larger environmental noise produces larger inequality measures. Similarly, Fig. 5 shows the role of noise in amplifying differences between isolates and non-isolates: the larger the noise, the larger the separation between the mean lifetimes of the two populations. Also, as mentioned above, larger λ leads to a more egalitarian population (Fig. 4); concurrently, larger λ leads to larger separation between mean lifetimes of the two populations (Fig. 5).

Ideally, we would like to derive these results mathematically, especially the point beyond which the average lifespan of two populations change. Since the SDS has only a few parameters, it would easy to decompose the contributions of the various factors response for mean differential lifetimes as it is unclear whether the differences are due to wealth stabilization induced by proto formation in Stage 2, or in Stage 3. Another interesting question is whether the population can be further stratified along proto-size dimensions. We conjecture that protos with larger number of agents will have larger mean lifetimes than protos with smaller numbers in the aggregate. We also conjecture that the time spent in a proto positively influences the mean lifetimes of agents in it.

While the mathematical formulation of the model was presented for a single particle, the system consists of a large ensemble of particles. Many of the characteristics, the differentiation of the system into isolates and non-isolates depend on the order rank order statistics of wealth. Also, a more nuanced statistical analysis that goes beyond the mean analysis presented here is required to tease out the necessary and sufficient conditions for agents in protos (non-isolates) to consistently outperform the isolates. We are currently pursuing these questions.

5 Conclusion and Future Work

This is one of the first in an ongoing project and the next directions are too many to mention here. Still, we mention a few next steps that adhere to the spirit of our project. These directions all alleviate some of the obvious limitations of the current work.

While both the stochastic models and ER models are well understood, proto formation dynamics, as coagulation processes in continuous time, remain to be better understood, mathematically. Also, while the mathematical analysis of dynamics of

discrete-time stochastic models on networks is well studied, diffusion processes on networks seem underexplored in the literature.

Even with only a few parameters, the simulation results are hard to visualize and interpret. As the number of parameters increases, as expected in future models, ideas from design and analysis of (computer) experiments may be required. Also, our exploration of the time to death distributions of the population shows non-trivial structure. More careful statistical tests that go beyond the mean analysis presented here are required to tease out the necessary and sufficient conditions for agents in protos (non-isolates) to consistently outperform the isolates.

The models used for our simulation are at best *stylized* models of real-world social and economic systems, especially in anthropology [33] and historical and urban sociology [12, 27]. Research in economic anthropology [11, 14–16, 22, 24, 25, 29] suggests complex food and economic resource-sharing rituals among members of various communities. Such resource-sharing social networks do not look like any of the *textbook* models. Extending our analysis to more expressive network models like exponential random graph models, stochastic block models, and latent space models is an important research direction. This alongside the use of empirically observed cooperation and coordination protocols have potential in making our models better calibrated with real-world systems.

Despite their simplicity, models like the ones constructed here have several advantages. As ACE models, they offer insights about economic systems in which the majority of the assumptions of neo-classical economics like perfectly mixed agents [34] and presence of equilibrium [1] do not hold. As AS models [7], they offer an approach that adds models of social mechanisms to CSS models in a graded manner.

References

1. Brian Arthur, W.: Chapter 32 out-of-equilibrium economics and agent-based modeling. In: Handbook of Computational Economics, vol. 2, pp. 1551–1564. Elsevier (2006)
2. Benhabib, J., Bisin, A.: Skewed wealth distributions: theory and empirics. J. Econ. Lit. **56**(4), 1261–91 (2018)
3. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. SIAM Rev. **59**(1), 65–98 (2017)
4. Granovetter, M.: Soc. Econ. Harvard University Press (2017)
5. Granovetter, M.: The impact of social structure on economic outcomes. J. Econ. Perspect. **19**(1), 33–50 (2005)
6. Greif, A.: Cultural beliefs and the organization of society: a historical and theoretical reflection on collectivist and individualist societies. J. Polit. Econ. Polit. **102**(5), 912–950 (1994)
7. Hedström, P., Bearman, P.: The Oxford Handbook of Analytical Sociology. Oxford Handbooks, OUP Oxford (2011)
8. Hedström, P., Bearman, P.: What is analytical sociology all about? an introductory essay. In: Hedström, P., Bearman, P. (eds.) The Oxford Handbook of Analytical Sociology. Oxford Handbooks, chapter 1, pp. 3–24. OUP Oxford (2011)

9. Hedström, P., Udehn, L.: Analytical sociology and theories of the middle range. In Hedström, P., Bearman, P. (eds.) The Oxford Handbook of Analytical Sociology. Oxford Handbooks, chapter 2, pp. 25–47. OUP Oxford (2011)
10. Jackson, M.O., Rogers, B.W., Zenou, Y.: The economic consequences of social-network structure. J. Econ. Lit. **55**(1), 49–95 (2017)
11. Kasper, C., Mulder, M.B.: Who helps and why?: cooperative networks in mpimbwe. Curr. Anthropol. **56**(5), 701–732 (2015)
12. Katz, M.B.: The People of Hamilton, Canada West: Family and Class in a Mid-Nineteenth-Century City. Harvard Studies in Urban History Series. Harvard University Press (2013)
13. Keuschnigg, M., Lovsjö, N., Hedström, P.: Analytical sociology and computational social science. J. Comput. Soc. Sci. **1**(1), 3–14 (2018)
14. Koster, J., Leckie, G., Miller, A., Hames, R.: Multilevel modeling analysis of dyadic network data with an application to Ye'kwana food sharing. Am. J. Phys. Anthropol. **157**(3), 507–512 (2015)
15. Koster, J., Lukas, D., Nolin, D., Power, E.A., Alvergne, A., Mace, R., Ross, C.T., Kramer, K., Greaves, R., Caudell, M., et al.: Kinship ties across the lifespan in human communities (2019)
16. Koster, J.M., Leckie, G.: Food sharing networks in lowland Nicaragua: an application of the social relations model to count data. Soc. Netw. **38**, 100–110 (2014)
17. Macy, M., Flache, A.: Social dynamics from the bottom up. In: Hedström, P., Bearman, P. (eds.) The Oxford Handbook of Analytical Sociology, Oxford Handbooks, chapter 11, pp. 245–268. OUP Oxford (2011)
18. Macy, M.W., Willer, R.: From factors to actors: computational sociology and agent-based modeling. Ann. Rev. Sociol. **28**(1), 143–166 (2002)
19. Martin, J.L., Lee, M.: Social structure. In: Wright, J.D. (Ed) International Encyclopedia of the Social and Behavioral Sciences, vol. 22, 2nd edn., pp. 713–718. Elsevier (2015)
20. Mudigonda, S., Friesen, M.: Institutional emergence and the persistence of inequality. In: Proceedings of Computational Social Sciences Society of the Americas 2018 Annual Conference. Springer Press (2018)
21. Newman, M.: Networks. Oxford University Press, Oxford (2018)
22. Nolin, D.A.: Food-sharing networks in Lamalera, Indonesia: status, sharing, and signaling. Evol. Hum. Behav. **33**(4), 334–345 (2012)
23. Piketty, T., Goldhammer, A.: Capital in the Twenty-First Century. Harvard University Press (2017)
24. Power, E.A., Ready, E.: Cooperation beyond consanguinity: post-marital residence, delineations of kin, and social support among south Indian Tamils (in press). Philos. Trans. Roy. Soc. Biol. Sci. (2019)
25. Ready, E., Power, E.A.: Why wage earners hunt: food sharing, social structure, and influence in an arctic mixed economy. Curr. Anthropol. **59**(1), 74–97 (2018)
26. Redner, S.: A Guide to First-Passage Processes. Cambridge University Press, A Guide to First-passage Processes (2001)
27. Sampson, R.J., Wilson, W.J.: Great American City: Chicago and the Enduring Neighborhood Effect. University of Chicago Press (2012)
28. Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T.: Assessing "neighborhood effects": social processes and new directions in research. Ann. Rev. Sociol. **28**(1), 443–478 (2002)
29. Smith, D., Dyble, M., Major, K., Page, A.E., Chaudhary, N., Salali, G.D., Thompson, J., Vinicius, L., Migliano, A.B., Mace, R.: A friend in need is a friend indeed: need-based sharing, rather than cooperative assortment, predicts experimental resource transfers among Agta hunter-gatherers. Evol. Hum. Behav. **40**(1), 82–89 (2019)
30. Tesfatsion, L.: Chapter 16 agent-based computational economics: a constructive approach to economic theory. In: Handbook of Computational Economics, vol. 2, pp. 831–880. Elsevier (2006)
31. Tesfatsion, L.: Modeling economic systems as locally-constructive sequential games. J. Econ. Methodol. **24**(4), 384–409 (2017)

32. Venkatachalapath, R.: Revisiting Markov models of intragenerational social mobility. In: Proceedings of Computational Social Sciences Society of the Americas 2018 Annual Conference. Springer Press (2018)
33. White, D.R.: Kinship, class and community. In: Scott, J.C., Carrington, P. (eds.) Sage Handbook of Social Networks. Sage Publications, New York, pp. 129–147 (2011)
34. Wilhite, A.: Chapter 20 economic activity on fixed networks. In: Handbook of Computational Economics, vol. 2, pp. 1013–1045. Elsevier (2006)

Social Primitives: Exploring Spark of Life Collective Behavior in Agent-Based Models

Srikanth P. Mudigonda and Milton J. Friesen

Abstract This research is focused on finding the simplest possible agent-based model called SPECscape (Social Primitives Experimental Cohort) that can demonstrate the emergence of wealth inequality. Agents feature a simple North-South-East-West best sugar patch search function within a 2D grid style code environment that allows formation of a proto-institution (common pool resource capability) under certain conditions. A Nearly Orthogonal Latin Hypercube (NOLH) is used to explore the behavior space of the model's dynamics with four distinct sugarscape arrangements and introduction of exogenous shocks at specified stages of the model's evolution. Our results suggest that proto-institutions and moderate shocks, are beneficial for agent members, and play an important role in lowering wealth inequality when many institutions are present and increasing wealth inequality when only a few are allowed to form, thereby indicating the presence of such institutions have a significant effect on wealth inequality in a society of agents.

1 Introduction

This paper examines a simple agent-based model that can demonstrate the emergence of wealth inequality. Our interest is to clarify understanding about the primitive dynamics that may be implicated in human inequality. This work, however, is not an anthropological study and the term "primitive" is used in an artificial, computational sense. We are not attempting to demonstrate here the origins of inequality as an empirical, human cultural fact. Digital primitives are magnitudes of order different than even the most basic human exchanges, so that complex social issues such as social norms [1] are relegated to simple coded rules. We use an agent-based model to

S. P. Mudigonda (✉)
Saint Louis University, 1 N Grand Blvd, St. Louis, MO 63103, USA
e-mail: srikanth.mudigonda@slu.edu

M. J. Friesen
University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
e-mail: m2friese@uwaterloo.ca

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_19

study the dynamics of inequality as an emergent property of agent interactions under certain primitive conditions as a means of bench testing social dynamics [2–5].

Our empirical orientation to the issue of inequality arises from a study of public records research in the City of Hamilton, Ontario during the years 1851–1860 [6]. Development of an agent-based exploration of this work suggested that there was potential for more extensive scrutiny of simplification approaches along with expanded data output analysis [7]. The key empirical dynamic was the existence of a small group of well-off people who remained in Hamilton over the decade of the 1850s, in contrast to a class of people (~80% of the population) who grew in numbers as a totality but among which there was a turnover rate of 2/3 during the same decade. The presence and participation in stabilizing institutions for the wealthy (and relative absence of such institutions for the lower 80%) was hypothesized by Katz as a primary driver of those contrasting dynamics. The transient nature of low-income populations has been observed in many contemporary settings leading to weaker social ties and fewer institutional investments [8]. Inequalities appear to be integral to many aspects of social structures [9].

Inequality is not a simple concept and the dynamic may exist across a range of social, cultural, and economic axes without clear consensus or understanding [10]. Rousseau discussed inequality at length in the mid-eighteenth century [11]. Changing one kind of inequality dynamic can lead to inverse changes in another [12] while attempts to make all people equal in all ways introduces impossible negotiations and dynamics [13]. The very terms and cultural environments involved in studies, interventions, and explanations of inequality may include a range of built-in biases and misconceptions [14, 15].

When Schelling and Sakoda, respectively, [16, 17] developed simplified models of one kind of inequality, neighborhood segregation, our understanding of the relationship between micro-motives and macro-behaviors was advanced in important ways even if such dynamics did not fully explain the complexities of actual segregation.

In a similar spirit, we are endeavoring to understand the structural aspects of inequality that arise when agents are not conscious, operate by very simple, deterministic rules, but have a certain probability of forming collective functions. We undertake this by comparing resource levels of agents who are not affiliated with a collective mechanism (a proto-institution in our model) and agents who are affiliated with a collective mechanism. Agents do not have an ability to choose to be part of a proto-institution and there are no dynamics related to agent features or characteristics that shape their choice such as may be found in employment, labor, or standards-driven models [18].

Our experimental foray is made with an awareness of the growing scope and scale of computational economics that begin at the bottom and grow their way up [19]. We want to examine as fully as possible the simplest form(s) of this differentiating dynamic. The example of Hamilton, ON in the 1850s provided the initial case study for the role that collective memberships may play in the structural dynamics of economic inequality [6] but we are not seeking to simulate that historical scenario in what follows.

2 Examples of Agent-Based Social Primitives

Agent-based models have been used in simple forms for a range of purposes. In what follows, we will examine Ring World and the Game of Life in the context of reflections on complex adaptive systems and a smoking cessation case study that tested the value of agent-based models for policy decision input.

2.1 Ring World

The value of artificial societies can be understood by using a framework that begins with simple elements and builds complexity on top of that [20]. Epstein and Axtell began with the assumption that if you could begin with agents and grow an artificial society that mirrored known, empirical dynamics, then you would have understood the dynamics involved. The most simplest model they proposed was Ring World.

In Ring World, the landscape is simplified as a connected ring with sugar levels of 4 assigned to each space at each step. During each time step, an agent in Ring World looks for the nearest unoccupied site with maximum sugar. If there are two agents in the ring and they get near to each other, then the logic of the coding says that an agent behind another agent won't take the occupied site but will take the site just ahead of the other agent—nearest with most sugar. Now the agent that got leap frogged will do the same. This creates a dynamic where they move together, apparently forming a group or clique. The agents "do stay together once they randomly encounter each other" [20]. They discovered a wide range of behaviors for this simple artificial set-up.

The relationship between individual agents and the larger modeling environment is complex so that the effect of agent action on larger social structures at different scales represents a significant challenge [21]. It is possible to link the endogenous agent dynamics with larger systems effects [22], but the intricacies of dynamics between and across social structures require careful interrogation of the source of those dynamics.

Our interest is in the nature of inequality and the role of a collective function among agents which, in the case of Ring World, is an artifact of the systems code—an agent-environment coupling. Agents are not able to cooperate in any way—they do not communicate, exchange information, or decide it is better to be in a group. The cliques that form result from interaction effects, important as examples of how emergence can occur, but not an example of cooperative behavior.[1] This is a significant primitive dynamic. There are clues in Growing Artificial Societies that point to even more

[1] Although outside the scope of this paper, it is important to note John Von Neumann's work on **one dimensional cellular automata**: "John Von Neumann in the late 1940s undertook analysis of machine reproduction under the suggestion of Stanislaw M. Ulam. It was later completed and documented by Arthur W. Burks in the 1960s. Other two-dimensional cellular automata, and particularly the game of "Life," were explored by John Conway in the 1970s. Many others have since researched CA's. In the late 1970s and 1980s Chris Langton, Tom Toffoli and Stephen Wolfram

primitive dynamics (e.g., Appendix A: Software Engineering Aspects of Artificial Societies, p15 on "agent-environment and agent-agent" interactions, p19 note 23 on formal descriptions of landscape and agent states, etc.) but the primary focus was on adding complexity (reproduction, trade, conflict) rather than interrogating simplicity even further.

2.2 Conway—Game of Life

The famous "Game of Life" program developed by John Conway represents an early cellular automata model and a component of current agent-based models—typically the landscape or environment [20]. The discovery Conway made was that simple rules in a 2D grid where cells were affected by the state of their neighbors could lead to unexpected artifacts that would glide across the screen, blink, and undergo a range of other distinctive dynamics. Like the Ring World example above, the cells could evaluate the state of their neighbors but couldn't communicate in richer ways. These simple limits, however, gave rise to unexpected, complex behavior. Do those emergent patterns constitute cooperation? Do collections of cells in a common state ("alive") such as a glider represent a highly primitive form of inequality (vs. "dead" cells). Game of Life has been referenced in a wide range of disciplines as a source of insight about computational dynamics [23, 24] but more intensive interrogation about the "simplest possible states of inequality" may still be required.

2.3 Complexity Adaptive Systems—Page and Miller

Agent sophistication is a central consideration in developing social models [25]. In Ring World, the agents are very limited with only a simple set of instructions and very low dynamic range—move to nearest unoccupied space with most sugar. The problem is that model builders have not been as clear about the range of sophistication from the "hyperrational, hyperinformed, hyperable" agent to the "myopic simpleton" [25].

Moving from particle to fully functioning social system represents the introduction of many potentially confounding variables in virtual settings. The central goal of this paper is the identification of social primitives including critical analysis of the simplest possible model of institutions. There are not agreed to ways of establishing the exact nature of agent sophistication. That means, in turn, that we do not have an established means of differentiating levels of agent sophistication. Do we proceed by

---

did some notable research. Wolfram classified all 256 one-dimensional two-state single-neighbor cellular automata. In his recent book, "A New Kind of Science," Wolfram presents many examples of cellular automata and argues for their fundamental importance in doing science." (Description from the "Info" section of the CA 1D NetLogo Model Library, accessed June 2, 2019).

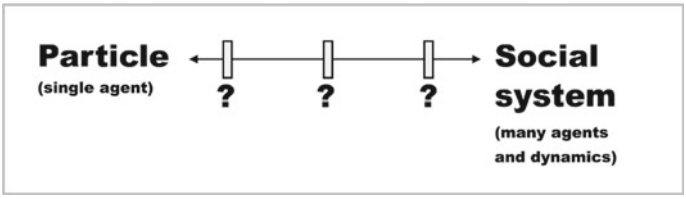


Fig. 1 Moving right increases sophistication and reflects emergence. Moving left reduces sophistication and reflects reduction. A taxonomy of critical transition points could be developed across this spectrum. We do not know if there are stopping points, if they are evenly spaced, or what effects may result from various interactions. It may be that emerging fields like chemoinformatics [28] that map massive possible chemical spaces could be instructive if applied to the design of new institutions, models, or feature sets [3]. Even simple physical systems like a sand pile make decomposition difficult—if you take sand away from a pile one grain at a time, at what point does it cease to be a sand pile [29]?

additive steps—absolute simplest and then add? Do we proceed by decomposition—take a suitable working model and then keep removing things until we are down to primitives, noting the effects along the way? Or are there other methods such as phase space/behavior space exploration but for sophistication rather than for known parameter variables in a model? (Fig. 1). Social primitives research requires that we interrogate the "myopic simpleton" side of the spectrum so that increases in sophistication are well understood [25]. Network science research discovered that a circle of nodes connected to first and second neighbors had a sudden decrease in average distance when nodes were randomly re-wired [26]. This dynamic was not owing to smarter or more capable nodes—they all remained homogenous. The sudden decrease in average distance between nodes was an artifact of the structural change in the network. Simple changes with significant dynamic impacts are an important feature of primitive systems [27].

2.4 Hammond—Tobacco and Policy—Low-Dimensional and High-Dimensional Models

A good example of a real-world problem being addressed by a very disciplined application of agent-based models is the tobacco use cessation study by Hammond that formally reviewed agent-based models in a highly disciplined way to answer an important public policy question about strategies to decrease tobacco use [30]. That discipline included identification of models that ranges from low-dimensional to high-dimensional approaches that are potentially useful and others that are potentially hazardous. Of particular interest was the effort to help individuals stop smoking. Aggregate data was useful descriptively but not in terms of understanding how smoking cessation actually worked for individuals. Agent-based models were identified as potentially relevant in understanding these dynamics.

The authors remark that the ability to run a given scenario over and over while changing variables can provide insight into either a strategic direction or introduce caution and more realistic expectations regarding various policy changes. For example, it is intuitive that increasing the legal age for smoking would prevent tobacco use but it may well drive up delinquent behavior as informal sources are pursued by underage smokers. Understanding the actual cessation causes would be necessary to determine which interventions would be most likely to yield desired results.

3 Taxonomy of Social Primitives

Examination of a relational dynamic requires elements that can interact. Agents, landscapes or environments, and some form of agent-level dynamic change are required. It can be challenging to fully specify the parameter space of a model even for simple dynamics. Implicit assumptions can also enter the model unwittingly through computing structures, coding approaches, language use and are of particular importance for the social primitives effort.

3.1 Agent Description

When we are interrogating the idea of social primitives, we must remain focused on systematic reviews of all features and elements. Agents are specific entities in an agent-based model. They are objects that are given characteristics, all of which are mathematically defined through the code. The idea of a relation using agents requires that there are at least two entities or objects that can be differentiated from each other and that they are able to interact in some way—either directly with each other or by means of something in the environment—such that the interaction changes something about a given agent (more or less of a resource, some characteristic of the agent, or a novel characteristic associated with the interaction such as becoming a member of a common group).

If we move beyond two agents, we must answer a range of questions: How many agents will populate a virtual landscape? How do they get there? What do they do when they arrive on the landscape? What cannot they do? How do they leave the simulation? What information do we collect on them? Agents can be in social networks without spatial definition, on landscapes that are constrained in two dimensions, others that operate in three dimensions, and they can possess an unlimited range of features that deeply impact the evolution and characteristics of the model.

3.2 Coding Agents and the Challenge of Implicit Assumptions

When a common term list of agent features is identified, translation of those terms into code suitable to run a model is another juncture where subtle influences may affect the effort to reduce unknown or unaccounted variables. Building agents from scratch is one way of minimizing this effect. Using code that is pre-built or models that are pre-built requires great watchfulness along this line. It would be useful to have a taxonomy of agents that range smoothly from the simplest possible agent up to the most sophisticated agent. Some earlier attempts to explore this can be glimpsed in Von Neumann's exploration of automata that are able to reproduce themselves [31] but computational sophistication has pushed us much more naturally to increasing agent, landscape, and social network features rather than making them more simple, fundamental and thoroughly understood.

3.3 Environment Description

Spatial, social, and a wide range of other real-world features are often built into our modeling environments with a noble intent—creating a model that is similar to the real world in ways that are useful, generate insight, and that can communicate results to those who are trying to understand a given problem. As with agents, these environments have grown dramatically in scope and scale so that whole cities, even the full global compliment of +7B people can be housed in silico for our experimental purposes [32].

3.4 Coding the Environment and the Challenge of Implicit Assumptions

As with agents, characteristics of an environment (whether social, spatial or otherwise) must be translated into computer code by programmers. Evolving platforms for collaboration reflect how challenging it is to balance domain expertise and knowledge with code-writing capabilities. Conceptual and aspirational hopes from people who know a social or cultural setting—e.g., people interested in effective policy change around smoking cessation—are converted into Python (via the Mesa framework), R, Julia, NetLogo, GNU SWARM, Java (via MASON or Repast) or similar programming environments so that iterations can be run and data collected. Coding and domain complexity means that no one person can fully synthesize these differences—they are negotiated among teams and groups who work together.

Our most difficult challenge is giving due consideration to the role played by the meta-control of the coding environment. This also includes the way in which models

are processed computationally, as serial systems, however quickly they are cycled. As noted above, the review of agent status and initiation of proto-institutions under certain conditions is the point at subtle details such as order, timing, and updating become much more important.

The gap is ripe for unwitting importation of assumptions; so, we need increased rigor to clearly understand the source of information coming out of models. It may be that fields such as swarm robotics hold promise for enacting more effective parallel agent interactions where each "agent" (individual swarm bot) operates without needing a procedure order from a controlling piece of code [33]. Robot swarms may be a way of operating truly parallel systems rather than the very fast but serial nature of many ABM computational model.

We are examining proto-institutions and their role in wealth inequality (resource distribution) without true agent cooperation. This is work contributes to the structural features of common pool resource dynamics but without the dynamics of cognition-driven agency. These simplifications are intentional, an experimental design approach focused on how proto-institutions change the evolution of a simple isomorphic system toward unequal distribution.

What is the operating framework doing, or, more specifically, how do we take stock of this functional framework as a contributor to the model's evolution? What order of operations do we follow when doing updates? Which agents go first? How are the patches in a 2D model seeded with resources—all at once or serially? In the notionally simple example of Ring World, the order of agent updates is noted as an important factor since in the execution of the operations, agents do move in a sequential set of turns but are updated randomly [20]. While Ring World may be robust to randomization of agent order, it suggests that this may not be a safe assumption in all cases and warrants more intensive examination.

