Social Simulation for Social Justice

Melanie Dickinson, Noah Wardrip-Fruin, Michael Mateas

mldickin@ucsc.edu, {nwf, michaelm}@soe.ucsc.edu

Expressive Intelligence Studio

Computational Media Department

University of California, Santa Cruz

Abstract

We argue that social simulation can help us understand social justice issues. In particular, modeling certain social dynamics within computational systems can be used to creatively explore and better understand the social and identity dynamics of oppression. Writing theories of oppression in code forces us to explicate everything, and question what we leave out or what we can't account for. As an early step in this direction, we present an in-progress social simulation of group discussion in activist meetings, developed in the already-existing AI system, Ensemble. Through this minimal, highly constrained social arena, we can explore wide-reaching phenomena like privilege, intersectionality, and power dynamics in nonhierarchical groups, but in a way that's grounded in concrete, person-to-person interactions. We propose that this kind of social simulation can aid in the process of unlearning hegemonic ways of being, and imagining liberatory alternatives.

Introduction

The Personal is Political

Interpersonal social dynamics can be invisible vehicles of oppression and inequality. Among many reasons oppression is damaging to individuals is that it is experienced everyday, in interactions with friends, strangers, colleagues, and even internal streams of thought. This kind of personally felt oppression doesn't require malicious agents, racists, or sexists, but rather societal constructs and patterns of thought that influence everyone invisibly (unless explicitly made visible, which can be done by scholarly analysis or works of media, including computational media).

A key tenet of feminism since the 1960s is that "the personal is political," or that the seemingly private issues experienced by marginalized individuals are often patterned and systemic, and therefore not *just personal*. In her seminal 1969 essay by that name, Carol Hanisch points to the consciousness-raising, analytical value of talking about personal problems in an activist group, defending it from some fellow activists who considered it mere "therapy," irrelevant to political action (Hanisch 1969).

Two decades later, Peggy McIntosh articulated visible, day-to-day evidence of privilege in her own life as a white

woman (McIntosh 1988). This evidence included many social phenomena, including whether other people are generally welcoming, pleasant, and supportive to her in daily interactions, rather than suspicious, closed off, or even violent. In her words, white privilege afforded her "comfort, confidence, and obliviousness," which manifested especially clearly in her interactions with others.

Since the 1970s, there has also been a vein of this thinking in psychology, called *pscicología de la liberación*, or liberation psychology, focused on the effects of oppression on the mental health of people and their communities (Martín-Baró, Aron, and Corne 1994; Burton and Kagan 2005; Moane 2010). This body of scholarship seeks to bridge the same gap that feminism observed between the personal and political, an isolating, depoliticizing assumption of mainstream academic knowledge and popular consciousness.

The Personal, as theorized in these and many other foundational feminist and social justice works, is the arena in which the effects of much larger systems of oppression are daily and viscerally experienced by ordinary people. One of the most important practical functions of critical race and feminist theory is to offer marginalized people a more liberatory, healing lens through which to see the world and themselves. Critical theories give people an opportunity to "break with the hegemonic modes of seeing, thinking, and being that block our capacity to see ourselves oppositionally, to imagine, describe, and invent ourselves in ways that are liberatory" (hooks 1992). This unlearning of hegemonic worldviews is a deeply personal process, that can helpfully inform interpretation of social interactions, identity, and personal difficulty. Even if these theories can't immediately dismantle entrenched oppression, their teachings of personal and interpersonal liberation can be internally healing, and can be enacted for mutual healing in our immediate social surroundings.

Personal-Political Simulation

The Personal is also the domain of expressive social simulation, which we argue can be designed into playful tools to explore and critically interrogate the political dimension of the Personal. We can create social simulation systems to augment and facilitate the process of unlearning hegemonic worldviews, and imagining liberatory alternatives.

Social simulation has been defined somewhat broadly and

informally for a game design context by Mitu Khandaker-Kokoris as "anything that allows social interactions with or between [non-playable characters] to meaningfully affect the outcome of a situation" (Khandaker-Kokoris 2015). These social interactions are computationally under-girded by some degree of either hand-written branching, as in Telltale's *The Walking Dead*, or emergent autonomous behavior, as in *The Sims*. In her talk, Khandaker-Kokoris advocates for greater investment in games about meaningful social interaction in general, whether precisely authored or procedural, but places particular emphasis on the latter end of the spectrum.

