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.

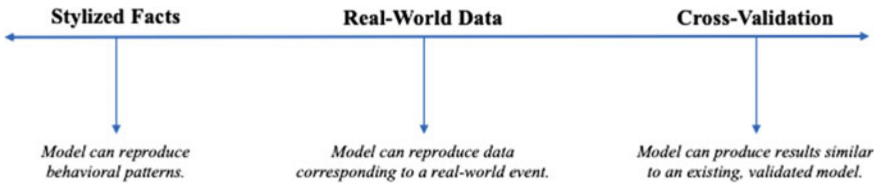

Fig. 1 Diagram of ABM validation as a spectrum defined by the data the models produce, ranging from reproduction of behavioral patterns, real-world event data, or data produced by a different model

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

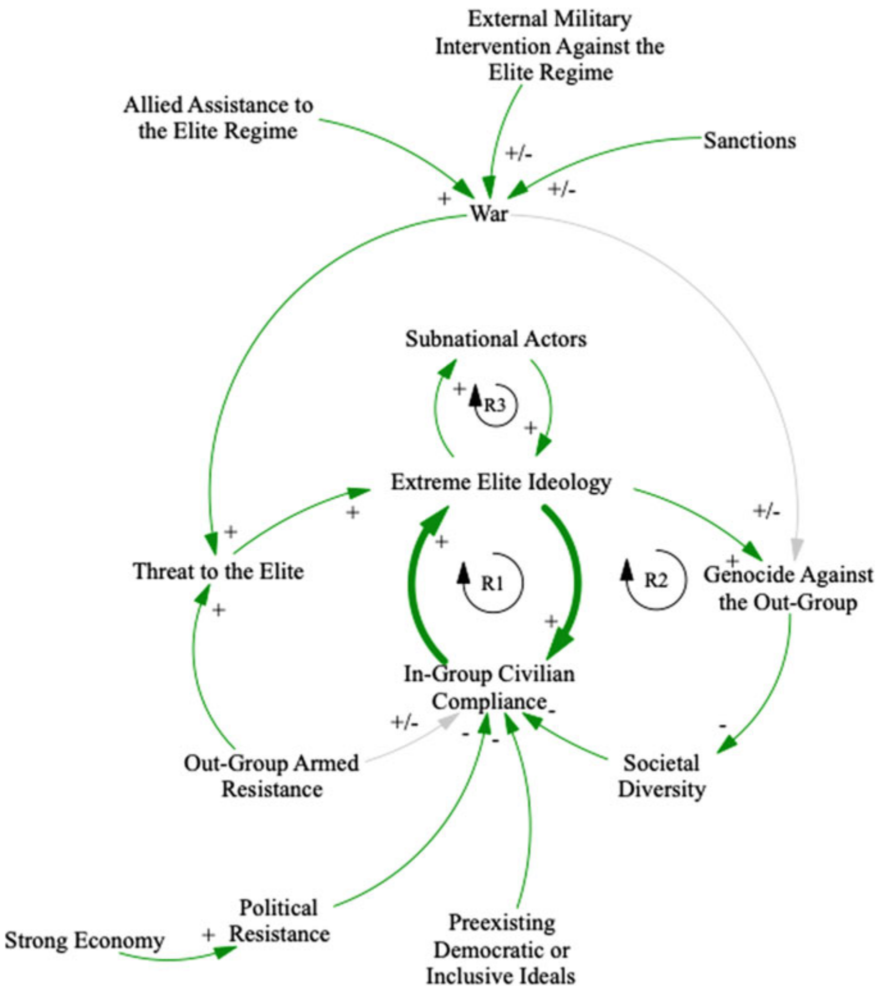

Fig. 2 Causal loop diagram of the system-level dynamics of genocide [10]

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."

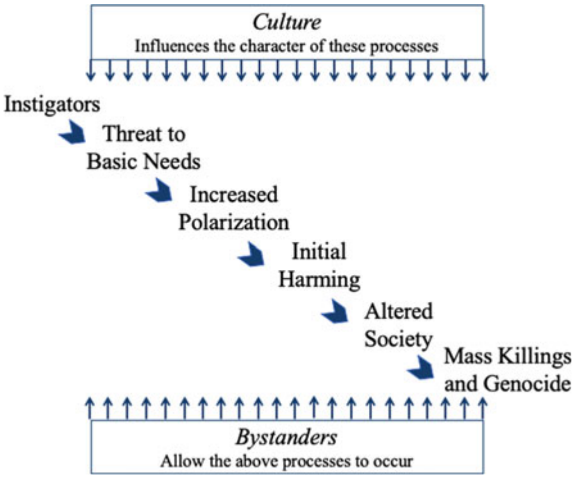

Fig. 3 Diagram of dynamics occurring in a society moving toward genocide, highlighting the influence of culture and bystanders on the process. Adapted from [25]