Inequality and institutions research may also benefit from finding ways to encode both the agents and landscapes in ways that allow the institutional function to emerge out of agent interactions without prior specification. Would it be more logical to put agents into a landscape with no pre-coded "proto-institution" function and then see if a common pool dynamic would eventually emerge? In a fully constrained environment like a computational language with a given logic structure, novelty and mutation may be introduced through features such as genetic algorithms that can modify themselves [34–36]. These adaptive software structures can be applied to problems as diverse as detecting community structures in data or playing solos in a virtual jazz quartet. Our model uses exogenous shocks and various landscape sugar distributions to introduce external variations that allow us to see the effects of proto-institution absence or presence. In these experimental design decisions, we have held to the basic explanatory framework:

> In Analytical Sociology, the phenomena to be explained typically are important aggregate or macro outcomes such as network structures, segregation patterns, inequalities, cultural tastes and common ways of acting. The entities we refer to in the explanation typically are individuals, and the activities referred to are the behavior of these individuals [3].

These disruptions or nudges along the run of the model lead to insight about common pool resources [37, 38] that may provide important clues.

4 Coding a Social Primitives Model: The SPECscape Model

Just as a detailed party plan for 3-year olds may not survive contact with actual humans [39], a conceptual model may have difficulty with translation into a functional coding environment. We detail below work that we did to translate the social primitives concepts into a coded model capable of iterating dynamics suitable for graphing and comparison with our "as simple as possible" goal.

4.1 Agent to Agent Interaction

The 2D landscape version of our Social Primitives Experimental Cohort (SPEC) model features a range of agent, landscape, and meta-variables (the **SPECscape** model). Agents are very simple with basic landscape search and consume features and no means of communicating with each other directly. The ability (or inability) to communicate is a central dynamic for social modeling that needs to be clearly described given the impact of interaction (communication?) on model dynamics [25]. Efforts toward development of true primitives must show the kind of decomposition of capabilities so that communication is either clearly explained as pre-determined (coded into the model) or is one of the emergent properties that a model seeks to explain through changes in model variables that lead to communication.

Our SPECscape model agents possess an information exchange capability that is triggered under certain conditions. Agents cannot occupy a space that is already occupied by another agent.[2] If, however, an agent moves into a space adjacent to another agent in a North/South/East/West (NSEW) configuration (no diagonals to maintain all possible simplicity), then a sugar-level evaluation is made by the "supervising" code. If they both have enough sugar above a threshold, then a proto-institution is formed and each agent is considered to be a member of that new "virtual" entity.

An important distinction in the SPECscape model is the nature of the proto-institution. The primary function of this entity is to hold surplus sugar that is deposited by agents who then become its "members", manage a list of its member agents it, and dispense sugar when member agents require it. The proto-institution does not occupy a space on the landscape and does not interact with agents except as a secondary function. It may be useful to think about the proto-institution as a ledger of agents that have met the conditions required for membership (viz., surplus sugar

[2] Simultaneous occupation of a cell by more than one agent is often not specified or noted in any way. In NetLogo, for example, this can occur as matter of course given the unseen background coding.

and proximity to another agent that is, or can become a member of, an institution). As a ledger, it does not have the features or characteristics of an agent and does not occupy space on the 2D landscape. If sufficient sugar for a given agent to survive is not obtained by consuming 2D grid sugar and that agent is on the ledger, the proto-institution tops up the agent to enable it to continue to function. If the proto-institution runs out of sugar to dispense, the ledger is reset by releasing the agents. Subsequently, these free agents may happen upon more sugar or die, as the case may be. If they survive and thrive long enough, they can join another institution (new or existing) using the same mechanism as earlier.

Two agents meeting the above conditions lead to formation of a proto-institution. An unaffiliated agent may join an existing proto-institution if it is in proximity to an agent that is already a member of a proto-institution (and if that unaffiliated agent meets the excess sugar requirement). Once the joining agent becomes part of the proto-institution, it can draw on proto-institution resources as needed.

This description is a form of "communication as information exchange" but it does not take place independently in an agent-to-agent mode. The framework coding provides the exchange and enacting of a proto-institution with membership. As members, agents cannot look to see what other agents are doing or gain any information about those agents. The proto-institution manages its membership lists, sugar-reading, and allocating functions.

4.2 Agent-Environment Interaction

The environment is an essential part of the 2D SPECscape model. As such, the dynamics of agent with landscape must be carefully interrogated. As outlined above, at the beginning of a simulation run, agents are randomly placed on the 2D grid landscape where each cell can be occupied by just one agent at a time. Once on the grid, agents are capable of examining each (NSEW) cell adjacent (or further as a function of "vision"), identifying the cell with the most sugar (or, if more than one are highest, randomly choosing from among equal high sugar cells), and then moving there. The sugar is "consumed" (added to the agents existing sugar level) and the agent repeats the process. Agents who are members of a proto-institution can access the sugar as described earlier from any point in the 2D landscape.

4.3 Anthropomorphic Crossover

As noted previously, one of the considerable challenges in the design, coding, and execution of models is the language we use to describe what is happening. It is difficult *not* to describe agents, landscapes, and functions in anthropomorphic terms even when the functions are devoid of anything human (e.g., suggesting agents "consume

sugar" is shorthand for a direct mathematical calculation within fully proscribed parameters).

4.3.1 Perception of Cooperation: Ring World, Game of Life, and Simulated Cooperation

When independent agents in a model begin to move together, it is natural to make a simple leap in language and call that motion "cooperative" or even a "behavior" even where the agents lack cooperative features. It may be that this is something like cooperative behavior. The explanatory power arises from the linkage between simple rules in behavioral entities like ants, birds and fish, where proximity, following, and adjusting path functions at an individual level produce swarms, flocks, and schools, respectively [40]. Ring World produced "cliques" of agents that moved around the ring together and even sped up so that a clique would overtake and incorporate solitary agents moving around the ring (up to the point where the clique size exceeded the "vision forward" parameter coded into the agents' search function).

The question for clarification is whether or not this qualifies as cooperation in the way that we usually understand that term. We would argue that it is not cooperation (whether in Ring World, Game of Life, or other simple models) but must instead be understood as an artifact of the coding because agents in Ring World have no awareness of themselves or other agents and no means of exchanging information with the other agents. The only signal an agent gets is the result of the rule that they cannot move to a space occupied by another agent. Again, the agent doesn't know the space is occupied by another agent, the code is designed such that the search parameters excluded consideration of a space if another agent is there. It is possible to reproduce the appearance of real-world phenomena via a model that cannot be equated with the empirical dynamic.

4.3.2 Actual Cooperation

A human understanding of cooperation is an evolved attribute that requires memory of past interactions and the potential to envision possible future outcomes. We learn that cooperation may lead to improved odds of individual-level survival or other kinds of benefits while non-cooperation leads to lower returns, loss of benefits, or even penalties [41]. In humans, this is a highly developed function, one that can be transferred across generations via culture (societal memory) and appears to be encoded as a potential in brain circuitry via genetic mechanisms leading to predispositions physically that can be amplified in practice [42, 43].

4.3.3 Coding Equivalents

We may need to be clearer when describing code functions in human language. For example, agents in models like Ring World or even our SPECscape arrangement do not cooperate as we usually think of cooperation—they simply reference conditions such as a space being occupied or not. Cooperation requires communication. Referencing is a more code aligned way to think about what is happening. When we write that an agent "looks to see if a space is occupied by another agent" there are many anthropomorphic assumptions we translate in reading that line which are not present in the agents themselves. When we code a 2D grid, we are not creating a landscape as that term is normally used. We are mathematically specifying potential X/Y coordinates that will form the basis for additional X/Y coordinate calculations.

When building automobiles, the effects of forces like surface tension are not a high priority even though the forces are present. In contrast, when designing nano-devices surface tension becomes a major concern because of the role that these "small" forces play at that scale [44, 45]. If we are pursuing rigorous understanding of social primitives, it will be vital to be much more aware of how much conceptual transference, attribution, and effect arises from language concepts rather than from the models' direct dynamics. These considerations apply directly to the visualization aspects of models as well where our propensity to stitch across gaps in information so that we can assemble meaningful patterns can get in the way of analytic clarity, shaping our perceptions about what is real and what is imagined [46].

5 Running the SPECscape Model

The SPECscape model was run with three operating modes to determine the dynamic characteristics of the model. We begin with general observations about the Gini index measures of the models, use four different landscape resource starting points, introduce exogenous shocks at various stages of model evolution to test resilience, and end by looking at the effect on models both with and without proto-institution capability.

5.1 SPECscape Model and Gini Index

Overall, it is evident that in sugarscape instances where agents have the possibility of forming proto-institutions in partnership with other geographically proximal agents, the inequality in sugar (wealth) reduces over time, as seen in the Gini index of sugar level distributions gradually decreasing and reaching an asymptotically stable value, when aggregated across all model parameter combinations. In contrast, the Gini in sugarscape instances where agents do not have the possibility of forming proto-institutions, a similarly computed Gini value is both higher and has a significantly

higher level, with a greater amount of fluctuation over time, aggregated across all model parameter combinations (Fig. 2). Consequently, we infer that the availability of sugar reserves stored in proto-institutions are likely to serve as buffers that help agents that are unable to find sugar on the sugarscape in meeting their sugar-related needs and increasing their likelihood of longer-term survival.

5.2 Four Landscape Starting Points

We next investigated whether the overall pattern of wealth inequality's evolution across time, where sugarscape instances that have proto-institution formations perform better (i.e., have a lower Gini coefficient) than those sugarscape instances where such possibility is lacking, holds true when disaggregated across the four specific types of sugarscape instances:

- one where the sugar levels across cells in the sugarscape are randomly drawn from a uniform distribution;
- one where the sugar levels across cells are distributed in a more deterministic manner, with a central region having the highest concentration of sugar, and the concentration of sugar decreasing symmetrically in all directions, varying (reducing) inversely based on the cubed distance of each cell from the central region of highest sugar concentration;
- one where the sugar levels across cells are distributed in a more deterministic manner, with two regions of maximal sugar concentration that are located at the centers of the north-east and south-west quadrants of the sugarscape, and the concentration of sugar decreasing symmetrically in all directions, varying (reducing) inversely based on the cube of the combined distance of each cell from the two regions of highest sugar concentration;
- one where the sugar levels across cells are distributed in a more deterministic manner, with four regions of maximal sugar concentration that are located at the centers of each of the four quadrants of the sugarscape, and the concentration of sugar decreasing symmetrically in all directions, varying (reducing) inversely based on the cube of the combined distance of each cell from the two regions of highest sugar concentration.

From the subplots, we infer that

- the overall pattern of sugarscape instances with proto-institutions having lower sugar level inequality, when compared to sugarscape instances without proto-institutions, holds true across the four types of sugarscape;
- the variation in the Gini coefficients, aggregated across all model combinations within each specific type of sugarscape is the least in the case of the 4-quadrant sugarscape instances, followed by the 2-quadrant sugarscape instances, and single area of concentration sugarscape instances, followed by the randomly-distribution sugarscape instances; the most stable, lowest inequality sugarscape instances

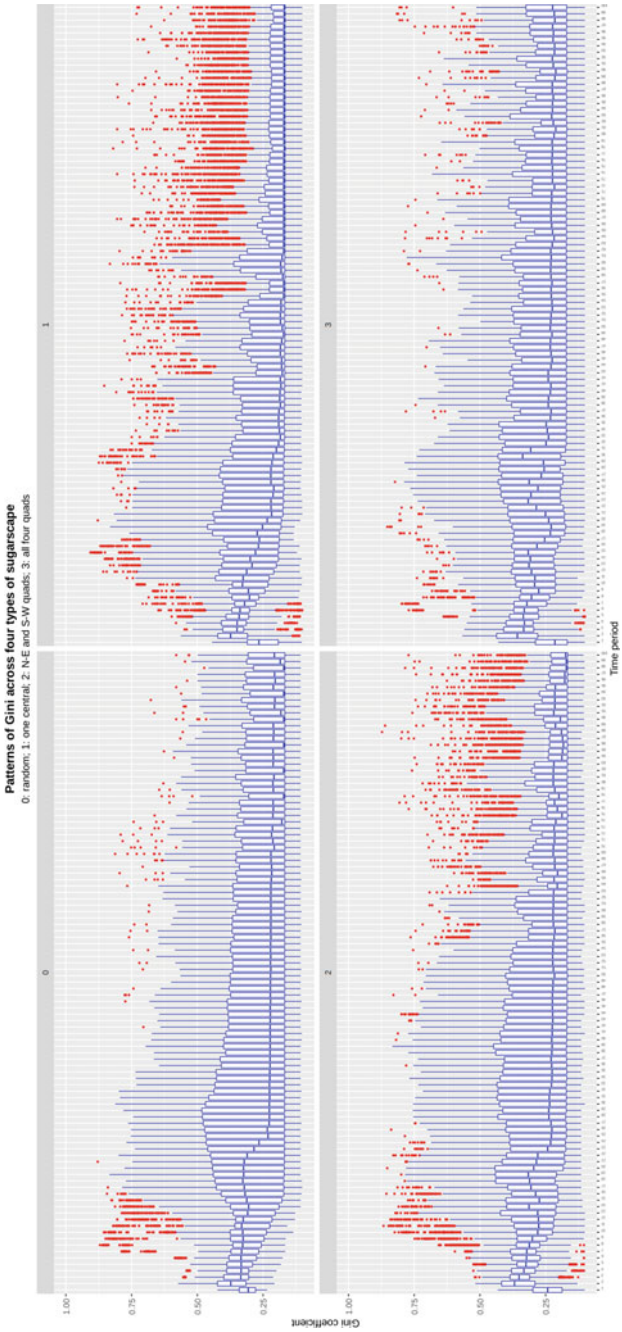


Fig. 2 Agent Ginis when landscape sugar value patterns are changed without proto-institutions

are to be found in the four-quadrant models where proto-institutions are present (Fig. 3).

5.3 The Effect of Exogenous Shocks at Various Stages of Model Evolution

Next, we analyzed the effect of environmental shock, represented in the form of a sudden decrease in the sugar level of every single cell in the sugarscape, affected the evolution of sugar (wealth) inequality in the sugarscape instances aggregated across all combinations of model parameters, but separated into three groups, one for each type of shock.

In the first group, no shock was present. In the second group, a single shock was delivered after roughly 1/3 of time periods have elapsed (so out of 100, at the end of time period 33) and in the third, a single shock was delivered after roughly 2/3 of time periods have elapsed (so out of 100, at the end of time period 66). We found that condition (a), wherein no shock was delivered, produced the most stable and lowest inequality in sugarscape agents' sugar levels followed by (b), and then by (c) (Fig. 4).

It appears that while having no shock produces the best outcome, having a shock earlier in their lifetime appears to allow sugarscape agents to recover and reach the same state as those in condition (a) during the last ~20 time periods. By contrast, when there are agents in sugarscape instances in condition (c), where a shock was delivered at the beginning of the last 1/3rd of the sugarscape instances' lifetime, the sugarscape instances' agents, as a whole, continued to experience fluctuations in both aggregate inequality, and in the asymptotic behavior prior to the end of the sugarscape instances' lifetime (of 100 periods). This provides further evidence that sugarscape instances require a long time to recover from the time when they experience shock, to settle down into a more stable distribution of sugar levels among their agents.

5.4 Models with and without Proto-institutions

We next sought to determine whether the patterns of perturbation and return to stability in the sugarscape instances agents' sugar levels' inequalities would differ between the conditions of proto-institutions could not be formed versus those where proto-institutions could be formed. We found that across all three shock types, sugarscape instances where proto-institutions could not be formed performed worse than those where proto-institutions could be formed. Additionally, these differences among the three shock types are not strongly distinguishable from each other, in sugarscape instances where proto-institutions could be formed (Fig. 5).

To verify the results intuited via the graphs, we performed a pooled panel regression analysis on the dataset consisting of agents' Gini coefficients across time as

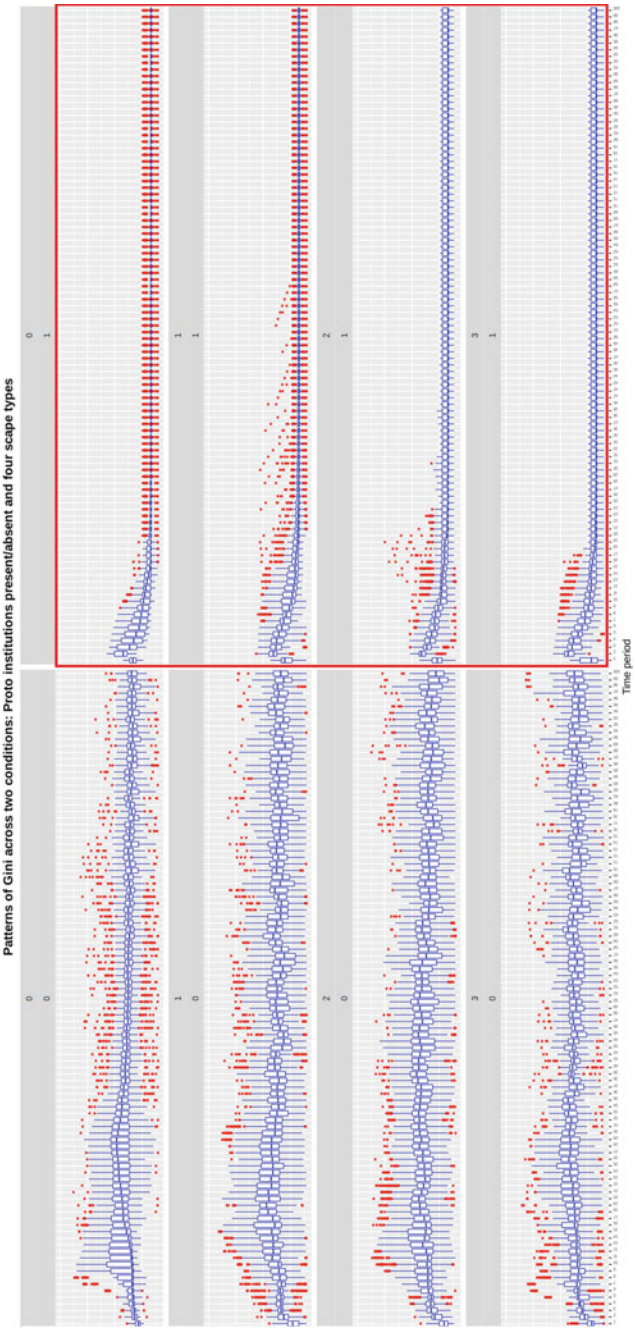


Fig. 3 Agent Ginis when landscape sugar patterns are changed and proto-institutions present (red box)

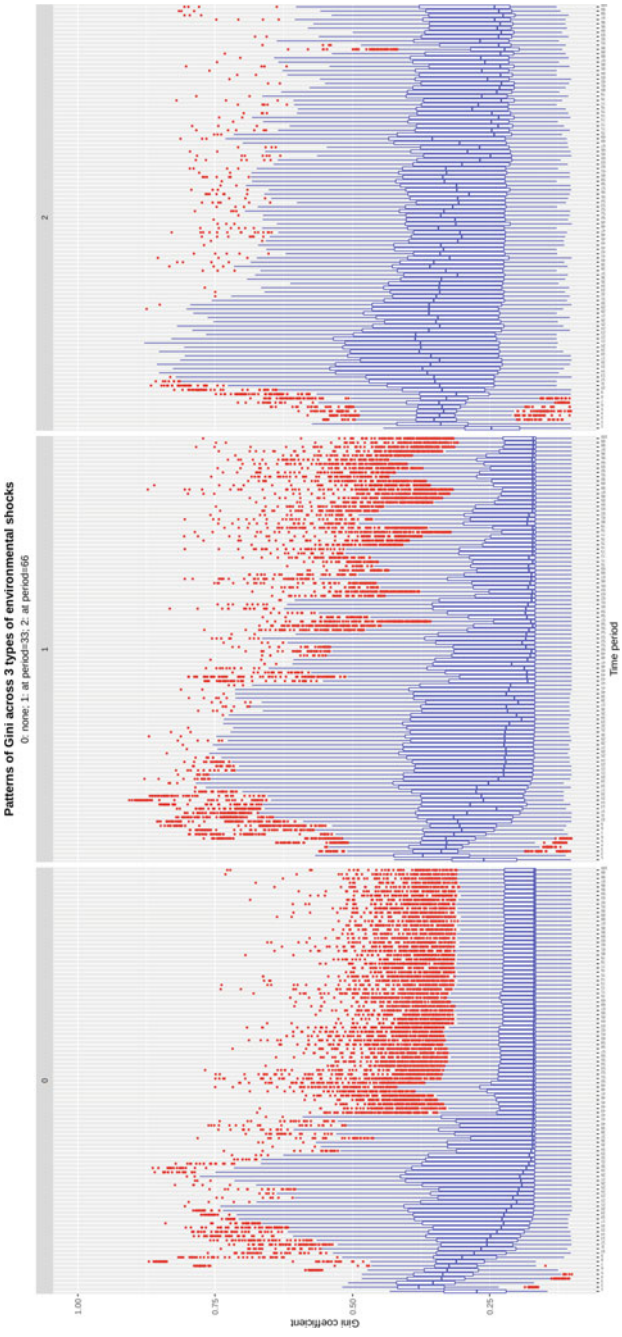


Fig. 4 Agent Ginis with exogenous shock (shortage) at beginning, time step 33, and time step 66 (of 100)

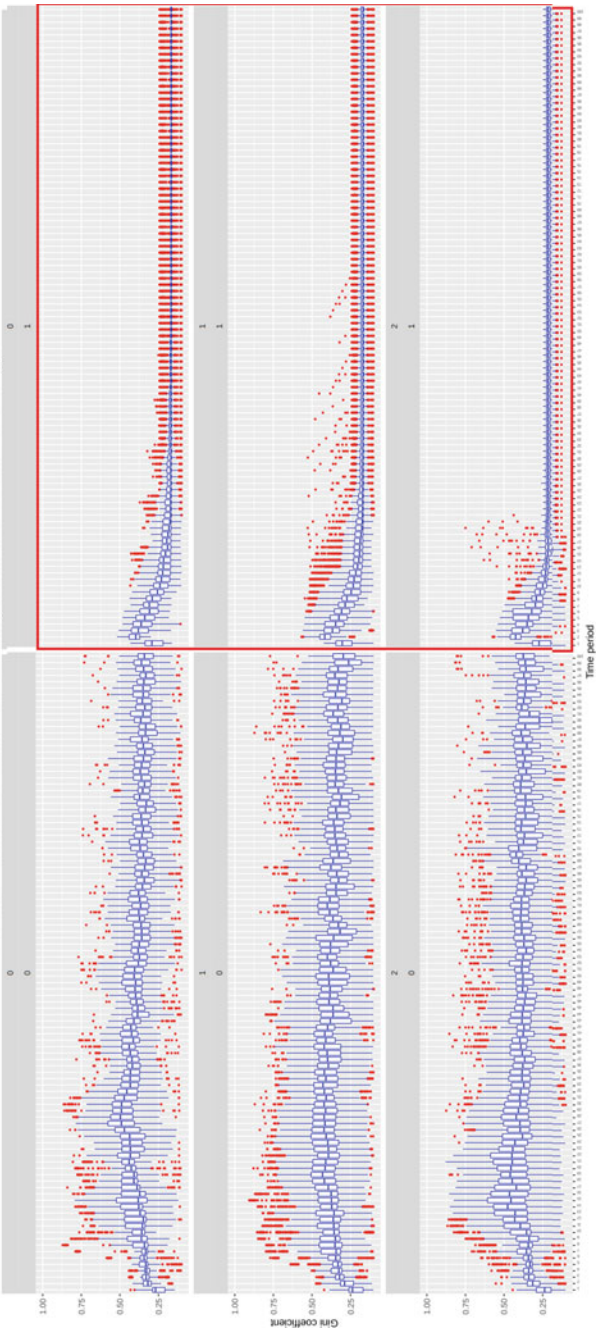


Fig. 5 Agent Ginis with exogenous shock (shortage) at beginning, time step 33, and time step 66 (of 100) but proto-institutions are present (red box)

the outcome variable, and all of the agent-based model's parameters as the predictor variables. Since we have different sugarscape instances, with different configurations and agent populations, for each combination of our simulation parameters, we analyzed the data using a pooled panel regression approach.