We take this further, defining a social simulation more narrowly as *a computational model which affords meaningful social interaction with or between autonomous characters*. For the purpose of this paper, a social simulation uses a simulation or emergent system under the hood, and is not hand-authored branching. It uses processes to determine what interactions occur between characters, the content of those interactions, and how the characters are affected by the interactions, which in turn impacts what interactions occur next. For our purposes, a social simulation also doesn't need to exist in a playable experience or game, though often does. Simulations about social interactions allow players or users to play or experiment with their existing ideas about the social world. They still afford experimentation and playful interaction without a corresponding game experience.

It is because of this procedurality that social simulation is uniquely positioned to mediate social justice and injustice. Even in her general overview of social simulation in games, Khandaker-Kokoris points to the potential for social simulation to discuss issues of privilege, and the importance of a diversity of characters in play centered around social interaction.

The Case for Social Simulation for Social Justice

Computational media, and social simulation in particular, can be uniquely leveraged in support of social justice.

The expansive systems of oppression which social justice seeks to dismantle are notoriously difficult to understand. There are many forces at play which are tightly interconnected and hard to observe holistically. Playable experiences afford iterative, pleasurable learning of underlying systems (Wardrip-Fruin 2009).

Furthermore, modeling issues of identity, power, and oppression in code can inform critical interrogation of mainstream technical applications. Insights gained in building and analyzing socially conscious social simulations can be applied to software which, for example, mediates virtual social interaction, or influences policy decisions based on large social data.

Social simulation is a **systems-based media genre**, which can enact the complex processes and interrelations of systemic oppression. These systems are hard to convey a nuanced representation of in prose or film, and hard to make accessible to a broad audience in scholarly, analytical writing.

Linear Storytelling about Social Issues Though engaging, other fictional media representations of these topics only show audiences brief, heavily filtered glimpses of an underlying social theory. They aren't working models of that theory. One piece of such media can show just one example of how specific characters in a specific context experience oppression or act for change. These are hugely valuable for social justice theory communication because they are often emotionally rich and grounded in a specific context. Audiences can therefore understand them and engage with them emotionally. These stories can *suggest* larger structures and more common experiences, but don't have to speak in a language that accounts for these larger, more abstract ideas, as more theory-centered work does.

But they also run the risk of audience over-generalization of the snapshot they present. Media about social injustice like *Precious*, *The Color Purple*, and *Slumdog Millionaire* are critiqued for their perpetuation of negative stereotypes about marginalized subjects, while they are acclaimed for amplifying stories about real pain that real people experience due to systemic, entrenched oppression. Both evaluations are justified. These kinds of media are often both deeply poignant as well as problematic for depicting only one kind of experience of people at an intersection of identities. If audiences are critical and conscious of the problematic, the snapshot mode of media-making is fine—they already have a relatively complex view of people, identities, and inter-personal politics. More insidious are media like these in front of audiences who aren't as aware of identity politics and non-hegemonic ways of thinking and being. They tend to see *only* the entrenched stereotypical representations of an oppressed group of people. And so, the depiction of a marginalized subject which should evoke empathy and make invisible experiences visible, instead reinforces harmful stereotypes.

Systems representations, on the other hand, could potentially generate multiple examples of how the same set of abstract dynamics can play out in grounded contexts, whether by player input, non-deterministic processes, given different initial states, or by tweaking system parameters. They are capable of dealing in both obtuse, abstract theory, and more broadly legible, enacted interactions between fictional people. Audiences can see multiple views of the same underlying theory through concrete character interactions. They can see the destructive effects of oppression on individuals, which may be consistent with stereotypes, as well as liberatory alternatives, which contradict and push against mainstream conceptions of marginalized peoples.

Without extreme care, diligence, and self-critical design, however, social simulation could provide multiple views of a problematic underlying model—which is perhaps even more harmful than a single problematic view. In our evaluation of systems which strive to model social issues, we need to account for this. How to go about measuring how problematic a social simulation is is a difficult, open question.

Analytical Essays about Social Issues Another way that people get exposed to ideas of social justice, outside of lived experience and word-of-mouth, is through critical theory

texts. Non-fiction writing on social justice topics is widely available, especially with the growth of socially critical academic fields (like feminist studies, critical race and ethnic studies, and *pscicología de la liberación*) in the past fifty years. These more analytical—rather than story or entertainment-driven—forms for social justice theory communication can cover theories in great depth, but are often difficult to internalize and put into practice for heady abstractions, jargon, and academic writing style. Much of the theoretical work to come out of feminist or womens studies is "too esoteric, too divorced from the reality of women's experiences, too inaccessible to the majority of women whom feminism is supposed to serve" (Rowland and Klein 1996).