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-

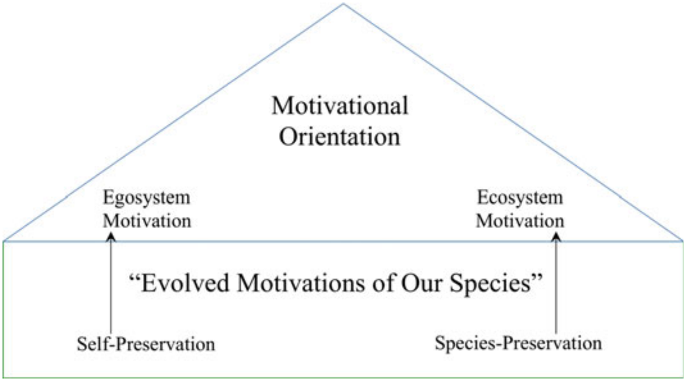

Fig. 4 Visualization of the scaffolding of motivational systems as defined in the ecosystem and egosystem theory of motivational orientations [14]

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

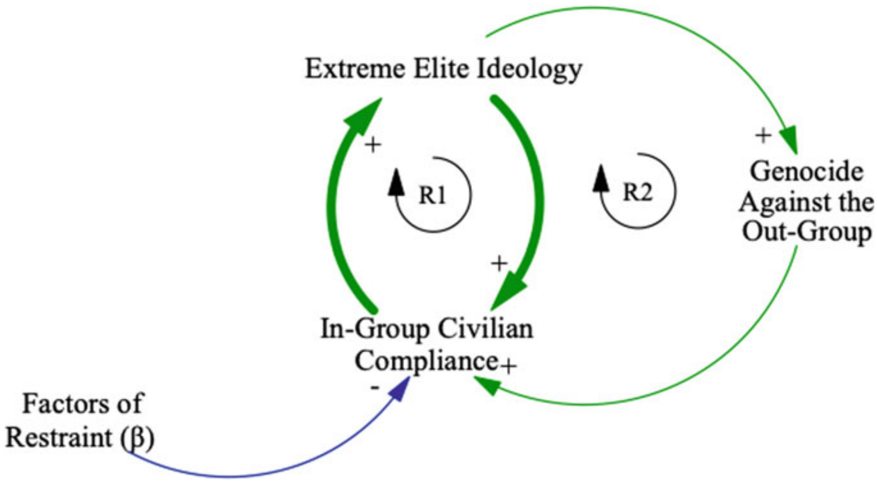

Fig. 5 Causal loop diagram showing elite and in-group civilian dynamics in the presence of generalized factors of restraint as a system-level function β

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.

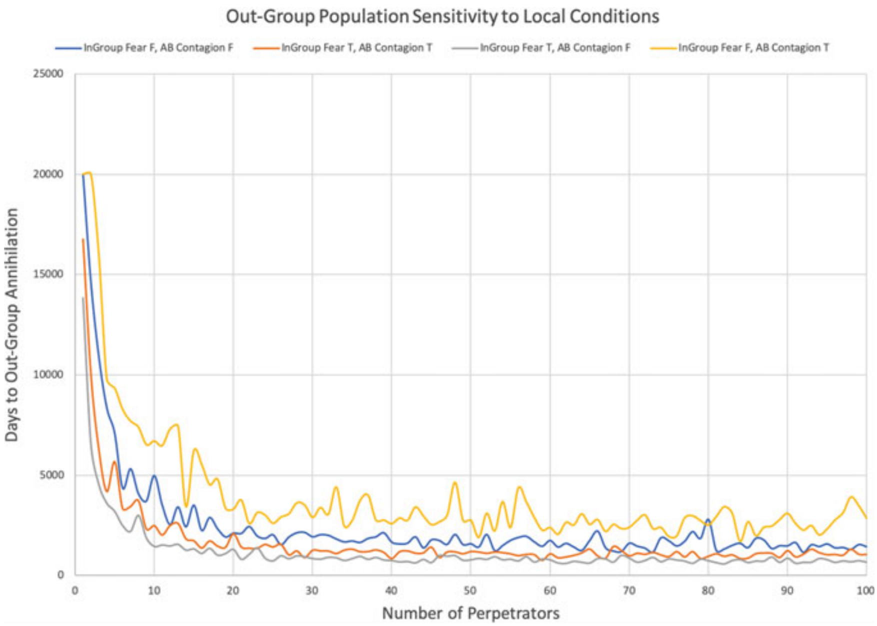

Fig. 6 Out-group population sensitivity to all combinations of use for Algorithms 3 (Active bystander contagion effect) and 5 (In-group fear in the presence of violence)

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.