The following is the output of the pooled panel regression model. From the results, it is evident that (Fig. 6):

- the patterns of evolution agents' sugar levels' Ginis in each of the three non-random-distribution-of-sugar sugarscape configurations are significantly different from the patterns observed in the random sugarscape;
- the patterns of evolution agents' sugar levels' Ginis in both types of environmental shock situations are significantly different from the evolution of Ginis in the baseline category;

```
Unbalanced Panel: n = 129, T = 376-2600, N = 281295

Residuals:
    Min.   1st Qu.    Median   3rd Qu.      Max.
-0.394748 -0.070325  0.003037  0.066521  0.563736

Coefficients:
                                                  Estimate  Std. Error
(Intercept)                                     0.39474762  0.00115417
as.factor(ScapeType)1                          -0.04544488  0.00112339
as.factor(ScapeType)2                          -0.00980199  0.00118004
as.factor(ScapeType)3                          -0.03129348  0.00125977
as.factor(ProtoPresent)1                       -0.21676892  0.00142485
as.factor(ShockType)1                          -0.04136747  0.00109462
as.factor(ShockType)2                          -0.01321831  0.00112347
as.factor(ProtoPresent)1:as.factor(ShockType)1  0.02489080  0.00131853
as.factor(ProtoPresent)1:as.factor(ShockType)2 -0.01020485  0.00146094
as.factor(ScapeType)1:as.factor(ProtoPresent)1  0.05204196  0.00147594
as.factor(ScapeType)2:as.factor(ProtoPresent)1 -0.00030962  0.00150192
as.factor(ScapeType)3:as.factor(ProtoPresent)1  0.00172500  0.00164786

                                                  t-value  Pr(>|t|)
(Intercept)                                     342.0194 < 2.2e-16 ***
as.factor(ScapeType)1                          -40.4532 < 2.2e-16 ***
as.factor(ScapeType)2                          -8.3065 < 2.2e-16 ***
as.factor(ScapeType)3                          -24.8406 < 2.2e-16 ***
as.factor(ProtoPresent)1                       -152.1349 < 2.2e-16 ***
as.factor(ShockType)1                          -37.7917 < 2.2e-16 ***
as.factor(ShockType)2                          -11.7656 < 2.2e-16 ***
as.factor(ProtoPresent)1:as.factor(ShockType)1  18.8777 < 2.2e-16 ***
as.factor(ProtoPresent)1:as.factor(ShockType)2  -6.9851 2.852e-12 ***
as.factor(ScapeType)1:as.factor(ProtoPresent)1  35.2602 < 2.2e-16 ***
as.factor(ScapeType)2:as.factor(ProtoPresent)1  -0.2061   0.8367
as.factor(ScapeType)3:as.factor(ProtoPresent)1   1.0468   0.2952
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Total Sum of Squares:    6982.6
Residual Sum of Squares: 4510.7
R-Squared:      0.35401
Adj. R-Squared: 0.35399
F-statistic: 14013.3 on 11 and 281283 DF, p-value: < 2.22e-16
```

Fig. 6 Summary statistics of data generated by model

- these patterns are significantly different across sugarscape instances where proto-institution formation was present when compared to those sugarscape instances where proto-institution formation was absent. The other interactions included in our model were not significant.

6 Conclusion

These are early results that are exploring important dynamics present in the germinal evolution of cooperative behavior among agents that produces primitive institutions. While a great deal more work remains to be done, salient observations at this point in time can guide future work.

6.1 Elements of Collective Function

There is still a lot of research to be done in developing the SPECscape version of social primitives. One of the most persistent challenges in looking at non-primitive models is that model builders and domain experts are eager to apply their tools to actual human and social problems. This interest in direct equivalence means that more fundamental research about dynamics that are not immediately transferable are precluded [25]. While primitive elements may not apply directly to any real-world problem, they may turn out to be vital ingredients for more powerful, explanatory models that we would do well to understand:

> The origin of life question has played a central role in the biological sciences. Alas, the origin of social life has had much less attention. Such questions lie at the heart of understanding our world. How do we recognize social life? What are the minimum requirements for it to arise? What are the deep, common elements in a social system that transcend time and agents? Is social life inevitable? (Op. cit.)

Collective functions, whether as potential outcomes of other interactions or coded directly into a model, are root issues for analytical sociology, anthropology, political science, and other broad fields of human interaction. Paring down our agents and landscapes with a "how simple can we get" motive could promote new insights that are valuable beyond single domain settings.

6.2 Determining Anthropomorphic Leaps

Our language is a challenge in this endeavor. We describe "cliques" when agents have no means of interacting. Dynamic patterns appear to us like purposeful direction even when it is not possible for agents to exercise foresight. We take an empirical example, build a model, and then use the language of the empirical setting to bridge

to our model. In the process, we smuggle more meaning into our models than we should. These anthropomorphic leaps represent a habit, witnessed in many modeling papers and presentations, which is clearly useful but may amplify unintentionally the meaning of a model beyond what is responsible.

6.3 Determining a Parameter-Based Measure of Model Complexity

Another challenge is careful delineation of model parameters. This builds on the noted differentiation between parameters that operate the modeling framework and parameters that alter the behavioural capabilities of agents. Parameter accounting needs a similar clarification (see Appendix for full description of SPECscape model parameters).

Our hope is that these types of clarifications would put the work of exploring social primitives more squarely in the path of error rather than confusion. From that clarity, better approaches have a greater chance of being developed.

6.4 Possibilities for Increasing the Complexity of SPECscape Model

6.4.1 Expanding on SPECscape with SPECnet Approaches

One of the ways by which to make the SPECscape model more complex would be the addition of stronger agent-to-agent ties. Using a network model, another group of researchers has adopted the SPECscape objectives and parameters as much as possible but without the 2D grid. Instead, a network model is used where there are no spatial considerations, only relational ties. This approach raises many of the same concerns and possibilities that were discovered in earlier attempts to determine formal means of comparing two models—what the authors called "docking" [2]. This is an evolving project; equivalence and alignment were pursued through the development of common agent and proto-institution capabilities in all respects except where it was necessary to deviate on the "relational vs. spatial" designs of the model structure.

6.4.2 SPEC [Other]

There are other ways to develop aligned or equivalent models with modalities that may be different than either SPECscape or SPECnet. Some of these possibilities include hybrids of these two early types, the addition of agent characteristics, proto-institution interactions both with agents and each other. These different

approaches could arise from empirically informed modifications such as repetitive transactions that lead to cooperative behavior that is formally "institutionalized" via proto-institution membership akin to the "community lending societies" like credit unions/chit funds/etc. This approach would be in contrast to the purely network-based approach that is being explored in a sibling project—SPECnet. Future work could include development of a set of primitives that could lead to more realistic cooperative behavior. These may be sets of rules based on game theory, network dynamics, interactive genetic algorithms, or other dynamic modeling approaches.

Appendix

SPECscape model parameters, ranges of values, and description. Given the significant number of possible combinations, the Nearly Orthogonal Latin Hypercube is used to sample the behavior space of the model.

Parameter	Min value	Max value	Description
Side	5	100	Side of the sugarscape. Total cells = side * side—e.g., 10 side = 100 cells
Capacity	1	12	Carrying capacity within each cell. Assigned individually for each cell, drawn from a range [1, maxval]. Max. value is determined for each combination via the NOLH design
RegRate	1	5	Regrowth rate for each cell. Assigned using the same approach as capacity
Adensity	0.1	0.4	Agent population's density: no. of agents at the start of a simulation run for a given parameter combination = area of sugarscape * agent density, rounded to nearest integer
MtblRate	1	3	Metabolic rate: Number of units of sugar an agent needs to consume to stay "active" (i.e., not starve) during each time step. Assigned from a uniform distribution [1, maxval], where maxval is determined for each combination via the NOLH design
VsnRng	1	7	Vision Range: Number of cells in the NESW directions (von-Neumann) an agent can see on the sugarscape (cannot see beyond the boundary). Assigned from a uniform distribution [1, maxval], where maxval is determined for each combination via the NOLH design
InitSgLvl	1	12	Initial sugar level with which an agent enters the sugarscape. Assigned from a uniform distribution [1, maxval], where maxval is determined for each combination via the NOLH design

(continued)

(continued)

Parameter	Min value	Max value	Description
Birthrate	0	0.3	Number of agents added during each time step, in proportion to the number of agents alive at the beginning of the time step. Assigned from a uniform distribution [1, maxval], where maxval is determined for each combination via the NOLH design
InbndRt	0	0.45	In-bound Rate: Number of agents migrating inwards (currently indistinguishable from birth; in future birth rate can be used to determine agents produced via "interaction" between a "couple" of agents). Assigned from a uniform distribution [1, maxval], where maxval is determined for each combination via the NOLH design
OtbndRt	0	0.2	Out-bound Rate: Probability value determining the likelihood of an agent that is currently in a starvation state migrates out of the sugarscape. Assigned from a uniform distribution [1, maxval], where maxval is determined for each combination via the NOLH design
Threshold	1	2	Amount of excess sugar needed (constant for all agents in a given sugarscape) to form/join a proto-institution by forming an alliance with a geographically proximal agent, who also has excess sugar available to contribute
ResilienceTime	1	25	Number of time steps an agent can remain in starvation mode, prior to departing the sugarscape via death. Assigned individually for each agent, drawn from a range [1, maxval]. Max. value is determined for each combination via the NOLH design
ScapeType	0	3	One of four sugarscape types constant across all cells. Each cell's initial sugar level is initialized to: (0) a random value, (1) a central region with maximal sugar value, and sugar level falling off uniformly as an inverse power of 0.3 with distance (2) two regions of maximal sugar value, located in the north-east and south- west quadrants (3) four regions of maximal sugar value, located roughly in the middle of each of the four quadrants

(continued)

(continued)

Parameter	Min value	Max value	Description
StrvFeedPcnt	0	0.2	Starvation Mode Feed Percent: Value by which a starving agent's metabolic rate (amount of sugar needed to be consumed in each time step) reduces when the agent is in starvation mode. This represents a continual reduction in the metabolic rate, as time progresses—eventually leading to death if the agent does not migrate out, or find a sugar-rich cell to come out of starvation. Assigned individually for each agent, drawn from a range [1, maxval]. Max. value is determined for each combination via the NOLH design
ProtoPresent	0	1	Identifier for whether proto-institutions can form or not in a given sugarscape—does not change over time. Determined at the beginning of a simulation via NOLH design
VisRedPcnt	0.01	0.6	Vision Reduction Percentage: Value by which a starving agent's vision rate reduces when the agent is in starvation mode. This represents a continual reduction in the agent's ability to look around its neighborhood for sugar rate, as time progresses. So the longer an agent exists in starvation mode, the narrower is the space it can investigate for sugar, thereby creating a vicious cycle of sugar poverty. An agent's vision is restored to its original value when it finds a sugar-rich cell and exits starvation mode. Assigned individually for each agent, drawn from a range [1, maxval]. Max. value is determined for each combination via the NOLH design

References

1. Saam, N.J., Harrer, A.G.: Simulating norms, social inequality, and functional change in artificial societies. J. Artif. Soc. Soc. Simul. **2**, 1–2 (1999)
2. Axtell, R., Axelrod, R., Epstein, J.M., Cohen, M.D.: Aligning simulation models: a case study and results. Comput. Math. Organ. Theory **1**, 123–141 (1996). https://doi.org/10.1007/BF01299065
3. Keuschnigg, M., Lovsjö, N., Hedström, P.: Analytical sociology and computational social science. J. Comput. Soc. Sci. **1**, 3–14 (2018). https://doi.org/10.1007/s42001-017-0006-5
4. Peregrine, P.N.: Toward a Theory of Recurrent Social Formations. Santa Fe Institute, Santa Fe, New Mexico (2017)
5. Watts, D.J.: Should social science be more solution-oriented? Nat. Hum. Behav. **1**, 0015 (2017). https://doi.org/10.1038/s41562-016-0015
6. Katz, M.B.: The People of Hamilton Canada West: Family and Class in a Mid 19th Century City. Harvard University Press, Cambridge, Massachusetts (1975)

7. Friesen, Mudigonda: Institutional emergence and the persistence of inequality in Hamilton, ON 1851–1861. In: Proceedings of the Computational Social Sciences Society of the Americas, Santa Fe (2018)
8. Sampson, R.J., Graif, C.: Neighborhood social capital as differential social organization resident and leadership dimensions. Am. Behav. Sci. **52**, 1579–1605 (2009). https://doi.org/10.1177/0002764209331527
9. Blau, P.: Inequality and Heterogeneity: A Primitive Theory of Social Structure. Free Press, New York (1977)
10. Forbes, A., Wainwright, S.P.: On the methodological, theoretical and philosophical context of health inequalities research: a critique. Soc. Sci. Med. **53**, 801–816 (2001). https://doi.org/10.1016/S0277-9536(00)00383-X
11. Rousseau, J.-J.: A Discourse on Inequality. Penguin Classics, Harmondsworth, Middlesex, England; New York, N.Y., U.S.A (1985)
12. Sen, A.: Inequality Reexamined. Harvard University Press, New York (1995)
13. Walker, R.: Consensual approaches to the definition of poverty: towards an alternative methodology. J. Soc. Policy **16**, 213–226 (1987). https://doi.org/10.1017/S0047279400020389
14. Bourdieu, P.: Language and Symbolic Power. Harvard University Press, Cambridge, Mass (1993)
15. Bourdieu, P., Wacquant, L.: An Invitation to Reflexive Sociology. University of Chicago Press (1992)
16. Sakoda, J.M.: The checkerboard model of social interaction. J. Math. Sociol. **1**, 119–132 (1971). https://doi.org/10.1080/0022250X.1971.9989791
17. Schelling, T.C.: Dynamic models of segregation. J. Math. Sociol. **1**, 143–186 (1971). https://doi.org/10.1080/0022250X.1971.9989794
18. Takács, K., Squazzoni, F.: High standards enhance inequality in idealized labor markets. J. Artif. Soc. Soc. Simul. **18**, 2 (2015). https://doi.org/10.18564/jasss.2940
19. Tesfatsion, L.: Agent-based computational economics: growing economies from the bottom up. Artif. Life **8**, 55–82 (2002). https://doi.org/10.1162/106454602753694765
20. Epstein, J.M., Axtell, R.L.: Growing Artificial Societies: Social Science from the Bottom Up. Brookings Institution Press, Washington, DC (1996)
21. Axtell, R.L.: 120 million agents self-organize into 6 million firms: a model of the US private sector. In: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, pp. 806–816. International Foundation for Autonomous Agents and Multiagent Systems (2016)
22. Axtell, R.: Endogenous firm dynamics and labor flows via heterogeneous agents. In: Handbook of Computational Economics, pp. 157–213. Elsevier (2018)
23. Bak, P., Chen, K., Creutz, M.: Self-organized criticality in the 'Game of Life'. Nature **342**, 780–782 (1989). https://doi.org/10.1038/342780a0
24. Reia, S.M., Kinouchi, O.: Conway's game of life is a near-critical metastable state in the multiverse of cellular automata. Phys. Rev. E **89**, 052123 (2014). https://doi.org/10.1103/PhysRevE.89.052123
25. Miller, J.H., Page, S.E.: Complex Adaptive Systems: An Introduction to Computational Models of Social Life. Princeton University Press (2007)
26. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998). https://doi.org/10.1038/30918
27. May, R.: Simple mathematical-models with very complicated dynamics. Nature **261**, 459–467 (1976). https://doi.org/10.1038/261459a0
28. Llanos, E.J., Leal, W., Luu, D.H., Jost, J., Stadler, P.F., Restrepo, G.: Exploration of the chemical space and its three historical regimes. Proc. Natl. Acad. Sci. 201816039 (2019). https://doi.org/10.1073/pnas.1816039116
29. Holland, J.: Complexity: A Very Short Introduction. Oxford University Press (2014)
30. Institute of Medicine: Assessing the Use of Agent-Based Models for Tobacco Regulation (2015). https://doi.org/10.17226/19018

31. Von Neumann, J., Burks, A.W., Arthur W.: Theory of Self-reproducing Automata. University of Illinois Press, Urbana (1966)
32. Miller, I.D., Cupchik, G.: A Synthetic World Population for Agent-Based Social Simulation (2015). https://www.researchgate.net/publication/309287649_A_Synthetic_World_Population_for_Agent-Based_Social_Simulation
33. Wareham, T., Vardy, A.: Putting it together: the computational complexity of designing robot controllers and environments for distributed construction. Swarm Intell. **12**, 111–128 (2018). https://doi.org/10.1007/s11721-017-0152-7
34. Goldberg, D.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Boston, MA (1989)
35. Holland, J.H.: New Perspectives in Nonlinearity or What to Do When the Whole Is More than the Sum of Its Parts. PSA Proc. Bienn. Meet. Philos. Sci. Assoc. **1976**, 240–255 (1976)
36. Holland, J.H.: Genetic algorithms. Sci. Am. **267**, 66–73 (1992)
37. Lohmann, R.A.: The Ostroms' Commons Revisited. Nonprofit Volunt. Sect. Q. 0899764016643613 (2016). https://doi.org/10.1177/0899764016643613
38. Raghavan, S.V.: Being a Force for Good with Advanced Analytics (2018)
39. Snowden, D., Boone, M.E.: Leader's framework for decision making. Harv. Bus. Rev. **10** (2007)
40. Slavkov, I., Carrillo-Zapata, D., Carranza, N., Diego, X., Jansson, F., Kaandorp, J., Hauert, S., Sharpe, J.: Morphogenesis in robot swarms. Sci. Robot. **3**, aau9178 (2018). https://doi.org/10.1126/scirobotics.aau9178
41. Sennett, R.: Together: The Rituals, Pleasures and Politics of Cooperation. Allen Lane, London (2012)
42. Philibert, R., Erwin, C.: A review of epigenetic markers of tobacco and alcohol consumption. Behav. Sci. Law. **33**, 675–690 (2015). https://doi.org/10.1002/bsl.2202
43. Roth, T.L., David Sweatt, J.: Annual research review: epigenetic mechanisms and environmental shaping of the brain during sensitive periods of development. J. Child Psychol. Psychiatry **52**, 398–408 (2011). https://doi.org/10.1111/j.1469-7610.2010.02282.x
44. Ardito, R., Frangi, A., Corigliano, A., De Masi, B., Cazzaniga, G.: The effect of nano-scale interaction forces on the premature pull-in of real-life micro-electro-mechanical systems. Microelectron. Reliab. **52**, 271–281 (2012). https://doi.org/10.1016/j.microrel.2011.08.021
45. Cho, Y.-S., Han, H., Kim, W.-D.: Numerical analysis of the adhesive forces in nano-scale structure. J. Bionic Eng. **3**, 209–216 (2006). https://doi.org/10.1016/S1672-6529(07)60004-3
46. Landy, D., Guay, B., Marghetis, T.: Bias and ignorance in demographic perception. Psychon. Bull. Rev. **25**, 1606–1618 (2018). https://doi.org/10.3758/s13423-017-1360-2

Capturing the Production of Innovative Ideas: An Online Social Network Experiment and "Idea Geography" Visualization

Yiding Cao, Yingjun Dong, Minjun Kim, Neil G. MacLaren, Ankita Kulkarni, Shelley D. Dionne, Francis J. Yammarino, and Hiroki Sayama

Abstract Collective design and innovation are crucial in organizations. To investigate how the collective design and innovation processes would be affected by the diversity of knowledge and background of collective individual members, we conducted three collaborative design task experiments which involved nearly 300 participants who worked together anonymously in a social network structure using a custom-made computer-mediated collaboration platform. We compared the idea generation activity among three different background distribution conditions (clustered, random, and dispersed) with the help of the "doc2vec" text representation machine learning algorithm. We also developed a new method called "Idea Geography" to visualize the idea utility terrain on a 2D problem domain. The results showed that groups with random background allocation tended to produce the best design idea with the highest utility values. It was also suggested that the diversity of participants' backgrounds distribution on the network might interact with each other to affect the diversity of ideas generated. The proposed idea geography successfully visualized that the collective design processes did find the high utility area through exploration and exploitation in collaborative work.

1 Introduction

Collective design and innovation have been an important research subject in social sciences and engineering, because its processes are needed for successful solution development for many real-world problems in an organization [9, 14]. It is there-

Y. Cao (✉) · Y. Dong · M. Kim · N. G. MacLaren · A. Kulkarni · S. D. Dionne · F. J. Yammarino · H. Sayama (✉)
Binghamton University, State University of New York, Binghamton, NY 13902, USA
e-mail: ycao20@binghamton.edu

H. Sayama
e-mail: sayama@binghamton.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_20

fore essential to consider how to improve the quality and efficiency of the collective design and innovation processes. Large-scale design processes that occur in the collective design and innovation involve interaction and interdependence among multiple individuals with task-related diversity [6, 21]. In particular, the interdisciplinary background within teams has been shown to be positively correlated with both quantitative and qualitative task performance [8, 10, 19]. This suggests that investigating the effects of expertise on collective performance would be a promising area of further research to improve the quality of collective design and innovation.

The organizational structure under which the collective design proceeds is usually complex, which makes it harder to investigate the collective dynamic performance in realistic organizational settings [4, 11, 16]. Previous studies on this problem were limited in several ways: (1) the network size was significantly smaller than that of most real-world collective design cases, (2) the duration of the collective task was significantly shorter than that of most real-world collective design cases, and (3) the tasks used in models or experiments were quite simple and not open-ended [2, 15, 22].

The objective of the present study is to experimentally investigate how the diversity of background of individual members will affect the effectiveness of design and innovation processes at collective levels. Our approach involves a combination of theoretical agent-based simulation models [5, 6, 21] and online human-subject social network experiments. This paper will describe our online social network experiments and report some initial findings.

Previous human subject studies typically evaluated team performance by basic variables such as the number of ideas generated, average scores, and win rates [20, 21]. In contrast, the present study not only measured such basic variables, like the number of posted ideas and the utility score of final ideas, but also measured quantitative similarity between produced ideas using "doc2vec," a text embedding machine learning algorithm [13]. We used the doc2vec algorithm to convert text-format design ideas to numerical vectors and performed further quantitative analyses. We also developed a new method called "Idea Geography" to visualize the idea utility terrain on a 2D problem domain. The utility terrain visualized with Idea Geography can help determine which region(s) in the problem space would be more (or less) promising for further exploration.

The rest of the paper is structured as follows. Section 2 introduces details of three online human-subject social network experiments we have conducted, including experimental procedures, data collection, and task descriptions for the three experimental sessions. Section 3 describes data analysis processes and methods, including the "Idea Geography" method we developed. Section 4 presents the results. Section 5 discusses the findings and concludes the paper with future directions.

2 Online Social Network Experiment

We designed and conducted three online experiments using a custom-made web-based computer-mediated collaboration (CMC) platform with an interface is similar to Twitter. This platform was implemented using Python and Flask. This platform allows participants to submit ideas in response to the assigned design task, see other participants' ideas, and add comments and "Likes" to those ideas. We recruited a multidisciplinary group of students at a mid-size US public university in the Fall 2018 and Spring 2019 semesters to participate in the experiments. Participants were undergraduate/graduate students majoring in Engineering or Management. They were allowed to sign up for one experimental session (i.e, they were not allowed to take part in this experiment more than once). Each experimental session involved 64~77 participants who worked on an open-ended collective design task for 2 weeks.

2.1 Experimental Procedure

To participate in an experimental session, participants were required to fill out an experimental registration form to provide their academic major and a written description of why they selected their major, as well as their academic knowledge, technical skills, career interest areas, hobbies or extracurricular activities, and/or any other information related to their background (this information is called simply "background" in this paper). This narrative information was converted into a numerical vector using the doc2vec algorithm. The academic major and vectors representing the background characteristics of the participants were used to allocate the participants into the following three groups: (1) spatially clustered, i.e., participants with similar background were placed together as social neighbors; (2) randomly distributed, i.e., participants were randomly placed regardless of their background; and (3) dispersed, i.e., participants with different background were placed together as social neighbors. These groups were configured to be similar to each other in terms of the amount of within-group background variations; they differed only with regard to spatial distributions of background variations. The underlying social network topology was a spatially clustered regular network made of 21~26 members with degree four, in all of the three groups. Participants could observe only their immediate neighbor's activities and would not directly see the activities of other nonadjacent participants.

At the beginning of each experimental session, participants were provided an overall objective of the collective design task and instructions of how to use the experimental platform. This online experimental session lasted for 2 weeks, during which participants were requested to log in to the experimental platform using anonymized usernames and spend at least 15 minutes each weekday, working on the assigned collective design task with their collaborators (i.e., anonymous neighbors in the social network). Their participation and actions were logged electronically in the server and monitored by the experimenters on a regular basis.