And yet, past the academese and paywalls, people can find work which provides an important basis for understanding their own experiences, those of others, and larger systems of oppression. Practical activism, personal liberation (or "decolonization of the mind" (hooks 1992)), and consciousness-raising all need theory, just as theory needs action.

Systems-based media is a way to marry accessible, concrete example with high-level theory. System-based media works are importantly still media; they are designed to engage with audiences in an expressive, emotional, aesthetic vein.

We submit a demo of a social system engaging issues of power and identity as proof of concept and early exploration of this position or genre. In particular, the social domain being modeled is that of small group meetings for social activism. This is a uniquely situated domain for this goal for a number of reasons. Due to the conscientiousness and self-critical organizing philosophies common to social justice activists, there is much written knowledge of how to create and uphold liberatory cultures in movements (Starhawk 2011; Chen, Dulani, and Piepzna-Samarasinha 2011; Smucker 2017).

Of course, there are countless other possible social simulations for social justice, with readymade critical sources to draw from. Other projects might model intimate relationships to test and explore theories of abusive partnerships and codependency (Chen, Dulani, and Piepzna-Samarasinha 2011). Another might model the disruptive effects of incarceration on close relationships and local social networks (Price 2015). Still others could explore alternatives to capitalism or prisons.

Related Work

While relatively few compared to social simulations in general, there are existing works of social simulation for social justice.

The Advanced Identity Research (AIR) project, led by D. Fox Harrell, is a significant provenance of this work, and has contributed many computational media experiences which systemically, expressively model issues of identity (Harrell 2010; 2013).

Parable of the Polygons (Hart and Case 2014) is a compelling simulation of segregation, that is interactive and playful, though not a traditional game. It is illustrative of the many different forms that social simulation can take.

And finally, Dog Eat Dog is a tabletop social simulation about colonialism (Burke 2012). While non-digital, social interaction between characters in this system produces emergent statements about the effects of colonialism.

Ensemble

We are creating the activist meeting demo in an AI architecture called Ensemble, which is a rare off-the-shelf system for authoring social simulations.

The Ensemble Engine is a Javascript, rules-based social physics engine for modeling autonomous characters and the interactions and relationships between them (Samuel et al. 2015). It is based off of Comme il Faut, the system underwriting Prom Week (McCoy et al. 2014).

Ensemble allows authors to create domain-specific social state, character volition rules conditioned on that state, and character actions which affect the state. The result of this authoring is a system which can be plugged into a game engine or interacted with through the command line or an authoring tool. The system can then be queried for what a character wants most at a given time, for actions that the character wants to do and can do to fulfill that desire, and for the state of the world after that action is performed.

The engine uses rules composed of logical predicates with dynamically bound variables. For instance, the author may write a rule such as: "If a character, X, is friends with another character, Y, and Y is going through a tough time, then X wants to provide emotional support for them." Giving emotional support would be a high-level character desire, or *volition*, that could be fulfilled with one or more *actions*. Given a particular game state and pair of characters for which the conditions hold, the system would increase the "provide emotional support" *volition score* for one specific character to another. Scores for this and other volitions would similarly be increased or decreased by rules like the above (called *volition rules*).

Once the top volition score is determined (i.e., what high-level desire one character most wants to fulfill), the system could be queried for the best action to fulfill it. In our example, it might return that the character bound to variable X wants to provide emotional support to the character bound to Y by offering them a hug. The offer may be accepted or rejected by the latter, and in either case, would affect the social state which is queried in future timesteps with respect to the conditions of our volition rules.

Current social state is used to condition volition rules, and is modified by action effects. The kinds of state that Ensemble maintains fall into three broad direction types: *reciprocal*, *directed*, and *undirected*. Reciprocal and directed state categories describe relations between two characters. Reciprocal state is for symmetrical relationships, while directed state is about how one character asymmetrically thinks or feels toward a second character. Undirected state captures information about a single character, not in relation to another. It is up to the author what specific kinds of reciprocal, directed, and undirected state information is captured and maintained by their social simulation.

A final aspect of authoring in Ensemble is designing how and when its functions (volition calculation, get actions, ap-

ply effects) are called and used in a game engine or other front-end interface.