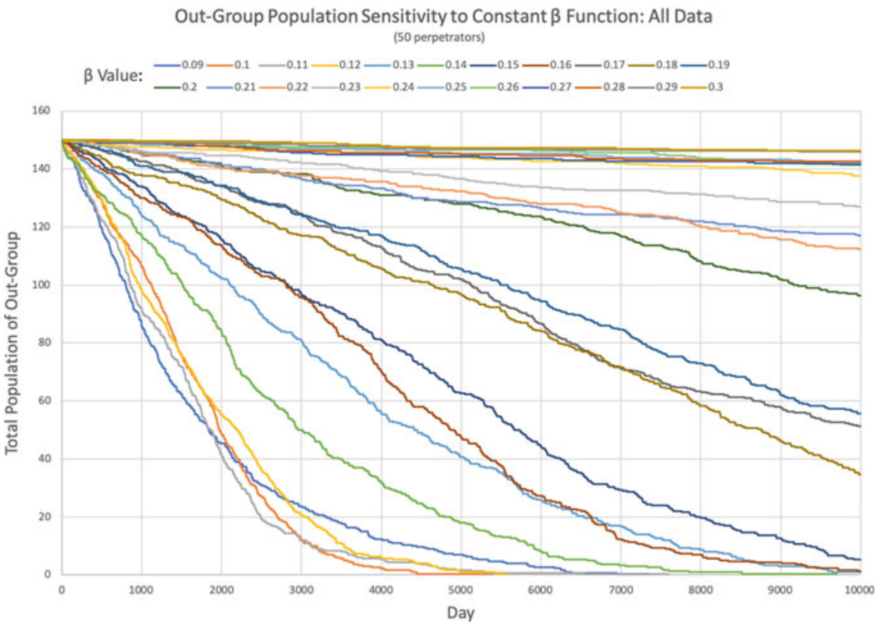

Fig. 7 Out-group population sensitivity to a constant β function providing system-level restraint against out-group violence

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-

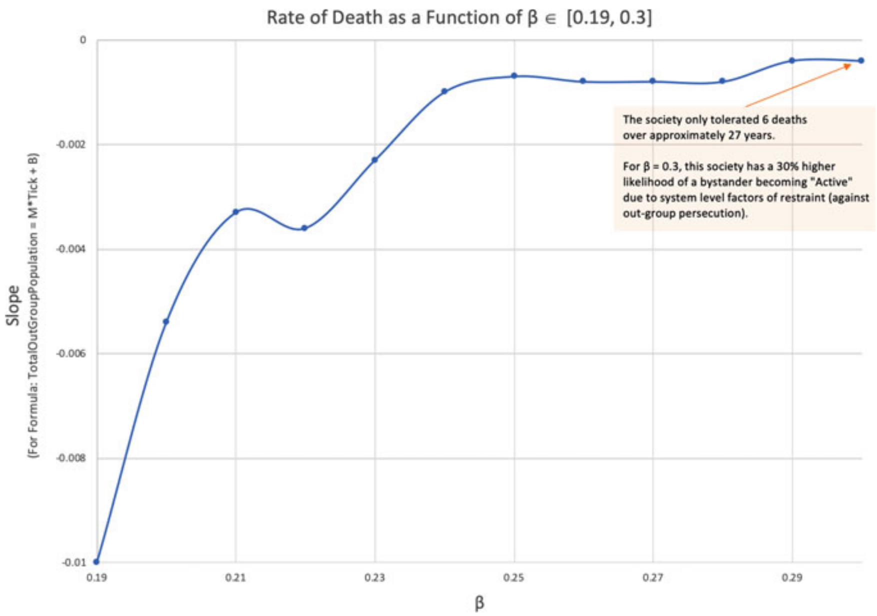

Fig. 8 Effect of β on out-group rate of death according to slopes of lines in Fig. 7 for $\beta \in [0.19, 0.30]$

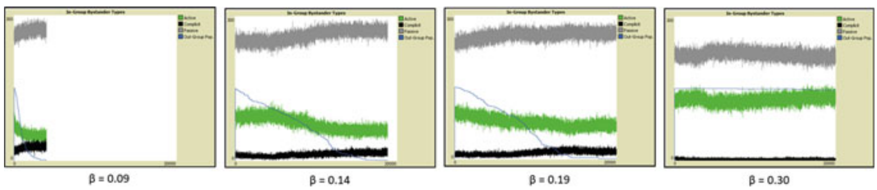

Fig. 9 Visualization of total number of bystanders according to type, over time, for given β values

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