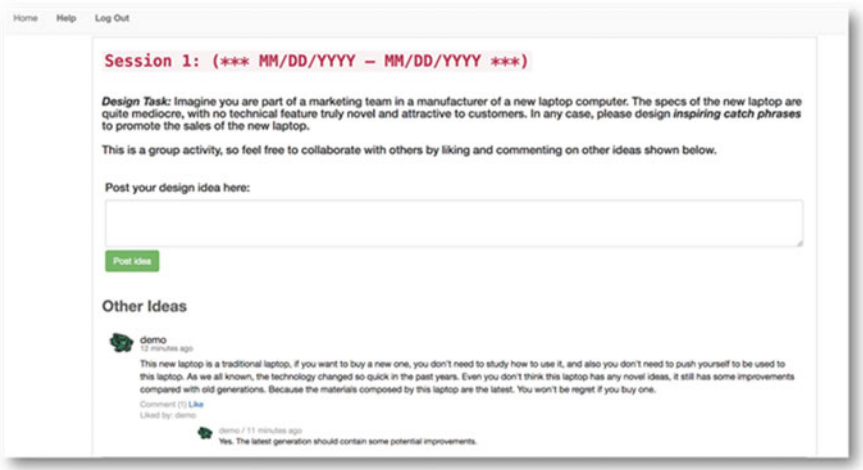


Fig. 1 Screen shot of the experimental platform

The collective design task description was displayed at the top of the experimental platform interface (Fig. 1). This interface allows participants to generate and post new ideas using the input box below the task description, read their collaborators' ideas in the timeline shown below the input box, and like and comment on others' ideas as well, like in Twitter. On each weekday during the experimental session, participants were requested to post ideas on the platform, discuss the task by reading, commenting, and liking their collaborators' ideas. By potentially utilizing their collaborators' ideas and comments, participants were expected to continuously elaborate and improve their idea quality over time on the platform.

After the 2-week experimental session was over, each participant was asked to submit an end-of-session survey form to provide three final ideas they chose for the assigned design task. These final ideas were later evaluated by third-party experts who were not involved in the experiments. These evaluation results were used to quantitatively assess the utility values of the final ideas made by each group. The participants were also asked in the survey form to answer questions about their overall experience, level of knowledge and understanding about their organizational neighbors, self-evaluation of their own contribution to the collaborative process, and personal evaluation of the final designs.

2.2 Collective Design Tasks

There were two different design tasks used for the experimental sessions. The tasks were open-ended textual design tasks with no obvious solutions immediately available to anyone.

The task for experiment session I was to create slogans, taglines, or catchphrases for marketing a laptop. This task worked successfully for students with diverse backgrounds in previous work [21]. This task was used in two experimental sessions conducted in two separate semesters (Fall 2018 and Spring 2019). In the Fall 2018 session I, participants were a relatively balanced mixture of Engineering and Management majors. A total of 64 students participated in this session. In the Spring 2019 session, participants were predominantly majoring in Management. A total of 66 students participated in this session. The final designs submitted by the participants who worked on this task were evaluated on a 5-point Likert scale by Marketing Ph.D. students who did not participate in the experiment.

The task for experiment session II was to write a story or a complete fiction within a word count limit. This experimental session was conducted once in Spring 2019. A total of 74 students participated in this session who were predominantly majoring in Management. The final designs submitted by the participants who worked on this task were evaluated on a 5-point Likert scale by professional staff members on campus who had the educational background and professional experience and expertise in creative writing and communication.

3 Data Analysis Methods

We applied a series of quantitative data analyses to investigate the effects of background distribution on group performance. The participants' activity records were utilized as the dataset for these analyses. First, the numbers of daily posts and submitted final ideas were measured as a characteristic of each group. We compared both time-series and distribution of numbers of daily posts among the three groups for each experimental session. We also measured the Euclidean distances between the posted ideas converted into numerical vectors using doc2vec and then created a distance matrix of the ideas. Furthermore, the ideas were visualized as points in a 2D problem space using the dimension-reduced idea vectors, which provided the basis to construct the "Idea Geography." Figure 2 shows an overview of our data analysis methods.

3.1 Doc2Vec

The data acquired from the registration forms, experimental records, and the end-of-session survey forms were mostly in plain text format, which would be hard to analyze using traditional quantitative analysis methods. Therefore, it was necessary to convert the text data to numerical data using text representation algorithms. There are many text embedding algorithms, such as Bag of Words (BOW) [7], Latent Dirichlet Allocation (LDA) [3], word2vec [17], and doc2vec [13] which was used in this study. Doc2vec, an adaptation of word2vec, is an unsupervised machine learning algorithm

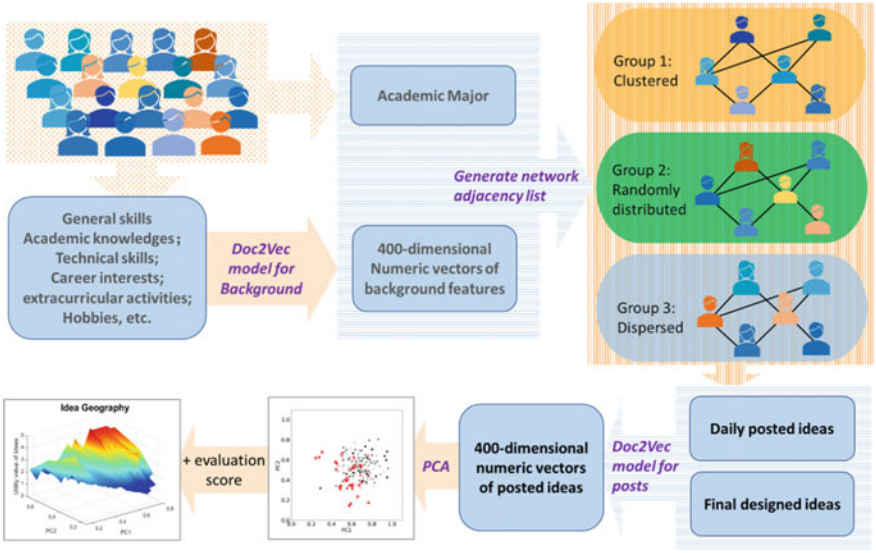


Fig. 2 An overview of data analysis methods

that can generate numerical vectors as a representation of sentences, paragraphs, or documents. Compared to other algorithms, doc2vec can provide a better text representation with a lower prediction error rate, because it can recognize the word ordering and semantics of words which are not accounted by other algorithms [12].

In this study, the whole set of written descriptions of background submitted by participants in the registration forms for each experimental session was used to build a doc2vec model of the participants' background for that session. The outputs of this doc2vec model were given in the form of 400-dimensional numerical vectors, which were combined with the self-reported academic major information to quantitatively represent the background features of the participants.

The daily posts and final ideas were also converted to 400-dimensional numerical vectors using doc2vec. The doc2vec model for the ideas posted during experiment session I was generated using the combined set of all the ideas obtained from both Fall 2018 and Spring 2019 session I. The doc2vec models for the ideas posted during experiment session II were generated using the sets of ideas obtained from the Spring 2019 session II.

3.2 Principal Component Analysis

Many of the 400 dimensions in the vectors obtained with doc2vec were undoubtedly correlated with each other, making the dataset highly redundant. Principal component analysis (PCA) was therefore applied to the set of idea vectors for each experimental

session to reduce dimensionality and visualize the idea distribution in a 2D space using the first two principal components (PC1, PC2). The 2D principal component space offered an efficient way to monitor the locations of ideas and also provided the basis to construct the "Idea Geography" method explained in the following section.

3.3 Idea Geography

The "Idea Geography" visualization method was developed specifically for this study. The average evaluation scores of the final ideas were used as the elevations at the 2D idea points to construct a utility terrain for each experiment session in the 2D principal component space, which we call "Idea Geography." From the landscape of idea geography, we can find mountain areas which represent regions in the problem space populated by ideas with high evaluation scores. We can also find valleys where the ideas would have low evaluation scores. In terms of physical geography and environmental studies, the terrain structure of a region is important for determining its suitability for human settlement [18], water flow patterns [1], and other properties of the region. In similar ways, with idea geography, the terrain of a problem space can intuitively reveal which areas in the problem space would have high or low utilities and would be suitable for further exploration. The visualization using idea geography may also help a leader or an organization manager to monitor the status of collaborative activities of a group.

4 Results

4.1 Experiment Session I: Catch Phrase Design

The end-of-session survey forms of Fall 2018 session I and Spring 2019 session I revealed that 89% and 75% of the participants, respectively, stated that they had a good overall experience in the experiment.

Figure 3 compares the numbers of daily ideas and final designs among three groups for the Fall 2018 and Spring 2019 session I. In both sessions, there was no statistically significant difference regarding the number of ideas among the three groups.

Figure 4 compares the average distance of ideas of each working day using the 400-dimensional idea vectors. In the Fall 2018 session I, a statistically significant difference was detected in terms of the average distance of ideas between Group 1 and Group 2 ($p = 0.012$), and Group 1 and Group 3 ($p = 0.0091$). In the Spring 2019 session I, there was no statistically significant difference regarding the average distance of ideas among the three groups.

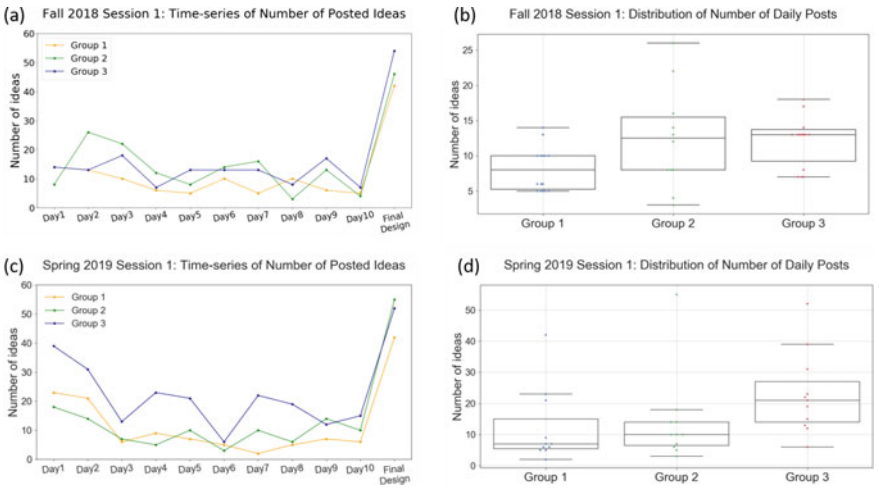


Fig. 3 Fall 2018 and Spring 2019 session I: number of daily posts and final ideas

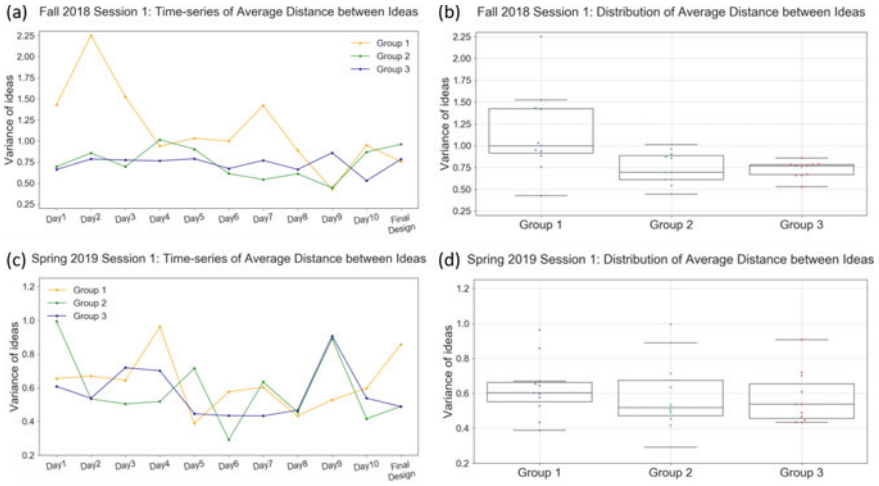


Fig. 4 Fall 2018 and Spring 2019 session I: average distance of ideas

The daily ideas and final designs of three groups were visualized as idea points in the 2D principal component problem space for two experimental sessions (Fig. 5). Group 1 with clustered background allocation produced a broader daily idea distribution but a more concentrated final design distribution than the other two groups. In Spring 2019 session I, three groups all have concentrated daily idea and final design distributions.

Figure 6 shows the idea geography visualizations of the Fall 2018 and Spring 2019 session I. There is a clearly identifiable utility mountain area in each idea geography, where most of the submitted final designs were concentrated. In both sessions, Group

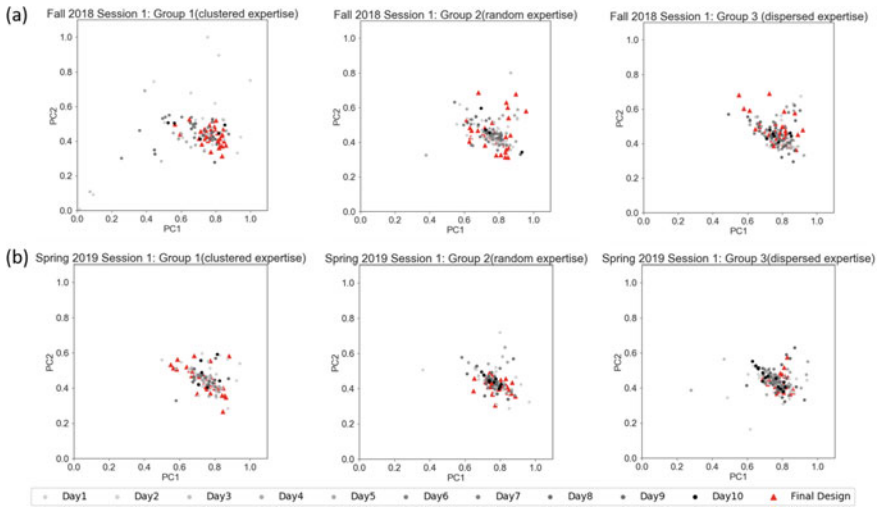


Fig. 5 Fall 2018 and Spring 2019 session I: distribution of idea points in 2D problem space. *Note*: Daily ideas from Day1 to Day10 are marked as circles with colors ranging from light gray to black; Final designs are marked as red triangles

2 with random background allocation produced the best final design with the highest utility value (Fall 2018: 4.75; Spring 2019: 4.667).

4.2 Experiment Session II: Story Design

In the end-of-session survey form of this experimental session, 71% of the participants stated that they had a good overall experience in the experiment.

Figure 7 compares the numbers of daily ideas and final designs among three groups. There was no statistically significant difference regarding the number of ideas among the three groups in this session.

Figure 8 compares the average distance of ideas of each working day in the Spring 2019 session II. There was no statistically significant difference detected regarding the average distance either.

The daily ideas and final designs generated by the three groups in the Spring 2019 session II were visualized in the 2D principal component problem space (Fig. 9). The three groups showed similar patterns of both the daily idea and final design distribution in the problem space.

Figure 10 shows the idea geography visualization of the Spring 2019 session II. We can see again that there is a clearly identifiable utility mountain area, where most of the submitted final designs were concentrated. Interestingly, Group 2 with random background allocation again produced the best final design with the highest utility value (5.0).

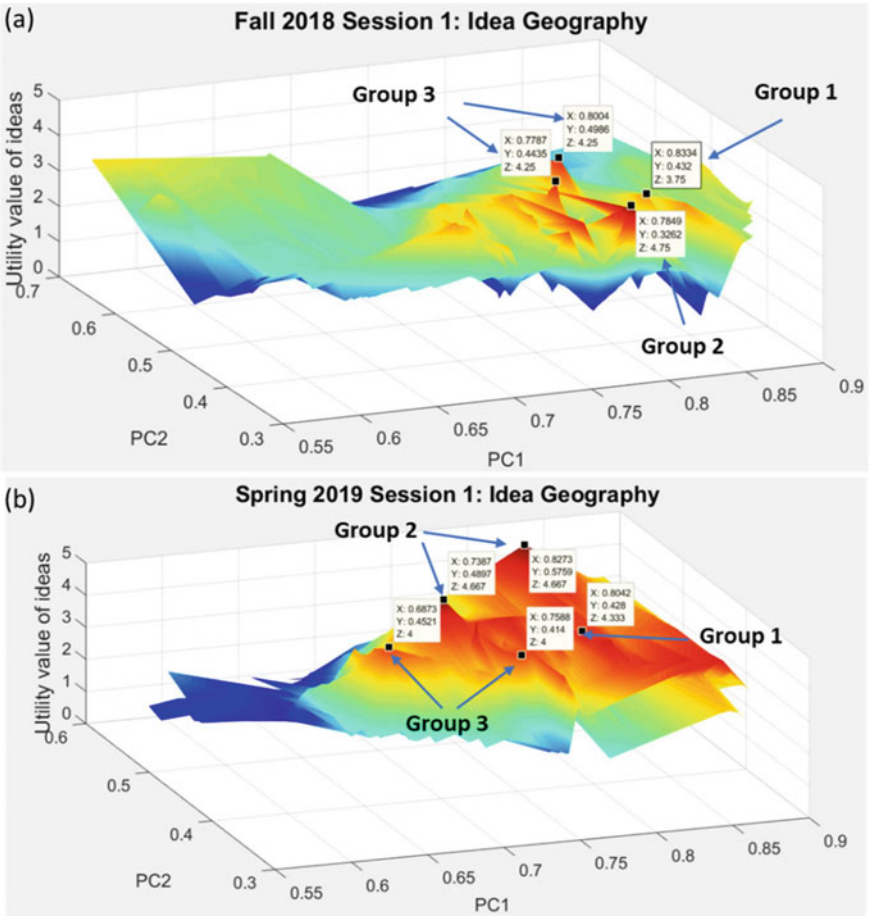


Fig. 6 Fall 2018 and Spring 2019 session I: Idea Geography

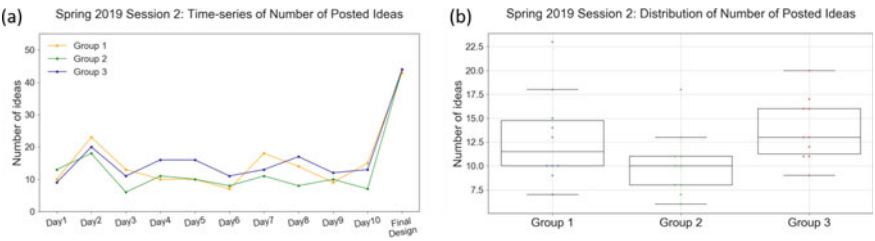


Fig. 7 Spring 2019 session II: number of daily posts and final ideas

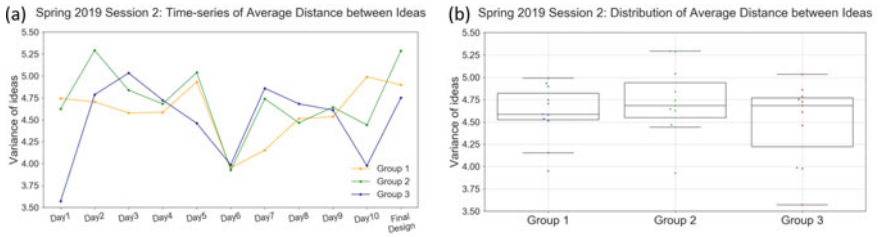


Fig. 8 Spring 2019 session II: average distance of ideas

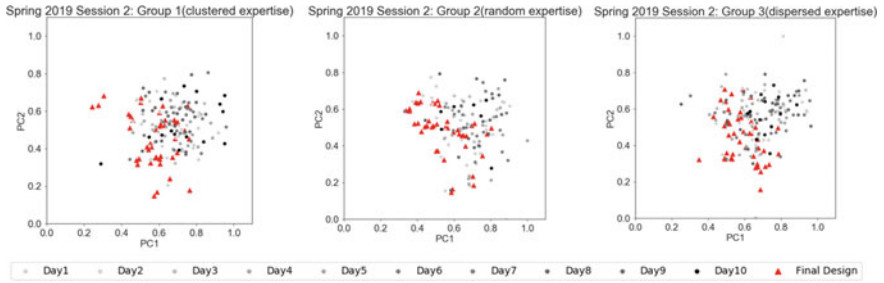


Fig. 9 Spring 2019 session II: Spring 2019 session II: distribution of idea points in 2D problem space. *Note*: Daily ideas from Day1 to Day10 are marked as circles with colors ranging from light gray to black; Final designs are marked as red triangles

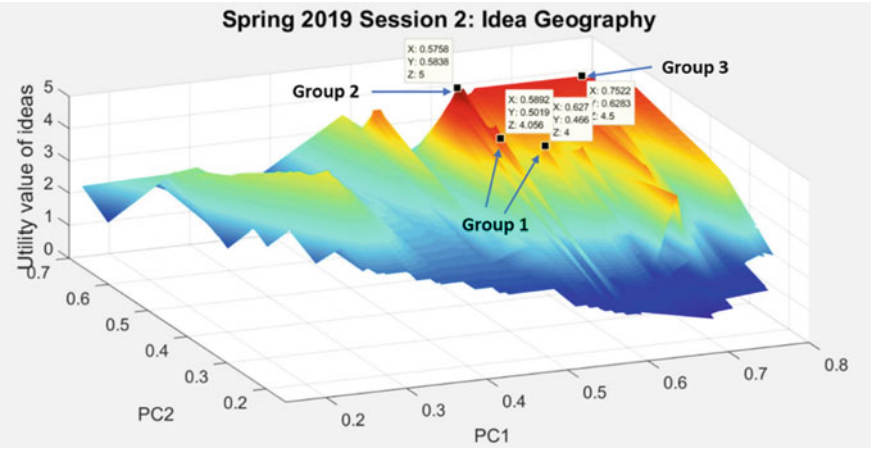


Fig. 10 Spring 2019 session II: Idea Geography

5 Discussions

Our results showed no statistically significant difference in the numbers of daily ideas among three groups in any of the three experimental sessions. This may imply that the activity level of individual participants was not affected by the allocation of background in the social network.

Meanwhile, a clear statistical significance was detected in terms of the average distance of ideas for the Fall 2018 session I, in which Group 1 (with clustered background allocation) produced more diverse ideas than the other two groups. However, this was not observed in the Spring 2019 session I. The key difference between Fall 2018 and Spring 2019 was in the variety of participants' backgrounds, i.e., Fall 2018 had a balanced mixture of Engineering and Management majors, while Spring 2019 was predominantly Management majors. We hypothesize that, when the participants' backgrounds were diverse and they were spatially clustered based on their similarity (this occurred only in Group 1 of the Fall 2018 Session I in our study), different parts of the network would explore the problem space in different directions without much mixture, and therefore, the average distance between generated ideas would naturally go larger. This also explains why the same phenomenon was not observed in Groups 2 and 3 in Fall 2018 or in any groups/sessions in Spring 2019.

The most interesting finding obtained so far was that, in all the experimental sessions for which idea geography was generated, Group 2 with random background allocation produced the best final design with the highest utility value. This seemingly puzzling observation may be explained by considering how much background diversity each participant was exposed to locally. Namely, in either Group 1 or Group 3, each participant would be connected to their neighbors that were relatively homogeneous background-wise (i.e., the neighbors should be similar to the focal participant in Group 1, while they should be the opposite of the focal participant in Group 3). Group 2 with random background allocation should have realized the most diverse local neighbors around each participant, which, we hypothesize, may have contributed to the enhancement of innovation search processes in Group 2.

Our experiments are still ongoing to collect more data. With additional experimental data, we hope to test these findings and hypotheses described above.

6 Conclusions

In this paper, we conducted a series of online human-subject experiments to examine and monitor the collective design and innovation processes on two different open-ended text design tasks. The doc2vec algorithm was used to quantify text-based information, including participants' backgrounds and ideas generated, which allowed for numerical characterization and control of similarities and differences between different backgrounds or ideas. The results we obtained so far indicated that, when participants with various backgrounds were randomly placed on the network, the group tended to find the best design ideas. The results also indicated potential

interaction between background distribution and the diversity of participants' backgrounds. More data are needed to confirm these observations and test our hypotheses to explain them.

This study also proposed the "Idea Geography" method, which was successfully demonstrated as an effective way to visualize the behavior and performance of the group's collaborative work. The fact that most of the submitted final designs clustered near the mountain areas in the idea geography indicates that the collective innovation and design processes were able to find the high utility regions through exploration and exploitation.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant #1734147.