Social Justice Domain Authoring

In this section, we describe the Ensemble domain created for this project. We first explain the decision to model activist meetings, then give a broad overview of the social world and afforded space of actions, followed by a tour of its key observable phenomena as informed by activist and feminist studies literature.

As stated earlier, this project models small group meetings for social justice activism. This is designed to emulate local groups composed of people with multiple identities, central goals and issues for activism, levels of commitment and connection to the group, and degrees of socio-political awareness. Imagine people in a community renting or commandeering a meeting space or member's living room, arranging chairs in a circle, and talking about issues that effect them and how to act around them.

Picking a Domain

We chose this cultural setting because it is concrete, realistic, readily connected to various topics in social justice ideology and theory, and personally meaningful to and experienced by the authors.

The meeting domain is *concrete*, in that it models character actions at high enough fidelity that players can observe analogous actions in everyday situations. For example, people in group meetings interrupt each other, listen more or less intently when other people speak, and sometimes clap or snap when they agree with what someone has just said. These are all easily observable actions in activist meetings "in the wild" (Della Porta and Rucht 2013), just as they are highly visible in the simulation. We argue in the introduction section that this concreteness is one of the affordances that differentiates social simulation from traditional analytical theory writing, but that simulation shares with linear storytelling.

We further propose that concreteness of domain in a social simulation is especially important if it is a simulation for social justice. A concrete domain enables players to see sequences of actions and events that could happen, and gives space for players to empathetically relate to those actions, much like a narrative does. They may place themselves in the scenario, imagine themselves as social justice actors if they hadn't before, imagine themselves doing particular kinds of activism, and perhaps even be more willing to act or organize in the future. Less optimistically, concreteness in a social simulation for social justice (like the activist group meeting) gives players concrete examples of action sequences that they might fundamentally disagree with, or fail to connect empathetically to. In this case, there are examples that players or fellow researchers can point to as feeling 'wrong' to them. Furthermore, it is easier to explain why a particular example of behavior is wrong than to do so for an abstract model of behavior.

Parable of the Polygons, for instance, is far less concrete, and therefore suggests less immediate action from its reader-players. The main character action is to move to a random place until happy, where happiness is defined by a threshold of neighbor similarity to you and (later in the article) difference from you. While a player can occasionally observe people moving houses in their real world, moving to a random neighborhood until happy it is not a realistic action sequence to emulate. While an engaging, powerful account of self-segregation, it's hard to know what to do about it from the simulation alone.

The activist group meeting domain was also chosen for being a *realistic* depiction of social justice topics. It isn't veiled in fantasy or science fiction theming. It is explicitly, directly about contemporary concepts of human race, gender, and politics. While fantastical metaphors for real-world identity issues are useful for the safe distance they create between players and difficult issues, unabashedly realistic, modernized treatments are far less explored in games, and present their own design challenges. We hope to see many more realistic, direct treatments of social justice and identity issues in games in the future.

This domain is heavily drawn from one of the author's personal experiences in autonomous students groups, anarchist groups, student media collectives, and graduate student governance bodies. The central locus of involvement with these groups is always group meetings, with more or less frequent bursts of individual work and larger collective events. Group meetings, while less dramatic and less publicized than marches and direct actions, are where crucial group identity development and strategic planning take place. If groups don't have regular conversations, where everyone is allowed to talk and feels comfortable doing so, with a place for hard, ideological and practical discussion, it is difficult to do anything, let alone maintain a cohesive, committed, and ever-critical whole.

These meetings can also serve as a grounded microcosm for larger-scale discussions in and about entire social movements (e.g., conversations about race and gender that storm and percolate across channels around things like #BlackLivesMatter, #SayHerName, and the 2017 Women's March).

To inform the knowledge engineering work described in the following sections, we drew from personal experience and from activist and social science literature. Three especially helpful sources were *The Empowerment Manual* (Starhawk 2011), *Beautiful Trouble* (Mitchell and Boyd 2013), and *Meeting Democracy* (Della Porta and Rucht 2013).