References

1. Baker, N.T., Capel, P.D.: Environmental factors that influence the location of crop agriculture in the conterminous United States. U.S. Geological Survey Scientific Investigations Report, 2011-5108 (2011)
2. Becker, J., Brackbill, D., Centola, D.: Network dynamics of social influence in the wisdom of crowds. Proc. Natl. Acad. Sci. **114**(26), E5070–E5076 (2017)
3. Blei, D., Ng, A., Jordan, M.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
4. Braha, D., Bar-Yam, Y.: Topology of large-scale engineering problem-solving networks. Phys. Rev. E **69**(1), 016113 (2004)
5. Dionne, S.D., Sayama, H., Hao, C., Bush, B.J.: The role of leadership in shared mental model convergence and team performance improvement: an agent-based computational model. Leadersh. Q. **21**(6), 1035–1049 (2010)
6. Dionne, S.D., Sayama, H., Yammarino, F.J.: Diversity and social network structure in collective decision making: evolutionary perspectives with agent-based simulations. Complexity **2019**, 7591072 (2019)
7. Harris, Z.: Distributional structure. WORD **10**(2–3), 146–162 (1954)
8. Horwitz, S., Horwitz, I.: The effects of team diversity on team outcomes: a meta-analytic review of team demography. J. Manag. **33**, 987–1015 (2007)
9. Kerr, N.L., Tindale, R.S.: Group performance and decision making. Annu. Rev. Psychol. **55**, 623–655 (2004)
10. Koh, W.T.H.: Heterogeneous expertise and collective decision-making. Soc. Choice Welf. **30**(3), 457–473 (2008)
11. Kijkuit, B., van den Ende, J.: With a little help from our colleagues: a longitudinal study of social networks for innovation. Org. Stud. **31**(4), 451–479 (2010)
12. Lau, J.H., Baldwin, T.: An empirical evaluation of doc2vec with practical insights into document embedding generation. In: Proceedings of the 1st Workshop on Representation Learning for NLP, pp. 78–86, Berlin, Germany (2016)
13. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning (ICML 2014), pp. 1188–1196, Beijing, China (2014)
14. Lu, S.Y., ElMaraghy, W., Schuh, G., Wilhelm, R.: A scientific foundation of collaborative engineering. CIRP Ann. Manuf. Technol. **56**(2), 605–634 (2007)
15. Mason, W., Watts, D.J.: Collaborative learning in networks. Proc. Natl. Acad. Sci. **109**(3), 764–769 (2012)

16. McCubbins, M.D., Paturi, R., Weller, N.: Connected coordination network structure and group coordination. Am. Politics Res. **37**(5), 899–920 (2009)
17. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of phrases and their compositionality. Adv. Neural Inf. Process. Syst. (2013). arXiv:1310.4546v1
18. Olwig, K.R.: Recovering the substantive nature of landscape. Ann. A.A.G **86**(4), 630–653 (1996)
19. Salas, E., Rosen, M.A., DiazGranados, D.: Expertise-based intuition and decision making in organizations. J. Manag. **36**(4), 941–973 (2010)
20. Sapienza, A., Zeng, Y., Bessi, A., Lerman, K., Ferrara, E.: Individual performance in team-based online games. R. Soc. Open Sci. **5**(6), 180329 (2018)
21. Sayama, H., Dionne, S.D.: Studying collective human decision making and creativity with evolutionary computation. Artif. Life **21**(3), 379–393 (2015)
22. Woolley, A.W., Chabris, C.F., Pentland, A.: Evidence for a collective intelligence factor in the performance of human groups. Science **330**, 686–88 (2010)

An Agent-Based Model to Understand a Simple Theory of Mind: Belief Representation Systematic Approach (BRSA)

Zahrieh Yousefi, Dietmar Heinke, Ian Apperly, and Peer-Olaf Siebers

Abstract The natural ability of humans to mentally process others' behavior is central to their social life. This mental process relies largely on understanding the essential constituents of others' behavior such as their beliefs, desires and goals. This ability, called Theory of Mind (ToM), is the complex cognitive ability to attribute mental states to others, and to understand, predict, infer and justify people's actions by reasoning about their mental states such as their beliefs, desires and intentions. This study presents an original agent-based model for two simple levels of ToM ability. Our model aims to (i) develop a general set of basic processes underlying a simple ToM ability, and (ii) compare this process with Belief Representation Systematic Approach (BRSA). Introduced in an earlier study on an agent-based model for a false belief task, the BRSA is a robust cognitive framework for the processes of a simple ToM task, including many false belief tasks, and is comprised of four main cognitive phases: (*1) Perception, (2) Memory, (3) Reasoning beliefs and desires, and (4) Expressing others' beliefs and desires in an action.* The BRSA is capable of analyzing ToM tasks within a network of cognitive structures, which facilitates the constructing and deconstructing of ToM general processes in a classified scheme. The BRSA as an applicable analytical framework is able to filter unrelated abilities from ToM tasks, leading to more accuracy and transparency in the results of the experimental tasks. In addition, the BRSA has the potential to represent more complex ToM tasks through adjustments to the features within its phases.

Z. Yousefi (✉) · D. Heinke · I. Apperly
School of Psychology, University of Birmingham, Birmingham, UK

D. Heinke
e-mail: d.g.heinke@bham.ac.uk

I. Apperly
e-mail: i.a.apperly@bham.ac.uk

P.-O. Siebers
School of Computer Science, University of Nottingham, Nottingham, UK
e-mail: peer-olaf.siebers@nottingham.ac.uk

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_21

Keywords Theory of mind · Minimal theory of mind · Belief representation · Agent-based model

1 Introduction

People infer others' mental states in everyday life. They make inferences about others' unobserved mental states from their observed behavior. These inferences are often precise and indispensable for human social life. Infants start from tracking eye direction effortlessly and automatically while adults are competent in flexible and complex social reasoning [3]. From the simple to the complex, mental states' processes influence the actions people take to reach their goals. A large body of research suggests that humans are capable of Theory of Mind (ToM), the ability to take one's own and others' mental states such as beliefs and desires into account, and to understand, predict and infer people's actions by reasoning about their mental states. Correspondingly, people may reason backwards to infer others' beliefs and desires from their actions [7].

Theory of mind research has been considerably increased through various fields recently. However, the areas of confusion and gaps in the literature are increasing and a lack of standardization has been identified by researchers (e.g. [2, 27, 28]). The underlying processes of ToM are still under considerable debate. For example, some studies of ToM experimental tasks and methods have included abilities that are more than ToM ability [8] and such studies have therefore misinterpreted ToM processes. The problem arises because some researchers tacitly regard ToM as a single, indivisible process, some consider it as a single brain network and other researchers combine varieties of ToM into one process [27]. Schaafsma et al. [27] argue that a scientific concept of ToM requires a set of simpler processes rather than its current definition as the essence of a mental representation of minds which does not permit an easy breakdown into its basic components. Consequently, they suggest the reconstruction of a concept of ToM with the necessary links to its more basic processes. To achieve this, Schaafsma and her colleagues propose two steps: breaking down ToM and its associated concepts into cognitive components that describe more basic processes, and then reassembling these basic blocks into different features of ToM.

The aim of this paper is to develop a structured understanding of the principles of a simple ToM process and also analyze how this process is similar to the steps in the BRSA. For this purpose, an agent-based model has been implemented—a rather different approach from traditional experimental and cognitive methods. Agent-based models (ABMs) provide a computational platform to simulate individual or collective autonomous entities as agents and replicate their actions. Agent-based simulations are able to evaluate the effect of agents' interactions within the environment. ABMs are capable of representing the rules and relationships between individuals' mental states, their actions, the environment and the procedure that they use to infer

others' mental states on the micro level, while the macro level represents the aggregated effect of these actions, for example, how successful individuals perform in the environment. Cognitive science can greatly benefit from ABMs in the analysis of cognitive processes and social aspects of cognition through agents' interactions [30]. These potential characteristics of ABMs offer a reliable framework for understanding the underlying cognitive processes of ToM.

This study offers a systematic set of basic processes underlying a simple ToM ability and examines the consistency of this structure with the Belief Representation Systematic Approach (BRSA) in the false belief task model [32]. Yousefi et al. [32] define the BRSA as a "simple and robust approach that breaks down false belief tasks into four fundamental cognitive phases, including Perception, Memory, Reasoning beliefs and desires, and Expressing others' beliefs and desires in an action".

1.1 Theory of Mind

Theory of mind was defined as a system of inferences about the mental states of others that are not directly observed, and it can be used as a theory to make predictions about others' behavior; it was first developed by Premack and Woodruff in 1978. In an article, they asked "Does the chimpanzee have a theory of mind?" and they suggested that one has a theory of mind if he/she could impute mental states to self and others [24]. While psychologists frequently approve of this definition as a standard, some researchers raise questions. For example, Sharrock and Coulter ([28], p. 69) argue that experimental tasks "can only efficiently address the question of what specific capacities may legitimately be ascribed to non-human primates". Moreover, Apperly argues that "the appearance of consensus on what theory of mind is, and how it could be studied is misleading" ([2], p. 2).

Experimental research shows that language, memory and executive functions are critical in ToM competence (e.g. [1, 14, 22]). The link between executive functions and ToM ability, and particularly the role of inhibitory control in children and adults, has been studied extensively. Russell [26] argues that deficits in executive control in autistic children hinder the emergence and expression of their ToM ability while Wellman [31] states that ToM development corresponds with executive function but not directly. MacLeod ([21], p. 3) explains that "cognitive inhibition is the stopping or overriding of a mental process, in whole or in part, with or without intention. The mental process so influenced might be selective attention or memory retrieval or a host of other cognitive processes". Also, the study by Leslie and Polizzi [19] offers the mechanism for ToM. They explain that, to be able to present others' beliefs, it is necessary to inhibit the default content of their own true belief and change the focus of attention to the alternative belief. The concept of inhibition is identified through the algorithms of the model in this paper and will be considered in the discussion section.

1.2 Minimal Theory of Mind

Butterfill and Apperly [9] developed a distinctive minimal form of ToM cognition that involves representing "belief-like" states without any cognitive demands or conceptual sophistication. This is an elaborated version of the first system of their two systems account. The first system is fast and cognitively efficient and capable of tracking others' registration of an object rather than belief representation as such but inflexible and limited. They suggest that such a system may account for the success of some theory of mind tasks by human infants, some non-human animals such as chimpanzees and human adults under cognitive load, whereas the second system associates with cognitively demanding but flexible and slow processing. The second system exists in human adults, parallel with the first system.

Butterfill and Apperly's argument starts with a fundamental question "what could someone represent that would enable her to track, at least with limits, other's perception, knowledge states and beliefs including false beliefs?" ([9], p. 1). They then formulated four principles to answer this question. The first principle relates to "a minimal grasp of goal-directed action" ([9], p. 10). In the second principle, the concept of field relates to a set of objects at any given time that is determined by factors such as proximity, eye direction and barriers. The concept of encountering involves the relation between the agent and an object in the agent's field. The third principle, "registration", is the relation between an agent, an object and its location. The agent registers the location of an object as it encounters the object. A correct registration is a precondition for a successful goal-directed action. The fourth principle involves a shift to thinking of a successful registration as a causal factor for the agent's action.

They present an insightful example of a minimal theory of mind. Suppose Hannah is able to distinguish whether someone can see her while she is stealing from others. She wants to escape others' detection by "exploiting a fact about other's mental states (namely that they usually cannot see Hannah's acts of theft when Hannah does not have their eyes in view). Then Hannah has a theory of mind ability" ([9], p. 606). They conclude that Hannah is able to use others' visual perspective without any theory of mind that requires complicated cognitive ability.

1.3 Belief Representation Systematic Approach

The general framework for the underlying blocks and basic processes of belief presentation is called the Belief Representation Systematic Approach (BRSA), which is illustrated in Fig. 1. As the figure shows, the BRSA contains four cognitive phases including:

- *Perception phase*: collecting information about the others' perspective regarding the goal from the environment.

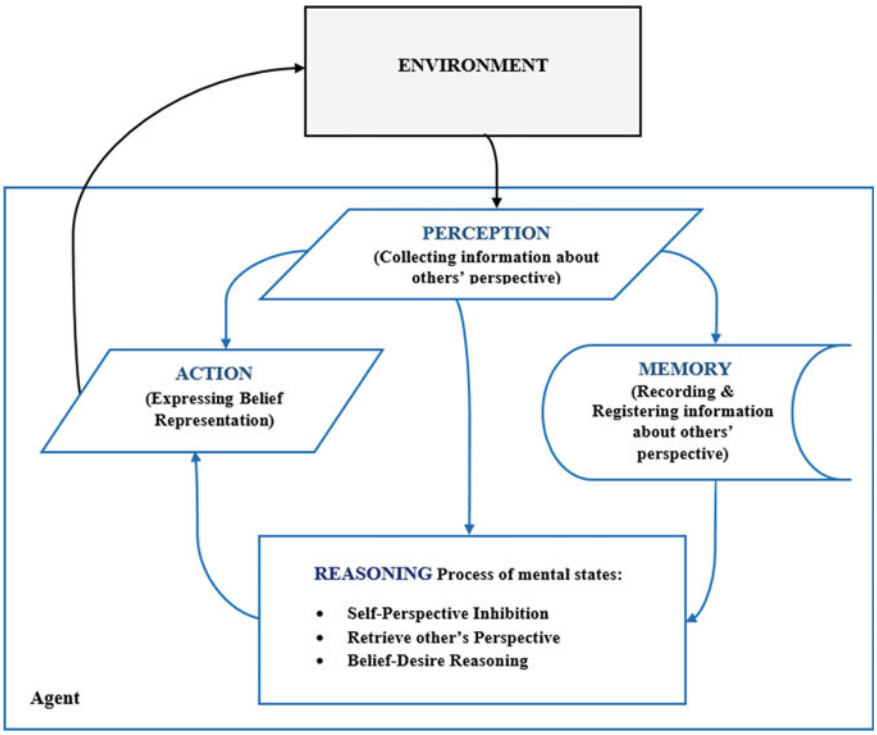


Fig. 1 Belief representation systematic approach (BRSA) [32]

- *Memory phase*: storing information about the others' perspective regarding the goal into short-term memory for use at the current time step.
- *Reasoning Process of Beliefs and Desires phase*: includes self-perspective inhibition, retrieving the protagonist's perspective data from memory, and selective processing of others' beliefs and desires subroutines. By the end of this phase, the process of understanding others' perspective is completed.
- *Action (Expressing others' mental states) phase*: expressing the understanding of others' beliefs and desires through an action.

2 Mental State Model

An agent-based model for a simple ToM ability is implemented in the Repast Simphony platform [25] called the Mental State Model (MSM). The MSM comprises interactions between three types of agents, representing different capabilities of understanding others' desires and beliefs with their heterogeneous set of rules: Infer agents with ToM ability, MinToM agents with minimal ToM ability and Control agents which do not possess ToM ability. The goal of agents in each time step is

to achieve a target. The detailed MSM Overview-Design-Details (ODD) protocol document, including the methodology and implementation of the model, is provided in Appendix.

2.1 Mental States (MS)

Two mutually exclusive mental states are designed for agents; an *active mental state* represents the necessity of achieving a target whereas a *passive mental state* signifies that agents are not able to achieve any target in the current time step. At the start of the simulation, all agents are active. However, their mental state changes upon achieving a target. Unless an agent achieves a target, its mental state remains active. The mental state of an agent that achieves a target changes to passive and it remains passive for two time steps.

Agents observe their neighborhood to search for targets. Active agents move to the nearest cell that contains a target while the Passive agents move to a cell with a target in its vicinity but which does not itself contain a target.

2.2 Egocentric Agents (Control Agents)

Control agents are egocentric simple reactive agents that ignore other agents and only consider their own mental states. Control agents collect information about the location of the nearest target in their field of view. Thus, they make decisions about achieving a target based on two factors; their belief about the location of the target, and their own mental state.

Control agents employ a target sensor and memory sensor at each time step, as shown in Table 1. The arrow and box diagram of the Control agents is illustrated in Fig. 2. The figure shows that Control agents collect the information about the target and make decisions about where to move to achieve a target.

2.3 Minimal Theory of Mind Agents (MinToM Agents)

MinToM agents possess a minimal theory of mind ability such that they are able to register the location of the target from other agents' perspectives as well as to take their own mental states into account. They are capable of tracking other agents' fields of view. They collect information about the target and also other agents' perspective regarding the location of the target. Figure 3 shows the arrow and box diagram of the MinToM agents; MinToM agents' procedure includes collecting information, registering others' perspective and an action to move toward a target.

Table 1 The agents' perceptions, understanding MS and memory capabilities

Agent	Perception (sensor)		Understanding MS			Memory		
	Target	Other agent	Own MS	Tracking others' field	Others' MS	Sensor memory	Short-term memory	Long-term memory
Control	✓		✓			✓		
MinToM	✓	✓	✓	✓		✓	✓	
Infer	✓	✓	✓	✓	✓	✓	✓	✓

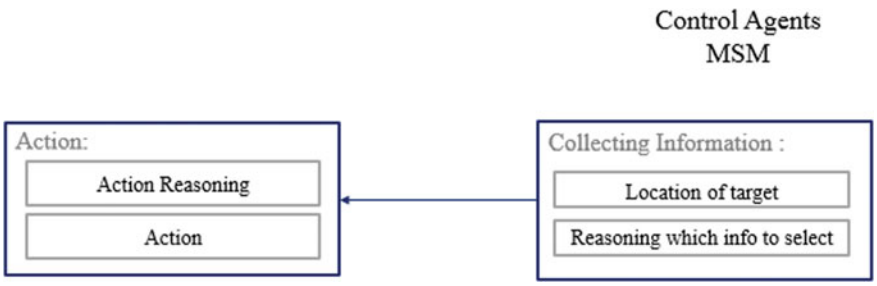


Fig. 2 Egocentric control agents' arrow and box diagram

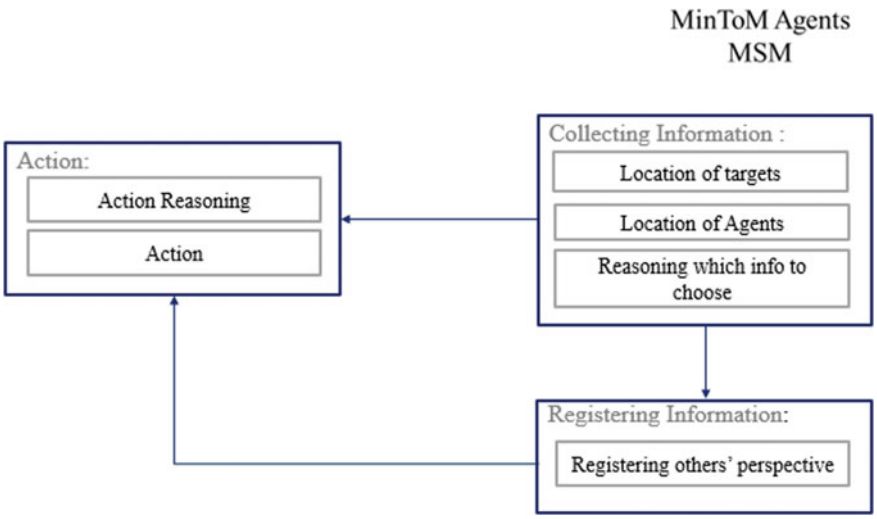


Fig. 3 MinToM agents' arrow and box diagram

The decisions of MinToM agents about achieving a target depend on three factors: their own belief about the location of the target, their own mental state and their belief about others' perspective regarding the location of the target. In order to track others' fields of view, MinToM agents require target and agent sensors, sensor memory, as well as a short-term memory to store the information in the current time step, as shown in Table 1. Short-term memory facilitates the information processing in each time step.

2.4 Simple Theory of Mind Agents (Infer Agents)

Agents with simple ToM ability, called Infer agents, are able to infer others' mental states. Infer agents collect information about the location of targets, the location of agents and agents' actions. The way that Infer agents infer others' mental state is simple; when they observe that an agent achieves a target, they infer that agent becomes Passive. Otherwise, they assume the agent is Active. Figure 4 shows an arrow and box diagram of Infer agents' decision-making processes which consist of the following four phases:

- ***Collecting Information Phase***: Infer agents collect information about the location of targets and other agents' perspectives from their field of view. More importantly, they collect information regarding other agents' actions in reaching a target. Collecting Information is a dynamic and online procedure, which is parallel with the changes of the environment over time and highly interconnected with other phases.
- ***Recording Information Phase***: Infer agents store other agents' desires and beliefs regarding the targets in their memory. They are able to store the information by exploiting three relevant types of memory, as shown in Table 1. The first type, sensor memory, relates to simple information from the neighborhood such as the location of a target in the current time step. The second type, short-term memory, is designed to store simple calculated information about different agents' perspectives in the current time step, for example, in registering and tracking others' fields of view. The third, long-term memory, stores the inferred information about

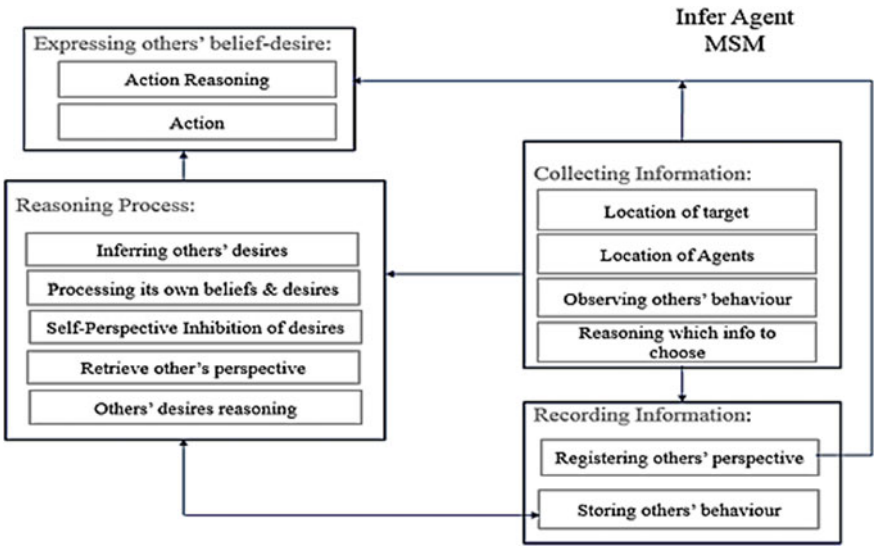


Fig. 4 Arrow and box diagram of infer agents' processes (IAP)

other agents' desires and beliefs for future access. The two main distinctions between these types of memory relate, firstly, to the length of time that information remains in the memory and secondly to the volume and complexity of their content, for example, information about others' perspectives. Memory demand is indispensable in order to understand others' mental states.

- ***Reasoning Process of Beliefs and Desires Phase***: This phase incorporates a central information processing unit for the purpose of belief-desire reasoning. Infer agents initially choose agents with the same beliefs as their own about the location of a shared target. Once the Infer agent perceives that another agent achieves a target, it concludes that agent's mental state was Active until it achieved the target. Thus, Infer agents infer that the agent's mental state changes to Passive in the current time step. This enables Infer agents to infer other agents' desires. Infer agents temporarily inhibit their own desire toward the target and retrieve the other agents' perspective from memory.

The reasoning phase for Infer agents involves the following five generic subroutines:

– Inferring others' desires from their observed behavior and storing the information in long-term memory;
– Processing its own beliefs and desires;
– Self-perspective inhibition of its own desires;
– Retrieving information about the subject agents' mental states from memory;
– Processing the subject agents' beliefs and desires.

By the end of this phase, Infer agents' reasoning about the other agents' mental states is completed.

- ***Expressing Others' Mental States (Actions as Output) Phase***: This phase is about deciding on actions by considering the mental states of others. There is a delicate, but important, distinction between understanding others' mental states and using (or expressing) this understanding in Infer agents' actions. In the first three phases of Infer Agents' Processes (IAP) shown in Fig. 4, agents understand others' mental states, and in the expressing phase they use and express this understanding.

In two identical situations in the environment, Infer agents act differently based on whether the other agent's mental state is Active or Passive. Infer agents use the information resulting from their ToM ability in this phase and express that use through their actions. A level of inevitable reasoning, action reasoning, in this phase is based on the principles of rational action in achieving goals. This action reasoning utilizes information about others' mental states and outputs a corresponding action.

To summarize, suppose F is the function that defines an agent's action and

$x_{1:}$ agent's own belief about the location of the target;

$x_{2:}$ agent's own desire;

$x_{3:}$ agent's belief about the other agent's perspective regarding the location of targets;

$x_{4:}$ agent's belief about the other agent's desire;

x_5: the action of the other agent.

For each type of agent, the action function, F, is defined as

$F_{Control}$ (x_1, x_2),

F_{MinToM} (x_1, x_2, x_3),

F_{Infer} (x_1, x_2, x_3, x_4, x_5).

Noticeably, the complexity of the action function inevitability grows as the level of ToM ability increases.

3 Results

Agents are competent in different levels of theory of mind ability including agents with simple ToM ability, agents with minimal ToM ability and agents with no ToM ability. The level of ToM ability is an independent variable and agents' performance is a dependent variable. The hypothesis is that agents' level of ToM has an impact on agents' performances in achieving their targets and the expectation is that agents' performances correlate with their level of theory of mind ability.

The total number of time steps in which Active agents fail to achieve a target is calculated as a measure of their performance. The agents' performances have been normalized to accomplish a standard scale. The results demonstrate a general pattern; Infer agents consistently perform more efficiently than the MinToM agents and similarly MinToM agents' performances are higher than those of Control agents. The results also demonstrate that the largest performance differences occur between MinToM agents and Infer agents.

The density of the number of targets, the number of agents and the ratio between them is critical to determine how agents use their strategies. Figure 5 demonstrates

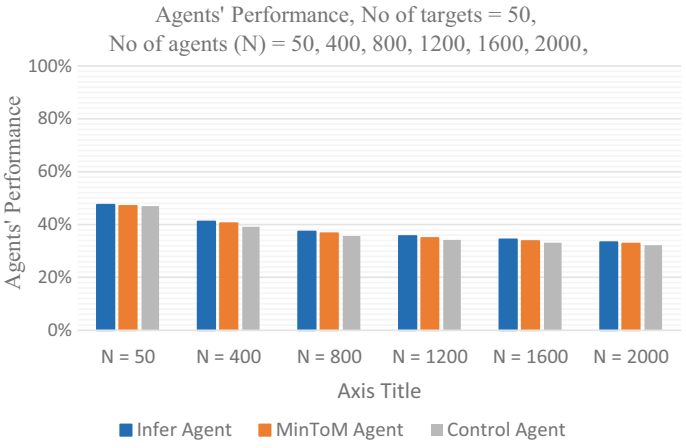


Fig. 5 The performance of agents in an extreme situation

the agents' performances in an extreme situation and shows that when the number of targets (T) = 50, and the number of agents = 50, the difference in agents' performances are not significant as we might expect. Both the density of targets and agents are very low. Thus, it is often less likely that the agents be able to use their ToM abilities. This extreme case is an example of the traditional use of sensitivity analysis that demonstrates the uncertainty caused by the parameters. Figure 6 shows the performance differences between agents rises due to the increased number of chances in which agents can apply their ToM ability when T = 400. As the figure shows, the performance differences between Infer agents and Min ToM agents vary from 3.9 to 4.4%.

The differences between Infer agents and Control agents fluctuate from 5.5 to 6.9% and between Min ToM agents and Control agents vary from 1.6 to 2.5%. Similarly, Fig. 7 for T = 800 demonstrates the same pattern of increasing performance differences between agents. The highest performance difference between Infer agents and Control agents is 6%, between Infer agents and Min ToM agents is 4.5% and between Min ToM agents and Control agents is 1.5%.

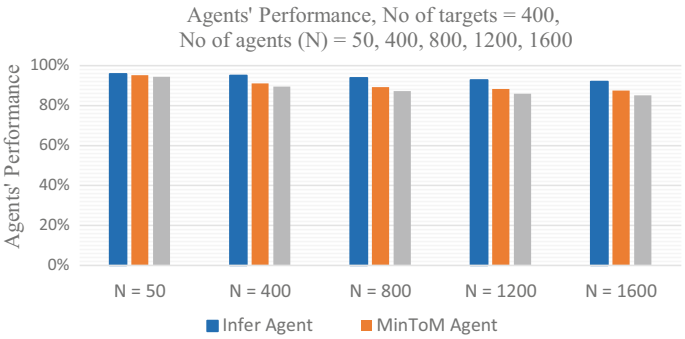


Fig. 6 The performance of agents, number of targets = 400

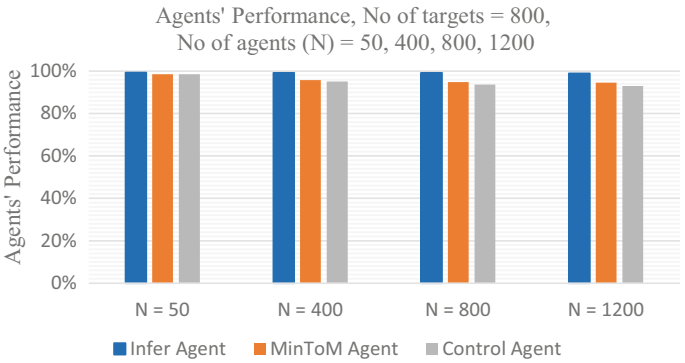


Fig. 7 The performance of agents, number of targets = 800

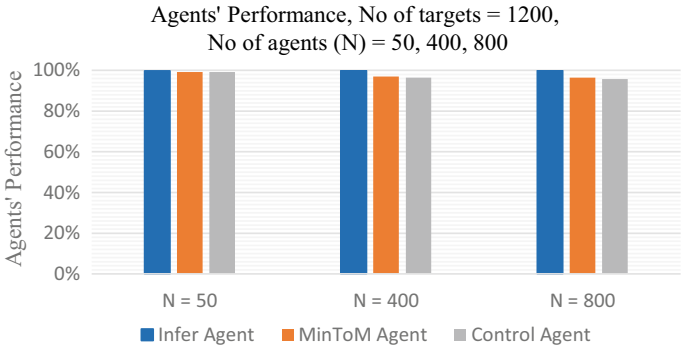


Fig. 8 The performance of agents, number of targets = 1200

The sensitivity analysis also shows that in extreme situations in which the ratio of targets to agents is high, agents achieve targets without exploiting their ToM abilities. For example, in Fig. 8, the agents' performances are less reflective of their ToM abilities due to the excessive number of targets, T = 1200.

3.1 The Cost and Resources of ToM

The ability of Infer agents to infer other agents' mental states demands more resources and costs. These resources based on the IAP diagram consist of a network of perception and attention (first phase), three types of memory for storing information (second phase), and inhibitory control and reasoning resources (third phase).

Moreover, in order to calculate the time consumed to infer others' mental states, the isolated processing time of the agents in nanoseconds is computed for each type of agent in the simulation. The results demonstrate that Infer agents have the longest processing time as shown in Fig. 9; one possible explanation for this result would be that there is additional reasoning time involved, additional information to process and also that the inhibition step consumes more time in retrieving information from memory. Hence, the infer system comes at the cost of more time and demands on interconnected resources including memory, perception, reasoning and inhibitory control.

4 Discussion

Experimental ToM research shows that preschool children fail in verbal false belief tasks, an acid test for ToM, whereas infants are able to pass non-verbal false belief tasks. How should we interpret this contradiction? Various valuable approaches

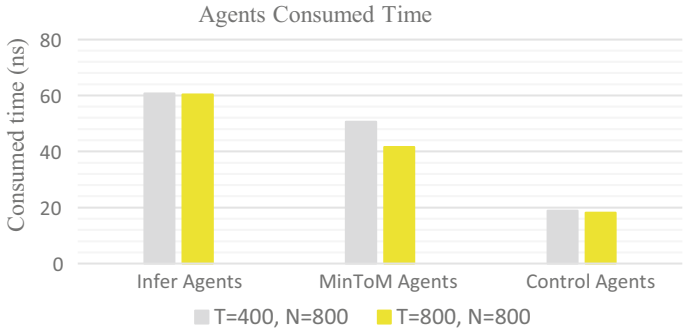


Fig. 9 The average consumed time by agents

explored their findings and suggested their own perspectives on these controversial issues. All of these different opinions agree on the experimental result that infants are able to pass non-verbal false belief tasks as young as 7 months, and in any case well below 4 years old, and also that preschool children fail verbal false belief tasks. Intriguingly, however, there is no consensus on whether infants truly understand non-verbal false beliefs. Although there is a lack of consensus, these plausible opinions would provide fresh insights in understanding belief representation. Broadly speaking, there are three different explanations for this problem.

The first explanation by Wellman suggests this is due to the developmental changes in preschool children. He argues that implicit false belief tasks require a simple attribution of desires and awareness rather than beliefs and false beliefs [31]. In general, he considers the theory of mind as a set of concepts to be acquired.

The second account argues that the belief representation ability of children under 4 years is obscured through the tasks' complexities and demands; for example, the child participants in false belief tasks need to override their own natural belief about the right location of the object and point out another's belief, which is not the right location. This may be a prepotent response that a child struggles to inhibit. According to the researchers of this line of argument (e.g. [6, 17, 18, 29]), both infants and preschool children are basically competent at belief-desire reasoning, the latter just fail to show this competence on tasks that exceed their capacity to inhibit inappropriate responses.

The third account, which was alluded to in the minimal theory of mind section, presents two processing systems: system one is an implicit system which is fast and limited, and enables infants to track others' belief-like states but not psychological representations as such, whereas system two is an explicit system which is flexible and cognitively demanding, and enables belief understanding [3, 13]. These limitations of system one will gradually be overcome by language and executive function development in childhood and equip children with system two. They suggest that infants may have an automatic emerging system, system one, which does not need much experience and persists into adulthood [3]. Recently there has been a growing

debate about the two systems in the ToM process. Some researchers argue that ToM is a single system [10] whereas others consider it a multi-system processing [11].

The focus of our study, however, is not to verify the above accounts, rather to explore the algorithmic mental process and representation mechanisms of two different levels of ToM ability from the perceptual level to output. It also investigates the structural relation between the main processes involved at each level of ToM ability and the BRSA. We now continue the discussion to explain our model and explore the processes of agents with simple ToM ability in more detail and provide brief evidence from the literature. Then we will present an investigation of the structural relation between the main processes involved at each level of ToM ability and the BRSA. This section will be completed by a short analysis of agents' performances.

Our model, the Mental State Model, consists of three types of agents with two mutually exclusive mental states regarding the target: Active or Passive. Control, MinToM and Infer agents move to achieve a target when they are in an Active state, whereas they cannot achieve any target while in a Passive state. Infer agents are able to infer other agents' mental states from observing their behavior in achieving a target. MinToM agents register other agents' perspectives regarding the location of targets. Control agents are able to only consider the location of targets and lack ToM ability.

The MSM presents the fundamental cognitive units of simple ToM processes through the IAP diagram. This set of basic processes addresses some of the ambiguity behind the ToM tasks by breaking a task into standard phases. It is also able to simplify and explain some of the complex characteristics within the tasks through the following four phases:

- ***Collecting Information Phase***. The first phase involves agents' perception (collecting information from the environment) which mainly corresponds to information about others' mental states. The IAP diagram shows that the perception phase is interconnected with all phases as a dynamic online access to the information. This phase utilizes reasoning to determine what information regarding the target is more important to select and in what priority. For example, the information about the nearest targets and the agents with a shared target has the highest priority. Infer agents collect information by observing other agents' actions as they are reaching a target, and this is analogous with the literature; one of the social information gathering skills in human infants in real life is decoding information regarding the social environment and discerning the information about an entity such as an object [5].
- ***Recording Information Phase***. The second phase involves memory and storing information regarding the mental state. The MSM elucidates memory demands on Infer agents to code and decode agents' desires and beliefs. There is extensive literature regarding the role of memory in ToM ability (e.g. [4, 12, 16]). The MSM also demonstrates that as the agents' ToM abilities develop, additional memory is required to store more complex information for a longer duration of time. Note that there is a subtle difference between registration and recording information,

which relates to the length of time and the type of memory that the information stores. In registration, the information is only accessible for the current time step, whereas in recording, the information is accessible for future time steps.

- ***Reasoning Process of Beliefs and Desires Phase***. The third phase, the reasoning process of beliefs and desires, encompasses five subroutines; Inferring others' desires from their observed behavior, Processing its own desires and beliefs, Self-perspective inhibition of its own mental states, Retrieving information about others' mental states and Subject agents' desires and beliefs processing. Infer agents start reasoning about their own beliefs and desires; they then temporarily inhibit their own perspective regarding the target and take protagonists' perspectives into account by retrieving the information from memory. The contribution of the executive function to ToM ability is explained by many researchers in the literature. For example, Leslie et al. argue that the executive function is necessary for ToM perhaps by inhibition of self-perspective [20]. Similarly, in this phase, the role of "memory" is identified in retrieving the information subroutine, the role of "inhibitory control" is presented in the self-perspective inhibition subroutine and the role of "reasoning" is revealed in the selective processing of beliefs and desires subroutine. Belief representation is accomplished by the end of the reasoning phase. Moreover, the IAP diagram is consistent with the developmental literature; it demonstrates a network of resources for Infer agents with a simple ToM competence, including perception, memory, self-perspective inhibition and reasoning resources.
- ***Expressing Beliefs and Desires of Others Phase***. The fourth phase relates to expressing the understanding of others' mental states as an action or behavior. Consistent with the literature, this phase refers to performing an action by using a mental representation [15]. In real life, the action of this phase might be as simple as an eye gaze and eye tracking or a more complicated action that requires additional reasoning. Based on the IAP diagram, there is a distinction between having ToM ability and using this ability in an action.

4.1 Simple ToM and the BRSA

The Infer agents' process in Fig. 10 demonstrates a simple ToM process within four basic cognitive phases, compatible with the BRSA. The underlying processes of a simple ToM, as we might expect, are equivalent to the BRSA phases.

The BRSA reveals the basic fundamental process for belief representation; the BRSA is flexible enough though for modification in more complex scenarios. As such, the more complex nature of the processes in a simple ToM process, rather than in a false belief task, demands the reasoning phase entails a new subroutine: the inferring others' mental states subroutine which is linked with the memory to store the information. Adding this subroutine to phase three of the BRSA and its link to the memory, the process of a simple ToM is otherwise fundamentally equivalent to

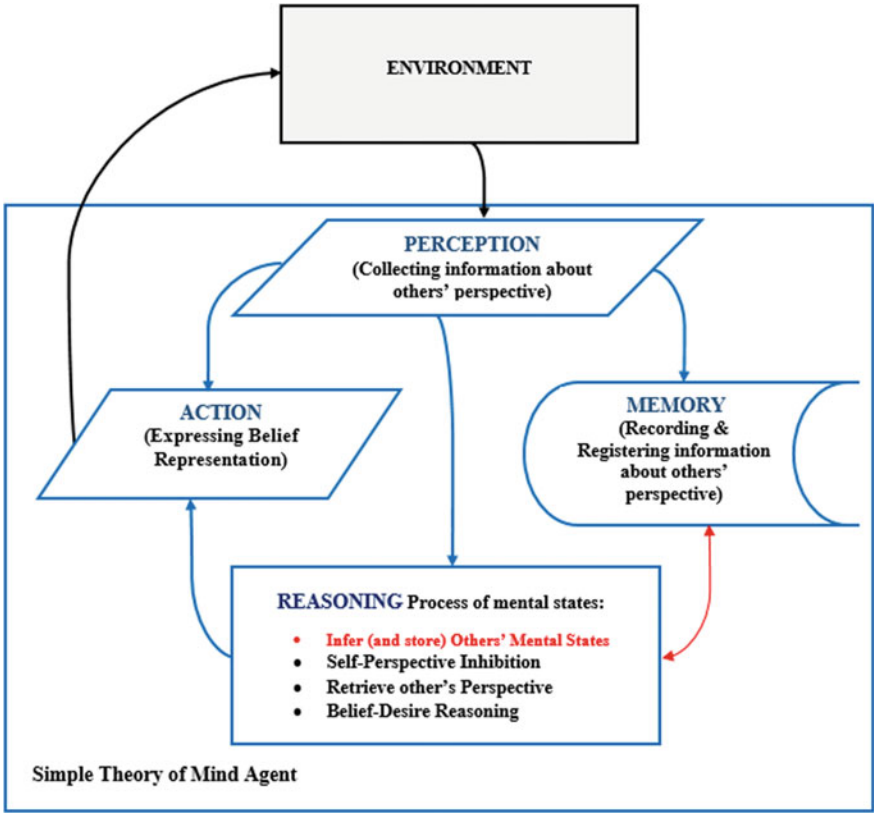


Fig. 10 The basic cognitive phases of a simple theory of mind

the BRSA. In this manner, BRSA is the basic foundation on which more complex ToM models can be built.

4.2 Minimal Theory of Mind and the BRSA

The first principle of the minimal theory of mind, goal-directed action, is a pre-assumption of our model. This includes the general rule that each agent has a goal, to achieve a target. The second principle, field and encountering, is parallel to the perception phase of the BRSA, in which agents encounter targets and other agents in their field of view. Registration and successful registration, the third principle of the minimal theory of mind, corresponds to the memory phase in the BRSA.

Action, the fourth principle in the minimal theory of mind, is equivalent to the expressing (action) phase in the BRSA. The minimal theory of mind phases are illustrated in Fig. 11. The figure, consistent with the literature, demonstrates that the

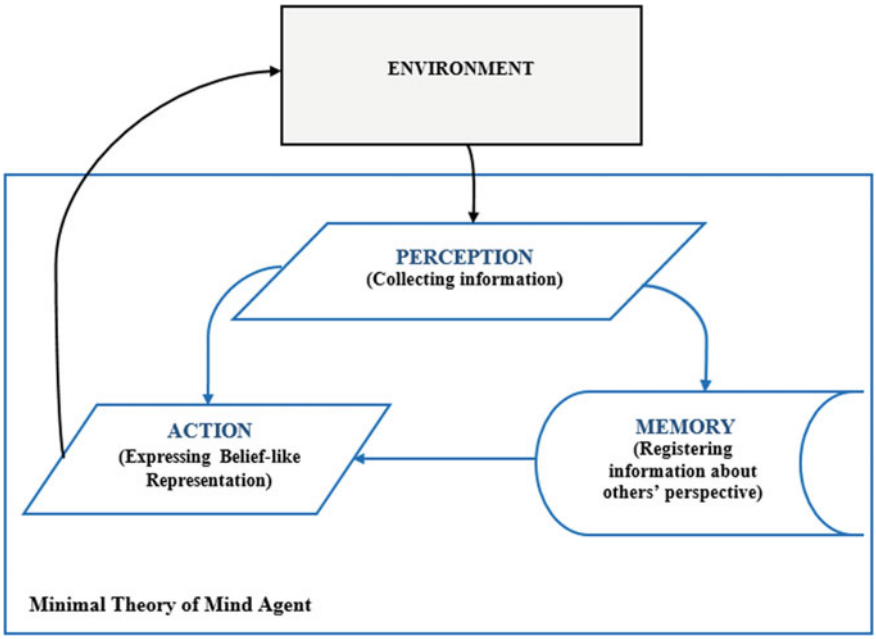


Fig. 11 Phases of a minimal theory of mind

third phase in the BRSA, the reasoning beliefs and desires phase, does not exist in the minimal theory of mind process. The minimal theory of mind process does not incorporate any reasoning or sophisticated concepts [9].

4.3 ToM Ability and Agents' Performances

The results of the simulation consistently suggest that there is a strong relationship between the agents' ToM capability and higher performance in a competitive society. As the ability of ToM develops from the simple level of understanding their own mental states to the level of tracking others' field of view and to the more complicated level of constructing inferences about other agents' mental states, the performance of agents effectively rises. However, the Infer agents' high performance comes at the cost of time because they use more time to collect information, reason, inhibit their own perspective and retrieve the information regarding others' mental states.

4.4 Presumptions of Mental States (Biases)

Infer agents have their presumptions about others' mental states. The presumptions are made when agents lack any reasons to think in a different way; these default assumptions (or biases) constitute the most suitable and productive way to make generalizations and often they are not correct [23]. Infer agents are capable of updating their presumptions regarding others' mental states as they observe other agents achieving targets. At the start of the simulation, Infer agents assume other agents' mental states are Active which is correct. As other agents' mental states change over time, the presumptions are not necessarily accurate anymore and will be updated by Infer agents' online inferences. MinToM agents' presumptions are that all agents are Active. Moreover, Control agents assume other agents are always Passive (ignoring other agents).

5 Conclusions

This study develops a novel cognitive framework for a simple ToM process. This framework validates the Belief Representation Systematic Approach (BRSA). The BRSA encompasses four phases consisting of collecting information as perception, recording information in memory, a reasoning process of beliefs and desires, and expressing the mental states of others in an action. The BRSA is a reliable systematic approach for simple ToM processes, and would contribute to analyzing ToM tasks within a network of cognitive structures. Since this cognitive structure enables construction and deconstruction of ToM processes, the BRSA can provide an analytical framework for distinguishing behavioral tasks from ToM tasks in future work. The BRSA provides standardization to ToM research that would help to reduce ambiguity and confusion in the ToM literature. The BRSA is a baseline capable of accommodating more complex ToM abilities by adjusting new features within its phases.

Appendix

The ODD Protocol Document for the Mental State Model

Overview

Purpose

The main purpose of the model is to study the fundamental process involved in a simple theory of mind (ToM) ability within a competitive and goal-directed environment. Specifically, it aims to explore the underlying processes of taking others' beliefs and desires into account in a goal-directed action. Moreover, the model aims to compare this process and the aggregated performance with the processes and performances involved in two other levels of ToM abilities including agents with the following ability:

(1) Minimal theory of mind: the ability to track others' fields of view
(2) Egocentric: Lack of understanding others' mental states.

State Variables and Scales

The Mental State Model (MSM) comprises interactions between three types of agents, representing different capabilities of understanding others' desires and beliefs: Infer agents (with Simple ToM ability), MinToM agents (with Minimal ToM ability) and Control agents (egocentric). The environment consists of a grid of 50 by 50.

The neighborhood (Ngh) of an agent is the area around it where the extent of its X-axis and Y-axis are equal. Field of movement consists of the vicinity cells around agents, the first neighborhood cells, in which agents are able to move. Field of view refers to the neighborhood in which an agent is able to perceive the environment within the fourth neighborhood (the square of 9 by 9 cells with the agent's location in the center).

The goals that agents are required to achieve are called targets and are illustrated by green cells in the environment. The number of targets in the environment remains constant through each time step of the simulation. Agents move to a cell which is not occupied by other agents. When an agent achieves a target, that target will be removed from the environment. However, in the next step a target will be generated in another random cell. The space is toroidal, therefore, if agents move to one border of the grid, they appear on the opposite border.

All agents are characterized by two mutually exclusive mental states: Active and Passive mental states. Each agent is required to achieve the goal. Thus, the target is used as a representative goal at the current time step and the underlying concept for agents' mental states. Tick is the time measurement unit; it is a step in the simulation in which agents simultaneously perform their actions. The default number of time steps is 1000 ticks.

Figure 12 shows the UML class diagram of the MSM including the classes, the relationships between them and their methods. The model comprises two abstract

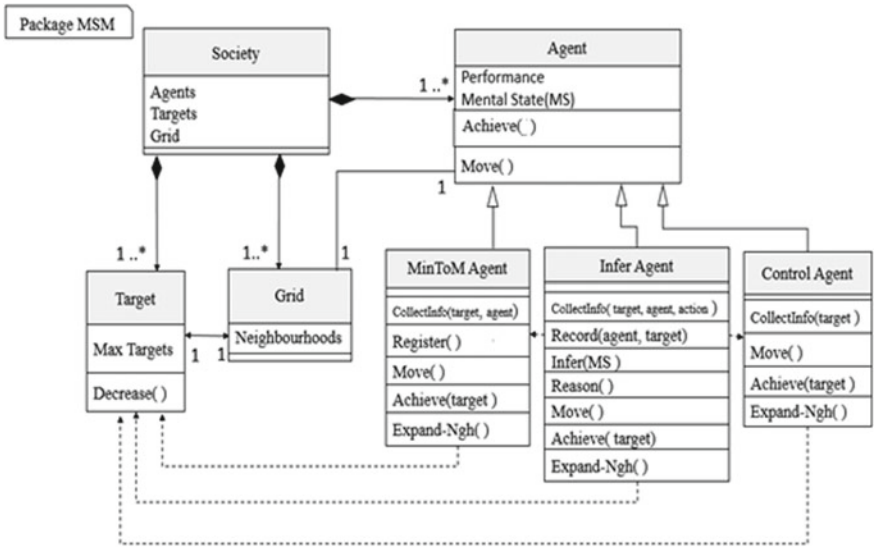


Fig. 12 Class diagram of the MSM model

classes: Society and Agent. Agents are moving to achieve their targets in a grid space where Society contains all of the classes including Agents, Targets and Grid.

The composition association between Society class and other classes, the black diamond sign in Fig. 12, shows that Society class is a container for objects of other classes. Agent class is an abstract class including three main subclasses: MinToM Agent, Infer Agent and Control Agent. The 1–1 association between Agent class and Grid class demonstrates that every agent is placed in one and only one cell of the grid. Likewise, the 1–1 association between Target class and Grid class indicates that each target can exist in one cell. The dependency relationship, shown by dots in Fig. 12, illustrates the flow of the information, about the target, from MinToM Agent to Grid class as they achieve a target. A similar dependency relationship occurs from the Infer Agent class to the Grid class and also from the Control Agent class to the Grid class.

The subclasses of the Agent class inherit its methods. In addition, they have their own specific methods. For example, Infer Agent class methods include CollectInfo, Record, Infer, Expand-Ngh and Reason whereas MinToM Agent class methods are the CollectInfo, Register and Expand-Ngh methods. Notice that although the name of some of the methods are the same and they are supposed to perform the same function, yet, they work differently with different parameters. The Decrease method of the Target class removes the target which has been achieved by an agent from the grid. Here in the model, the level of theory of mind ability is an independent variable and agents' performance is a dependent variable. The hypothesis is that agents' level of theory of mind ability has an impact on the agents' performances in achieving their targets in a competitive society.

Process Overview and Scheduling

In each time step, agents are randomly selected to collect the information from their neighborhood, and based on their strategies make a decision to move toward a target. Thus, the update of agents' states is synchronous. Figure 13 shows the search order for a target which starts from the first neighborhood to the wider area of the second neighborhood and continues to the third and fourth neighborhoods.

The shortest path toward the target is designed by using the intersection set between the agents' field of movement and the neighborhood of the target. Figure 14 shows an example of the intersection sets (red cells) between the agent (purple circle) and the target (green cross). The agent chooses one of the cells in the intersection set to move toward the target.

Active Control agents move to the nearest target in their neighborhood. If there is more than one target, they randomly choose one. Passive Control agents move toward an empty cell, which is in the vicinity of a target. Active MinToM agents move toward a target with the minimum number of agents around it with the desire to achieve the same target, whereas Passive MinToM agents move toward an empty cell in the vicinity of a target which is surrounded by the minimum number of agents with desire toward the same target.

For Infer agents, suppose ms is the variable for their desire, NoOfTargets is the number of targets in the first Ngh of Infer Agent and expandFieldOfView is the function that searches for targets in the first wider Ngh. Also, TargetWithMinActive is a function that calculates and returns which target has the minimum number of Active agents around it whereas TargetWithMaxActive is a function that calculates and returns which target has the maximum number of Active agents around it. The pseudo code of different cases are as follows:

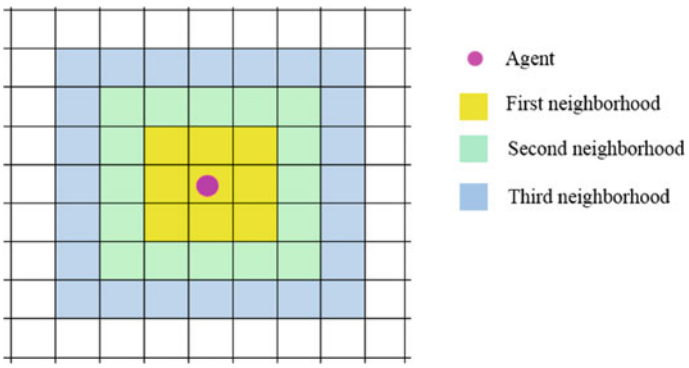


Fig. 13 The search order for a target

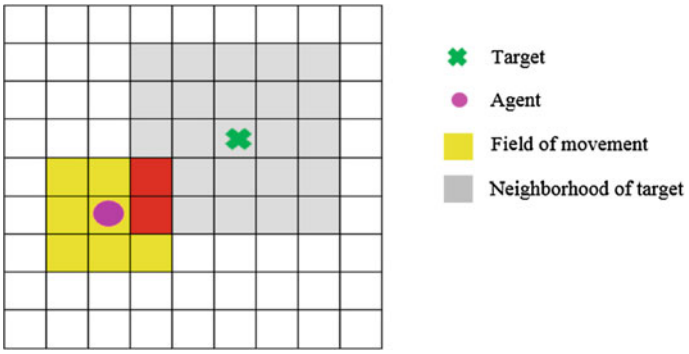


Fig. 14 The intersection set is shown by red cells

```
Case 1)  If (ms = Active & NoOfTargets = 1) Then
                                    Moveto (target)
                                    Achieve (target)
Case 2)  If (ms = Active & NoOfTargets = 0) Then
                                    expandFieldOfView()
                                    set output of TargetWithMinActive () to minTarget
                                    Moveto (A cell in Ngh of minTarget)
Case 3)  If (ms = Active & NoOfTargets > 1) Then
                                    TargetWithMaxActive()
                                    set output of TargetWithMaxActive () to maxTarget
                                    Moveto (maxTarget)
                                    Achieve (maxTarget)
Case 4)  If (ms = Passive & NoOfTargets = 1 Then
                                    Moveto (A cell in Ngh of target)
Case 5)  If (ms = Passive & NoOfTargets = 0) Then
                                    expandFieldOfView()
                                     set output of TargetWithMinActive () to minTarget
                                     Moveto (A cell in Ngh of minTarget)
Case 6)  If (ms = Passive & NoOfTargets > 1) Then
                                     set output of TargetWithMinActive () to minTarget
                                     Moveto (A cell in Ngh of minTarget)
```

Infer agents store the information regarding other agents' desires until the agent moves out of the Infer agent's field of view.

Design Concepts

The information about the type and number of agents, location of each agent and location of targets can be shown in the simulation run. At the end of the simulation, the number of time steps that Active agents did not achieve a target and the total process time consumed for each type of agent will be calculated and printed in the output file.

Table 2 Parameters

Parameter	Description
N	Number of agents: the agents' population in the environment
T	Number of targets in the environment
Ngh	Field of view (Neighborhood) of the agent
Tick	Total number of time steps in the simulation

Table 3 The values of parameters

Number of agents (N)	Number of targets (T)	Field of view (Ngh)	Tick
50	50	2	1000
400	400	3	
800	800	4	
1200	1200	6	
1600			
2000			

Tables 2 and 3 show the parameters and their values. This project was part of the primary author's PhD project and it was run with two more parameters, and 10 values for each parameter (N and T) resulting in a significant number of graphs and information. For the sake of simplicity, we are only using some of the values in this paper to demonstrate our points.

Furthermore, by altering the parameters' values in the initial conditions including number of targets, number of agents and the neighborhood of agents' field of view, the model sensitivity to its initial conditions is examined. Agents' performances are sensitive to the initial conditions, particularly where agents are unable to apply their rules and strategies in uncertain and extreme situations.

Details

Initialization

Simulation initializes with parameter values of N and T from Table 3 and the Field of view (Ngh) = 4 and Time Steps = 1000 Ticks. For the sake of simplicity, the simulation run for the other values of Ngh is not presented in this paper.

Input

The model does not include any input of external data.

Submodels

The basic functions include:

CollectInfo(): Agents collect the information from their field of view about the location of targets, agents and their actions depending on the type of agent.

Moveto(): Agent moves to a new location.

Achieve(): Agent achieves a target.

Expand-Ngh(): Agent expands its field of view (Ngh).

Register(): Agent stores the other agents' desires only for the current time step.

Infer(): Agent observes other agents' actions and keeps the information if they achieve a target.

Record(): Agent stores the information about other agents' desires.

DecisionMaking(): Agents choose a cell in their first neighborhood to move based on all of the information they collected and stored.

Each of these functions is tested by printing and checking the results in different scenarios.

References

1. Apperly, I.: Mindreaders: The Cognitive Basis of "Theory of Mind." Psychology Press, Hove (2011)
2. Apperly, I.: What is "theory of mind"? Concepts, cognitive processes and individual differences. Q. J. Exp. Psychol. **65**, 825–839 (2012)
3. Apperly, I.A., Butterfill, S.A.: Do humans have two systems to track beliefs and belief-like states? Psychol. Rev. **116**(4), 953–970 (2009)
4. Arslan, B., Wierda, S., Taatgen, N., Verbrugge, R.: The role of simple and complex working memory strategies in the development of first-order false belief reasoning: a computational model of transfer of skills. In: Proceedings of the 13th International Conference on Cognitive Modeling, pp. 100–105 (2015)
5. Baldwin, D.A., Moses, L.J.: The ontogeny of social information gathering. Child Dev. **67**(5), 1915–1939 (1996)
6. Baillargeon, R., Scott, R.M., He, Z.: False-belief understanding in infants. Trends Cogn. Sci. **14**, 110–118 (2010)
7. Baker, C.L., Saxe, R., Tenenbaum, J.B.: Action understanding as inverse planning. Cognition (2009)
8. Bloom, P., German, T.: Two reasons to abandon the false belief task as a test of theory of mind. Cognition **77**, B25-31 (2000)
9. Butterfill, S., Apperly, I.: How to construct a minimal theory of mind? Mind Lang. **28**(2), 606–637 (2013)
10. Carruthers, P.: Two systems for mindreading? Rev. Philos. Psychol. **27**(1), 141–162 (2016)

11. Christensen, W., Michael, J.: From two systems to a multi-systems architecture for mindreading. New Ideas Psychol. **40**, 48–64 (2016)
12. Fliss, R., Le Gall, D., Etcharry-Bouyx, F., Chauviré, V., Desgranges, B., Allain, P.: Theory of mind and social reserve: alternative hypothesis of progressive theory of mind decay during different stages of Alzheimer's disease. Soc. Neurosci. 1–15
13. Frith, C.D., Frith, U.: Implicit and explicit processes in social cognition. Neuron **60**, 503–510 (2008). https://doi.org/10.1016/j.neuron.2008.10.032
14. Hughes, C.: Executive function in preschoolers: links with theory of mind and verbal ability. Br. J. Dev. Psychol. **16**, 233–253 (1998)
15. Hughes, C.: Social Understanding and Social Lives: From Toddlerhood through to the Transition to School. Hove: Psychology Press (2011a)
16. Hughes, C.: Changes and challenges in 20 years of research into the development of executive functions. Infant Child Dev. **20**(3), 251–271 (2011b)
17. Kovács, Á.M., Téglás, E., Endress, A.D.: The social sense: susceptibility to others' beliefs in human infants and adults. Science **330**(6012), 1830–1834 (2010)
18. Leslie, A.: Developmental parallels in understanding minds and bodies. Trends Cogn. Sci. **9**(10), 459–462 (2005)
19. Leslie, A.M., Polizzi, P.: Inhibitory processing in the false belief task: two conjectures. Dev. Sci. **2**(1), 247–253 (1998)
20. Leslie, A.M., Friedman, O., German, T.P.: Core mechanisms in 'theory of mind'. TRENDS Cogn. Sci. **8**(12) (2004)
21. MacLeod, C.M.: In: Gorfein, D.S. (ed.) Inhibition In Cognition, vol. xvii. American Psychological Association, Washington, DC, USA (2007)
22. Marcovitch, S., O'Brien, M., Calkins, S., Leerkes, E., Weaver, J., Levine, D.: A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. Cogn. Dev. (2015)
23. Minsky, M.: The Society of Mind. Simon & Schuster, New York (1988)
24. Premack, D., Woodruff, G.: Does the chimpanzee have a theory of mind? Behav. Brain Sci. **1**, 515–526 (1978)
25. Repast.: The Repast Suite (2019). https://repast.github.io/. Accessed 24 July 2019
26. Russell, J.: Agency: Its Role in Mental Development. Erlbaum (UK), Taylor & Francis, Hove (1996)
27. Schaafsma, S., Pfaff, D., Spunt, R., Adolphs, R.: Deconstructing and reconstructing theory of mind. Trends Cogn. Sci. **19**(2), 65–72 (2015)
28. Sharrock, W., Coulter, J.: Theory of mind: a critical commentary continued. In: Leudar, I., Costall, A. (eds.) *Against Theory of Mind.* Palgrave Macmillan, Basingstoke (2009)
29. Southgate, V., Vernetti, A.: Belief-based action prediction in preverbal infants. Cognition **130**, 1–10 (2014)
30. Sun, R.: Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation. Cambridge University Press, New York (2006)
31. Wellman, H.: Making Minds. How Theory of Mind Develops. Oxford University Press, NewYork (2014)
32. Yousefi, Z., Heinke, D., Apperly, I., Siebers, P.O.: An agent-based model for false belief tasks: belief representation systematic approach (BRSA). In: Thomson, R., Dancy, C., Hyder, A., Bisgin, H. (eds.) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018. Lecture Notes in Computer Science, vol. 10899. Springer, Cham, pp. 111–126 (2018)

Decision-Adjusted Modeling for Imbalanced Classification: Predicting Rooftop Solar Panel Adoption in Rural Virginia

Zhihao Hu, Xinwei Deng, Achla Marathe, Samarth Swarup, and Anil Vullikanti

Abstract Understanding the factors that contribute to the adoption of rooftop solar panels can help design appropriate policies to convert potential adopters into actual adopters. This paper aims to predict the adoption of rooftop solar panels in the rural regions of Virginia using demographic features of households. A novel feature of this work is that it uses a decision-adjusted framework, which optimizes the specific goals of the study. Another novel feature is that we construct indicator features to improve the model performance. Traditional logistic regression models and decision-adjusted logistic regression models are compared. Results show that decision-adjusted models are better than traditional logistic regression models, and that indicator features can improve the performance of predictive models. The sum of true positive and true negative rates of the decision-adjusted model is 1.5973, while for the traditional logistic regression model it is only 1.

Z. Hu · X. Deng
Department of Statistics, Virginia Tech, Virginia Beach, VA, USA
e-mail: huzhihao@vt.edu

X. Deng
e-mail: xdeng@vt.edu

A. Marathe (✉)
Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
e-mail: achla@virginia.edu

S. Swarup
Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA
e-mail: swarup@virginia.edu

A. Vullikanti
Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia Department of Computer Science, University of Virginia, Charlottesville, VA, USA
e-mail: vsakumar@virginia.edu

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_22

1 Introduction

Fossil fuels are the primary source of energy in the world and are considered non-renewable since they take millions of years to form. They are also the primary source of human-caused emissions. According to the US Department of Energy, fossil fuels provided 80% of the total energy needed in the US and caused nearly 75% of the emissions in the last two decades. With the demand for energy ever-increasing and a need for decarbonization, the use of renewable energy must increase.

Among sources of renewable energy, solar is the most abundant. Due to a faster than expected decline in the cost of solar photovoltaics (PV), combined with economic incentives such as state and federal subsidies, solar is becoming an important mainstream electricity source. The economics of solar power generation has been steadily improving as the cost of panels and installation has gone down, and panel efficiency has gone up. The cost of solar PV panels has dropped more than 60% from 2010 to 2016 [1] and solar panel capacity has improved 22 times from 2008 to 2015 [2].

Solar energy has many advantages. It lowers greenhouse gas emissions, creates green jobs, and provides higher grid resiliency. It can be built as distributed generation or as a central station; consumption occurs near the point of generation which cuts down transmission and distribution losses. The use of solar is especially efficient in remote and rural regions where space is abundant and expansion of transmission capacity and grid maintenance enhancements are expensive due to low aggregate demand.

Though the use of solar energy has evident advantages, the adoption of solar energy is low in general, especially in remote regions. State and federal subsidies have helped but there is a lot more potential for increasing the use of solar power. If the likelihood of households adopting solar panels can be properly characterized, appropriate policies can be designed to convert potential adopters into actual adopters.

This research aims to understand the factors that contribute to the adoption of rooftop solar panels in the rural regions of Virginia in order to predict potential solar adopters. Due to low penetration of solar adopters, the data on adopters in remote and rural regions is sparse. The small number of positive samples (i.e., instances with the adoption of solar panels) make it difficult to predict true solar adopters using traditional statistical methods. The overall prediction accuracy of traditional methods is good, but they often predict almost all households to be non-adopters due to the class imbalance.

To overcome this drawback, our proposed method considers a novel "decision-adjusted" framework developed by Mao et al. [3], to enable the estimated model tailored for the decision of interest. Note that traditional evaluation methods use cross-validation and area under the curve (AUC) as the criteria for variable selection and parameter tuning. Such approaches cannot address the issue of class imbalance. Under the decision-adjusted framework, model estimation and parameter-tuning are conducted to optimize a specific decision-based model evaluation criterion. The

estimated data analytic model is then optimized for this specific objective instead of cross-validation error, likelihood, or AUC.

The remainder of this work is organized as follows. Section 2 reviews related studies. Section 3 describes the survey data, and details the decision-adjusted framework. Section 4 gives the performance results of different decision-adjusted and non-decision-adjusted models. Section 5 provides a discussion.

2 Related Work

Graziano and Gillingham [4] developed a diffusion model for the adoption of residential solar photovoltaic (PV) systems in Connecticut (CT). They found the primary determinants of diffusion of PV systems in CT are spatial neighbor effects and the built environment. Aklin et al. studied solar technology adoption in rural India. Their results showed that household expenditures and savings as well as the head of the households' entrepreneurial attitude are strong predictors of adoption [5]. Parkins et al. studied factors that affected solar adoption intention in Canada. Their results showed that the visibility of solar technology has a particularly strong effect on intention, and the perceived knowledge of energy systems and being publicly engaged in energy issues significantly increase adoption intention [6].

Alyousef et al. [7] modeled solar PV and battery adoption in Germany using an agent-based approach. They found that increasing electricity prices could result in improved PV-battery adoption in Germany better than reducing PV-battery prices could. Zhang et al. applied an agent-based modeling framework to forecast individual and aggregate residential rooftop solar adoption in San Diego county [8]. However, they focused on the prediction of solar adoption rate, instead of prediction accuracy.

Vasseur and Kemp [9] studied the motivations for adoption for the subgroup of adopters and compared the relevance of different factors for non-adoption for the subgroups of potential adopters and rejecters. They found that for adopters the costs of adoption are considered affordable, whereas for non-adopters they are viewed as being too high. Rai et al. [10] stated that system cost is not the only barrier faced by solar adopters. They found that installers and neighbors play important roles throughout the decision-making process and have an influence on both the decision to adopt as well as on the mode of adoption.

Recently, Gupta et al. studied the prediction of rooftop solar adoption in rural Virginia [11]. Their results showed that the demographics and neighborhood-level features influence the likelihood of adoption but social network-based features do not. They also reported the in-sample and out-of-sample classification errors. Not surprisingly, the true negative rate is very high while the true positive rate is zero because the proportion of adopters is very small in the dataset. Thus, they proposed a weighted model to solve the problem. A new dataset was created where the ratio of adopters and non-adopter households was 1:1. The number of adopters was increased by simply resampling from the adopter data with replacement until the ratio

of adopters to non-adopters became 1:1. The weighted model-based out-of-sample results improved the true positive rate.

This research aims to improve the prediction performance of the model by using a novel framework that was developed by Mao et al. [3]. They proposed a decision-adjusted approach to develop the optimal driver risk prediction model using telematics driving information. Their "decision-adjusted" prediction framework optimizes the prediction model with respect to a pre-specified percentage of the riskiest drivers as "high-risk drivers." The corresponding decision-adjusted objective function is to maximize the prediction precision, which is the percentage of correct identification among the drivers labeled as high risk. In our study, we aim to maximize the true positive rate plus true negative rate (TPR + TNR), i.e., we want to identify as many adopters as possible while maintaining a high level of overall accuracy.

3 Methodology

3.1 Data

The data used in this study were collected in February 2018 through a survey of members of four rural electric cooperatives in Virginia. A total of 1470 households participated in the survey, of which 604 responded to phone interviews and 866 participated online. All respondent households were residential members of one of the following Virginia electric cooperatives: A&N Electric Cooperative, Central Virginia Electric Cooperative, Rappahannock Electric Cooperative, and Southside Electric Cooperative.

The survey collected information on demographics of the households, perceptions about solar energy, potential influencers, neighborhood features, etc. The prediction model used variables such as household income, household size, built year, market value of the house, square footage of the house, pool present or not, number of bedrooms, and number of neighbors within a mile who have installed solar panels. Households that had missing data or incomplete information were removed from the analysis. This resulted in 675 households in total. Of these only 14 were solar adopters. Table 1 provides the list of the variables used.

3.2 Decision-Adjusted Model

When we apply a traditional statistical model to a dataset that has very few solar adopters in the training data, the model will predict most of the households to be non-adopters. If the response variable is binary, logistic regression will traditionally be used to build the model. The estimation of logistic model tries to maximize the likelihood of probability distribution instead of optimizing the model under our

Table 1 Explanations of variables used

	Variable	Description	Type
Y	PV adopter	Has rooftop solar panels	Binary
X_1	Income	Estimated household income	Continuous
X_2	Built	Year home was built	Integer
X_3	Sqfoot	Square footage of the house	Continuous
X_4	Pool	Swimming pool present at home	Binary
X_5	Bdroomcnt	Number of bedrooms in the house	Integer
Z_1	Neighbors	Number of households in the neighborhood (approximately within a mile of the house) who have installed solar panels	Integer
Z_2	Hhnum	Number of people in the family	Integer
Z_3	Mktval	Estimated market value of home	Continuous

specific objective, i.e., identify true adopters. Even though the overall prediction accuracy of logistic regression is very high, it fails to characterize the potential solar adopters, which is the primary goal of this work. To address this issue, we use the decision-adjusted framework which allows decision-driven optimization [3].

Our preliminary study using the logistic regression model finds that some of the features have constant coefficients, while others do not. For example, if the coefficient of a feature is positive, then a larger value of the feature will increase the likelihood of adoption. However, this positive relationship is often not constant. The relationship may be strong when the value of the feature is small, and weak when the value of the feature is large. This will be further verified in Table 3 in Sect. 4.4. The dataset is $\{X_i, Z_i, Y_i\}_{i=1}^n$, where X represents the demographic features of the households with constant coefficients, Z represents the demographic features of the households with non-constant coefficients, and Y is whether the household is a solar adopter ($Y = 1$) or not ($Y = 0$). The X and Z variables are described in Table 1. To address the issue of non-constant coefficients, we introduce three indicator variables $I_j(\delta_j)$, $j = 1, 2, 3$, to the model. An indicator variable is defined as

$$I_j(\delta_j) = \begin{cases} 1, \text{ if } Z_j > \delta_j; \\ 0, \text{ if } Z_j \leq \delta_j, \end{cases}$$

where δ_j, $j = 1, 2, 3$, are threshold values to construct the indicator variables. A fixed coefficient for Z_j is not able to catch all information from the data, while the indicator variable can help bring more information to the model. Let us consider 'neighbors' as an example; adoption by neighbors does have an influence on the household but the influence may not increase as the adoption among neighbors increases. Therefore, a fixed coefficient for the neighbors feature is not able to describe this influence comprehensively. The indicator feature allows having a large coefficient when the value of neighbors feature is small, and a small coefficient when the feature value is large. Thus, the indicator variable is able to describe this influence better and bring more flexibility into the model.

The decision-adjusted solar adopter prediction model is as follows. Suppose that a prediction model is obtained from data $\{X_i, Z_i, I_i(\delta), Y_i\}_{i=1}^n$. Logistic regression models the response through the linear combination of coefficients and variables,

$$\Pr(Y = 1|X, Z, I(\delta)) = \frac{exp\Big(\big(X, Z, I(\delta)\big)^T \boldsymbol{\beta}\Big)}{1 + exp\Big(\big(X, Z, I(\delta)\big)^T \boldsymbol{\beta}\Big)},$$

where $\boldsymbol{\beta}$ represents the regression coefficients, and is estimated through maximizing the log-likelihood $l(\boldsymbol{\beta})$ or a penalized log-likelihood,

$$l(\boldsymbol{\beta}) - P(\boldsymbol{\beta}, \alpha, \lambda), \tag{1}$$

where

$$P(\boldsymbol{\beta}, \alpha, \lambda) = \lambda\big((1-\alpha)\frac{1}{2}||\boldsymbol{\beta}||_2^2 + \alpha||\boldsymbol{\beta}||_1\big),\ \lambda \geq 0,\ 0 \leq \alpha \leq 1$$

is the elastic net (enet) penalty [12]. The enet penalty can control the model complexity. The lasso part, $||\boldsymbol{\beta}||_1$, performs the variable selection. If there are two important variables with high correlation, the lasso penalty will keep only one of the variables to be nonzero. The ridge part, $||\boldsymbol{\beta}||_2^2$, will keep both important variables in the model. The enet penalty is a linear combination of the lasso penalty and the ridge penalty. The model predicts the probability of a household being a solar adopter, $\Pr(Y_i = 1)$. The predicted label for this household is determined by the following, where τ is the threshold probability for determining whether a household is adopter or not.

$$\hat{Y}_i = \hat{Y}_i(X_i, Z_i, I_i(\delta), \alpha, \lambda, \tau) = \begin{cases} 1, \text{ if } \Pr(Y_i = 1) \geq \tau \\ 0, \text{ if } \Pr(Y_i = 1) < \tau \end{cases}$$

Note that $\boldsymbol{\beta}$ is the model parameter, which is estimated by optimizing the penalized likelihood in Eq. (1), given certain α, λ, and δ. The α, λ, δ, and τ are hyperparameters to be tuned. Traditionally, the hyperparameters are tuned to optimize the cross-validation error or AUC. Under the decision-adjusted framework, the hyperparameters are tuned to optimize the specific objective function. Then the optimization

problem for our decision-adjusted model is

$$\max_{\delta,\alpha,\lambda,\tau} \Omega\left(\hat{Y}_i(X_i, Z_i, I_i(\delta), \alpha, \lambda, \tau), Y\right), \tag{2}$$

where $\Omega(\cdot)$ is a function of the hyperparameters. It is determined by the specific goal of the study.

Usually, the threshold τ is chosen as 0.5 for traditional logistic regression to classify 0s and 1s; this threshold works well when the dataset is fairly balanced. However, in our data, the proportion of solar adopters is only about 0.02. If we set $\tau = 0.5$, the prediction model fails to identify true solar adopters as shown in Table 2, which provides both the in-sample and out of sample performance. M_0 is the logistic regression model with $\tau = 0.5$. Although the model has a very high prediction accuracy because it correctly identifies most non-adopters, but not the true adopters.

4 Case Study and Results

We divide the data into adopters and non-adopter classes. From each class, we randomly select half of the data to include in the training set and the other half to make a test set. This ensures that each set includes 7 of the 14 adopters. The training set has a total of 337 households, and the test set has 338 households. To obtain in-sample results, the model is trained and tested on the whole dataset. To obtain prediction (out of sample) results, the model is trained on the training set and tested on the test set.

The decision-adjusted model is compared against the baseline logistic model and several other models. In the decision-adjusted model, we want to maximize ($\Omega =$ TPR + TNR) given τ and δ. Here 'TP' refers to true positives, 'TN' refers to true negatives, 'FP' refers to false positives, and 'FN' refers to false negatives. The TPR + TNR is defined as follows:

$$\Omega = \text{TPR} + \text{TNR} = \frac{TP}{TP + FN} + \frac{TN}{TN + FP}.$$

One can see that this objective is better than the accuracy (ACC), which is defined as

$$\text{ACC} = \frac{TP + TN}{TP + FN + TN + FP} = \frac{TP + TN}{n},$$

where TP + FN + TN + FP is the total number of data and is fixed. The ACC puts the same weight on TP and TN. In our imbalanced data, the number of positives (TP + FN) is much less than the number of negatives (TN + FP), and the TPR from the traditional logistic regression is close to 0. Compared to ACC, the new objective

(TPR + TNR) puts more weight on TP than on TN since $\frac{1}{TP+FN} >> \frac{1}{TN+FP}$ and enables us to achieve higher TPR while keeping TNR relatively high.

4.1 Model Comparison

We consider nine different models; these are:
M_0: Logistic regression model.
M_1: Logistic regression model with enet (elastic net).
M_2: Weighted model used by Gupta et al. [11].
M_3: Decision-adjusted Logistic regression model.
M_4: Decision-adjusted Logistic regression model with enet.
M_5: Decision-adjusted Logistic regression model, with neighbors indicator feature.
M_6: Decision-adjusted logistic regression model, with 'hhnum' indicator feature.
M_7: Decision-adjusted logistic regression model, with 'mktval' indicator feature.
M_8: Decision-adjusted logistic regression model, with all indicator features.
M_9: Decision-adjusted logistic regression model with enet, and all indicator features.

M_0 is the logistic regression model, and M_1 is the logistic regression with enet penalty. M_2 is the weighted model as used by Gupta et al. [11], which increases the number of adopters in the training data by sampling them with replacement. M_3 and M_4 are decision-adjusted versions of M_0 and M_1. We would like to compare M_0 with M_1, and M_3 with M_4 to see if enet improves the models' performance. The difference between M_3 and M_5, M_6, M_7 is that M_5, M_6, M_7 include an indicator feature. The difference between M_3 and M_8 is that, M_8 includes all three indicator features.

First, we compare the decision-adjusted models and non-decision-adjusted models, M_3 to M_9 vs. M_0 and M_1. Second, we compare the decision-adjusted models and weighted model, M_3 to M_9 vs. M_2. Third, we compare the models with enet and models without enet, M_1 vs. M_0, M_4 vs. M_3, and M_9 vs. M_8. Last, we compare the models with indicators and models without indicators, M_5 to M_8 vs. M_3.

4.2 True Postitive Rate (TPR) + True Negative Rate (TNR)

In this subsection, we compare different models in terms of TPR + TNR. For M_0, logistic regression model, and M_2, weighted model, we don't have any parameters to tune. For M_1, logistic regression model with enet, we tune two parameters, α and λ, from enet penalty to maximize the penalized likelihood. For decision-adjusted models, we tune the threshold, τ, δ's from indicator variables, and, α and λ from enet penalty to optimize TPR + TNR. The results are shown in the Table 2.

In terms of out-of-sample TPR + TNR, the decision-adjusted models, M_3 to M_9, are much better than non-decision-adjusted models, M_0 and M_1. The decision-adjusted models with indicator features, M_5, M_8, and M_9, are better than weighted

Table 2 In-sample and prediction (out-of-sample) TPR + TNR of different models

	In-sample			Prediction		
Model	TPR + TNR	TPR	TNR	TPR + TNR	TPR	TNR
M_0	1.0714	0.0714	1	1	0	1
M_1	1.0714	0.0714	1	1	0	1
M_2	1.6801	0.8571	0.8230	1.4324	0.5714	0.8610
M_3	1.5917	0.7143	0.8775	1.4264	0.5714	0.8550
M_4	1.5933	0.7143	0.8790	1.4234	0.5714	0.8520
M_5	1.6780	0.7143	0.9637	1.5451	0.7143	0.8308
M_6	1.6423	0.8571	0.7872	1.4294	0.5714	0.8580
M_7	1.6741	0.8571	0.8169	1.4324	0.5714	**0.8610**
M_8	1.7985	0.9286	0.8699	1.5602	0.7143	0.8459
M_9	1.7970	0.9286	0.8684	**1.5973**	**0.8571**	0.7402

model, M_2. By comparing M_1 vs. M_0, M_4 vs. M_3, and M_9 vs. M_8, the models without enet are close to the models with enet. The decision-adjusted models with indicator feature for neighbors are better than decision-adjusted models without indicator feature, while the decision-adjusted models with indicator feature for hhnum and mktval are not better than decision-adjusted models without indicator feature.

Figure 1 shows the comparison between different models. The model M_3 has higher TPR + TNR than M_0, and M_4 is better than M_1, which indicates that the decision-adjusted models are better than the traditional logistic regression models. The model M_0 is close to M_1, $M3$ is close to M_5, and M_8 is close to M_9, which means the models with enet are not better than the models without enet. The model M_5 is better than M_3, which indicates that the model with indicator feature is better than the model without indicator feature.

4.3 Comparison Among Different Decision-Adjusted Models

The performance of different decision-adjusted models is compared given a certain target percentage of solar adopters. For example, if we want to predict at most 10% households as solar adopters, 68 out of 675 households will be selected as solar adopters, some of them will be true positives and the rest will be false positives. Then we compare TPR + TNR among different models and we have corresponding curves in the figures below. Figure 2 shows the in-sample results for decision-adjusted models with and without indicator features, Fig. 3 shows the in-sample results for decision-adjusted models with and without enet, Fig. 4 shows the prediction results for decision-adjusted models with and without indicator features, and Fig. 5 shows the prediction results for decision-adjusted models with and without enet. Both in-sample and out-of-sample results show the decision-adjusted logistic regression models are

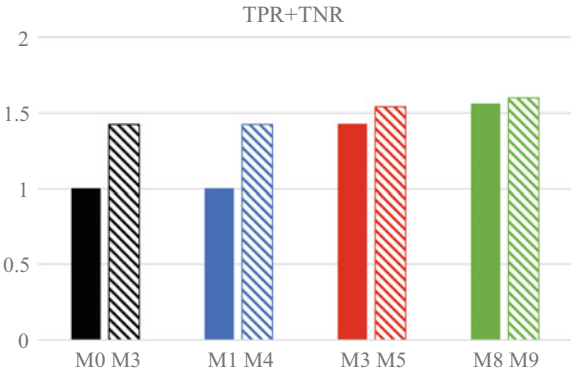


Fig. 1 Prediction performance comparison of different models. The model M_0 is the logistic regression model, and M_3 is the decision-adjusted logistic regression model. The model M_1 is the logistic regression model with enet, and M_4 is the decision-adjusted logistic regression model with enet. The model M_3 is the decision-adjusted logistic regression model, and M_5 is the decision-adjusted logistic regression model with an indicator feature for the neighbors. The model M_8 is the decision-adjusted logistic regression model with all indicator features, and M_9 is the decision-adjusted logistic regression model with enet and all indicator features

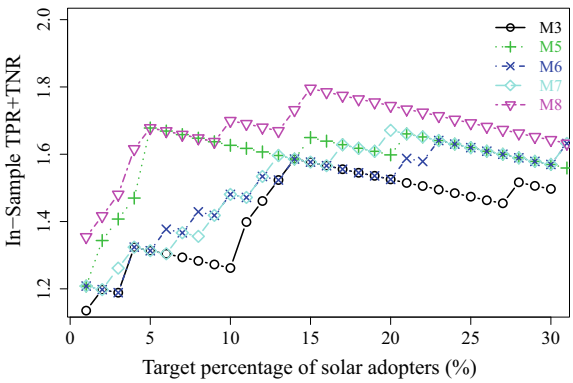


Fig. 2 **In-sample** performance comparison of decision-adjusted models **with indicator features** (in color other black, M_5, M_6, M_7, and M_8) and **without indicator features** (in black, M_3)

better than the traditional logistic regression models. The in-sample results show the decision-adjusted models without enet and decision-adjusted models with enet are close to each other, while the out-of-sample results show the decision-adjusted model with enet, M_9, is better than the decision-adjusted model without enet, M_8.

4.4 Model Coefficients

To study the influence of indicator features on the likelihood of solar adoption, the model with indicator features and the model without indicator feature are compared.

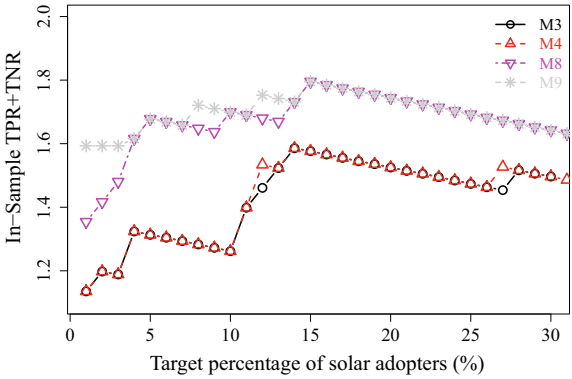


Fig. 3 In-sample performance comparison of different decision-adjusted models **with enet** (in color other than black, M_4, M_8, and M_9) and **without enet** (in black, M_3)

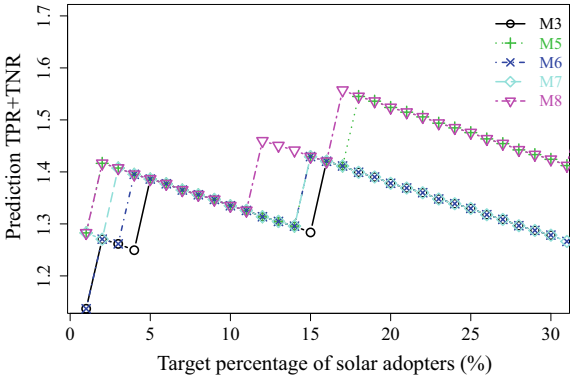


Fig. 4 Out-of-sample performance comparison of different decision-adjusted models **with indicator features** (in color other black, M_5, M_6, M_7, and M_8) and **without indicator features** (in black, M_3)

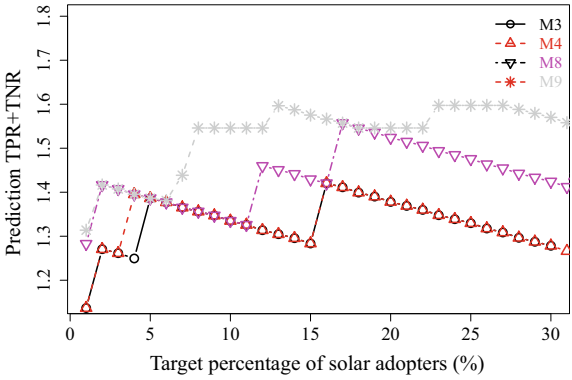


Fig. 5 Out-of-sample performance comparison of different decision-adjusted models **with enet** (in color other than black, M_4, M_8, and M_9) and **without enet** (in black, M_3)

Table 3 Coefficient estimates and their p-values are shown for Models M_0 and M_8. Note that Model 8 includes the demographics and indicator predictors

		Model M_0		Model M_8	
		Estimate	P-value	Estimate	P-value
X	(Intercept)	15.40	0.4302	16.95	0.4102
	income ($\times 10^5$)	0.74	0.3406	0.44	0.6046
	built	−0.01	0.3382	−0.01	0.3107
	sqfoot ($\times 10^3$)	−0.17	0.7127	−0.30	0.5112
	pool	0.64	0.3744	1.20	0.1044
	bdroomcnt	−0.34	0.3843	−0.45	0.2610
Z	neighbors	0.58	**0.0002**	1.28	**0.0000**
	hhnum	−0.48	**0.0498**	−0.23	0.4594
	mktval ($\times 10^5$)	0.29	**0.0247**	0.26	0.0963
I	$I^{neighbors}$			−4.49	**0.0059**
	I^{hhnum}			−1.10	0.4903
	I^{mktval}			1.24	0.1195

The logistic regression with enet penalty does not have legal significance testing, because the parameter estimators are not Maximum Likelihood Estimators (MLE), they don't have the asymptotic properties which MLEs have. The coefficients in model M_8 (without penalty) are better interpreted than coefficients in M_9. Thus, model M_8, a decision-adjusted logistic regression model with indicator features but without enet, is chosen over M_9.

Table 3 summarizes the coefficients estimates and their p-values for M_0 and M_8. M_0 is the traditional logistic regression model, and M_8 is the decision-adjusted model with all indicator features. In M_0, 'neighbors,' 'hhnum,' and 'mktval' are significant, while in M_8, neighbors and its indicator predictor are significant. This demonstrates that 'neighbors' indicator feature is more important than the other two features, and the corresponding decision-adjusted model M_5 is better than M_6 and M_7. Additionally, both 'neighbors' and its indicator predictor are significant, which implies that the indicator feature does explain what the original feature did not explain. None of the other two features, 'hhnum,' 'mktval' are significant, which means their indicator features also would not improve the model. For M_8, the threshold values δ for indicator features are

$$\begin{aligned} \delta_{neighbors} &= 3, \\ \delta_{hhnum} &= 4, \\ \delta_{mktval} &= 3.25. \end{aligned}$$

5 Discussions

This research focuses on building a decision-adjusted model of solar adoption. This model is able to precisely identify solar adopters with low false negative rate. In the meantime, we also identify predictors which improve the model fit.

All decision-adjusted logistic regression models, both with and without indicator features, are much better than the traditional logistic regression model in terms of TPR + TNR. The decision-adjusted logistic regression models with indicator features are better than the decision-adjusted model without indicator features, and the decision-adjusted logistic regression model with all indicator features is the best. Interestingly, the decision-adjusted logistic regression model with neighborhood indicator feature is better than decision-adjusted models with other indicator features, and its performance is close to the model with all indicator features.

Table 2 summarizes the TPR + TNR, TPR for different models in terms of in-sample and prediction performance. The results clearly show that the decision-adjusted models outperforms the logistic regression model, which performs poorly in terms of TPR. The decision-adjusted model makes a trade off between TPR and TNR, still it remarkably improves the TPR and keeps a low false negative rate. Comparison of models with and without enet, i.e., M_0 vs. M_1, M_3 vs. M_4, and M_8 vs. M_9, show that the enet does not improve the performance much. This is because the number of features in our model is not very large. However, Fig. 5 shows that M_9 is better than M_8. It indicates that for the decision-adjusted models with all indicator features, the enet penalty improves the performance given different target percentage of solar adopters. Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 in the Appendix show the confusion matrices of different models.

When the number of features is more than 20 or 30, the enet is expected to make a significant improvement in the performance of the model. Comparison of the decision-adjusted model with weighted model shows that the weighted model is not better, especially in terms of prediction. Also, the decision-adjusted approach has an advantage over-weighted model because it does not manipulate the input data. Table 2 shows that M_8, M_9 and M_5 have higher TPR + TNR than other decision-adjusted models. M_8, M_9 are the models with all three indicator features, and M_5 is the model with the indicator feature for neighbors. It tells us that the neighbors' indicator feature is more important than the other two indicator features.

Table 3 summarizes the coefficient estimates and their p-values for M_0 and M_8. For the model M_0, without any indicator features, the results show that neighborhood adopters, hhnum, and mktval are significant. A household with more neighbors who have installed solar panels and with large house market value is more likely to be a solar adopter. While a household with large household size is less likely to be an adopter. For the model M_8, which has three indicator features, the neighborhood adopters and its indicator feature are still significant. This means that the neighbors is more important than the other two predictors. The effects from neighbors, hhnum, and mktval are the same as in M_0. The estimated coefficient for 'neighbors' is 1.28, while the coefficient for the indicator feature of neighbors is −4.49. The 'neighbors' feature

suggests that the likelihood of adoption increases when there are more neighbors with solar panels, while the likelihood of adoption decreases when the number of neighbors who installed solar panels is very large (> 3). Hence, neighbors i.e., peer effect, play a significant role in adoption decision and this finding is consistent with the results of Rai et al. [10].

The coefficient of hhnum (-0.48 in M_0, -0.23 in M_8) shows that larger family size decreases the likelihood of adoption. The coefficient of I^{hhnum} (-1.10 in M_8) indicates that when the household size is very large (> 4), the likelihood of adoption will drop by a larger amount.

The coefficient of mktval (0.29 in M_0, 0.26 in M_8) shows that high market value of the house increases the likelihood of adoption. The coefficient of I^{mktval} (1.24 in M_8) indicates that a very high market value (> 3.25) will highly increase the likelihood of adoption. Also, high income increases the likelihood of adoption. A household with high house market value and high income is more likely to find the cost of solar affordable. Work by Vasseur and Kemp [9] also found similar results.

The neighbor feature is the most significant feature in predicting solar adoption. Considering the low penetration of solar adopters (0.02) in rural Virginia, the diffusion of rooftop solar panels is also low. This finding is consistent with Muaafa et al. [13] and Vasseur and Kemp [9].

In our study, we maximize TPR + TNR, where TPR and TNR have the same weights. Depending upon what information is more important to policymakers, different weights can be assigned to them and the objective function can be accordingly altered and optimized.

Acknowledgements The authors thank members of National Rural Electric Cooperative Association for conducting the survey of the electric cooperatives in Virginia. This work was supported in part by the Defense Threat Reduction Agency under Grant HDTRA1-17-0118, the Department of Energy SEEDS-II Grant DE-EE0007660, the National Science Foundation grant NSF-NRT DESE DGE-154362, NSF Grant CMMI-1745207 EAGER:SSDIM: Ensembles of Interdependent Critical Infrastructure Networks, and NSF IBSS Grant SMA-1520359.

Appendix

Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 summarize confusion matrices for five models under precision and TPR + TNR.

Table 4 In-sample confusion matrix of M_0

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	1	0	1
	Non-adopter	13	661	674
	Sum	14	661	675

Table 5 In-sample confusion matrix of M_1

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	1	0	1
	Non-adopter	13	661	674
	Sum	14	661	675

Table 6 In-sample confusion matrix of M_2

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	12	117	129
	Non-adopter	2	544	546
	Sum	14	661	675

Table 7 In-sample confusion matrix of M_3

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	10	81	91
	Non-adopter	4	580	584
	Sum	14	661	675

Table 8 In-sample confusion matrix of M_4

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	10	80	90
	Non-adopter	4	581	585
	Sum	14	661	675

Table 9 In-sample confusion matrix of M_5

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	10	24	34
	Non-adopter	4	637	641
	Sum	14	661	675

Table 10 In-sample confusion matrix of M_6

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	12	142	154
	Non-adopter	2	519	521
	Sum	14	661	675

Table 11 In-sample confusion matrix of M_7

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	12	121	133
	Non-adopter	2	540	542
	Sum	14	661	675

Table 12 In-sample confusion matrix of M_8

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	13	86	99
	Non-adopter	1	575	576
	Sum	14	661	675

Table 13 In-sample confusion matrix of M_9

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	13	87	100
	Non-adopter	1	574	575
	Sum	14	661	675

Table 14 Prediction confusion matrix of M_0

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	0	0	0
	Non-adopter	7	331	338
	Sum	7	331	338

Table 15 Prediction confusion matrix of M_1

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	0	0	0
	Non-adopter	7	331	338
	Sum	7	331	338

Table 16 IPrediction confusion matrix of M_2

		True condition		
		Adopter	Non-adopter	Sum
Fitted condition	Adopter	4	46	50
	Non-adopter	3	285	288
	Sum	7	331	338

Table 17 Prediction confusion matrix of M_3

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	4	48	52
	Non-adopter	3	283	286
	Sum	7	331	338

Table 18 Prediction confusion matrix of M_4

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	4	49	53
	Non-adopter	3	282	285
	Sum	7	331	338

Table 19 Prediction confusion matrix of M_5

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	5	56	61
	Non-adopter	2	275	277
	sum	7	331	338

Table 20 Prediction confusion matrix of M_6

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	4	47	51
	Non-adopter	3	284	287
	Sum	7	331	338

Table 21 Prediction confusion matrix of M_7

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	4	46	50
	Non-adopter	3	285	288
	Sum	7	331	338

Table 22 Prediction confusion matrix of M_8

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	5	51	56
	Non-adopter	2	280	282
	Sum	7	331	338

Table 23 Prediction confusion matrix of M_9

		True condition		
		Adopter	Non-adopter	Sum
Predicted condition	Adopter	6	86	92
	Non-adopter	1	245	246
	Sum	7	331	338

References

1. Department of Energy (DOE). Soft costs 101: the key to achieving cheaper solar energy (2016). https://www.energy.gov/eere/articles/soft-costs-101-key-achieving-cheaper-solar-energy. Accessed 25 Feb 2016
2. Department of Energy (DOE). Top 6 things you didn't know about solar energy (2016). https://www.energy.gov/articles/top-6-things-you-didnt-know-about-solar-energy. Accessed 6 June 2016
3. Mao, H., Guo, F., Deng, X., Doerzaph, Z.: Decision-adjusted driver risk predictive model using kinematics information. In: Submitted to IEEE Transactions on Intelligent Transportation Systems (2019)
4. Graziano, M., Gillingham, K.: Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment. J. Econ. Geogr. **15**, 815–839 (2015)
5. Aklin, M., Bayer, P., Harish, S.P., Urpelainen, J.: Economics of household technology adoption in developing countries: evidence from solar technology adoption in rural India. Energy Econ. **72**, 35–46 (2018)
6. Parkins, J.R., Rollins, C., Anders, S., Comeau, L.: Predicting intention to adopt solar technology in Canada: the role of knowledge, public engagement, and visibility. Energy Policy **114**, 114–122 (2018)
7. Alyousef, A., Adepetu, A., de Meer, H.: Analysis and model-based predictions of solar PV and battery adoption in Germany: an agent-based approach. Comput. Sci.—Res. Dev. **32**, 211–223 (2017)
8. Zhang, H., Vorobeychik, Y., Letchford, J., Lakkaraju, K.: Data-driven agent-based modeling, with application to rooftop solar adoption. Auton. Agents Multi-Agent Syst. **30**, 1023–1049 (2016)
9. Vasseur, V., Kemp, R.: The adoption of PV in the Netherlands: a statistical analysis of adoption factors. Renew. Sustain. Energy Rev. **41**, 493–494 (2015)
10. Rai, V., Reeves, D.C., Margolis, R.: Overcoming barriers and uncertainties in the adoption of residential solar PV. Renew. Energy **89**, 498–505 (2016)
11. Gupta, A., Hu, Z., Marathe, A., Swarup, S., Vullikanti, A.: Predictors of rooftop solar adoption in rural Virginia. In: Proceedings of the Computational Social Science (CSS) Annual Conference. Santa Fe, NM, USA (2018)
12. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Society: Ser. B (Stat. Methodol.) **67**, 301–320 (2005)
13. Muaafa, M., Adjali, I., Bean, P., Fuentes, R., Kimbrough, S.O., Murphy, F.H.: Can adoption of rooftop solar panels trigger a utility death spiral? a tale of two U.S. cities. Energy Res. Soc. Sci. **34**, 154–162 (2017)