State

Now that we have notion of what kind of social world we want to reflect back to us in code, we can start developing in Ensemble. Generally, the first step in coding an Ensemble domain is to define the social schema, or what kind of state is maintained for each character or character pair. This state is what volition rules will reference in conditions, and what will be changed by action effects. An Ensemble schema is composed of some number of social state *categories*. Each category is a data object with the following properties: category name, category types list, direction type, whether the

Category	Types	Direction	Type	Actionable	Use Convention
definition	person identity issue group	undirected (only about one character)	boolean	not actionable (characters *cannot* change their definition)	Only one of these types is true at a time, the rest are false
trait	emotionalEnergy mentalEnergy anger guilt patience selfLove empowerment	undirected	scalar	actionable (characters *can* form volitions to change their trait values)	Only relevant for actual characters (where person=true)
relationship	respectFor closeness supportFrom supportTo	directed (from one character to a second character)	scalar	actionable	Only relevant for person → person relationships
politics	personallyInvestedIn activeIn awareOf	directed	scalar	actionable	Person → issue
identity	identifiesWith experiencesOppressionAs awareOfNormsOf	directed	scalar	actionable	Person → identity
group	memberOf associationWith perceivedEfficacy perceivedImportance perceivedEthics	directed	scalar	actionable	Person → group
state	inMeeting hasTheFloor	undirected	boolean	actionable	
discussion	speakingTo	directed	boolean	actionable	

Figure 1: The current Ensemble schema for the activist meeting domain. All scalar categories range from 0-10, with 0 as the default value. False is the default value for all boolean categories.

category is boolean or scalar-valued, whether characters will be allowed to form volitions about the category or not, and default, minimum, and maximum values.

Figure 1 describes each category defined for the activist meeting domain. Each character listed in the Ensemble cast file will have a value for every undirected category type, and each character pair will have a value for every directed type (there are no reciprocal categories in this schema). They will start at timestep 0 with the default value of each category (false or 0), unless specified differently in the history. The history is an array of data objects, each object containing a timestep, character, schema category type, and value for that type. The history object is automatically updated according to action effects throughout the simulation, but can be initialized with the state of the world before the simulation starts.

We model individuals as entities with *undirected* state involving variables like: emotionalEnergy, mentalEnergy, selfLove, anger, guilt, etc. While the variable content is unusual, this type of individual state-keeping is a common pattern in games, even those without social simulation.

Characters can also have directed relationships with other characters, another ubiquitous pattern in social simulation. Our character relationship content is closely tied to the activist domain.

Less common, even in Ensemble-based games, is our modeling of identities, issues, and the group as Ensemble "characters." This means that individuals can have relationships with them, form volitions about them, and direct actions toward them. These concept characters are first-class citizens in the computational model, just as actual characters are.

Individual characters can have *directed* relationships toward an identity, issue, or group (all scalar-valued). Each target implies different relevant types:

- **Person-to-identity** relationship types include: identifiesAs, experiencesOppressionAs, and awareOfNormsOf. Characters relate in these ways to target identities like race (e.g., Black, Latinx, white, etc), gender (nonbinary, feminine, masculine), and cultural background (immigrant, ESL, etc). We plan to add identity categories like class and ability in the future.

- **Person-to-issue** relationships: awareOf, personallyInvestedIn, and activeIn. Characters can relate to issues like anti-oppression, police brutality, colonialism, capitalism, food justice, and war. It is less important for this project, at least at its current stage, what the content of particular issues and identities is. We focus instead on

the relationships people have to those issues, in a fairly content-agnostic way.

- **Person-to-group** relationships: `memberOf`, `involvementIn`, `perceivedEfficacy`, `perceivedEthics`, etc. For simplicity, there is only one group in the current simulation, but adding more is clear future work. A simulation about multiple groups would presumably call for the addition of rules that interrelate different groups and their members. In such a setting, you might also author rules and actions that treat groups as characters with agency, that can form volitions and act on their own, perhaps representing collective action, which we currently do not do for our model.

We chose to define identities, issues, and groups as "characters" in order to take advantage of Ensemble's agent-centric modeling paradigm. Ensemble's primary reasoning functions are the calculation of character volitions (primarily targeted at themselves or a single other character), and the calculation of the best actions for a character to take to satisfy their volitions. By representing identity and ideology at a character level, we can write rules like "if character A identifies as identity I and character B does not, but is aware of the norms of I, and there has just been a misunderstanding from C to A, then B has a higher volition to clarify A." An alternative approach would be to contain identities, issues, and group information in undirected state for each character. But this quickly becomes unwieldy for both rule authoring and adding or removing identities, issues, and groups.

The representation of state concepts around identity is of special import in computational modeling of social justice issues. Because each identity is a separate, complex object in our model, one character can identify as more or less of multiple races, cultures, etc. For example, a mixed-race character with light skin may have an `identifiesAs` relationship with the identity `Black` of value 5, and value 2 with the identity `white`. Contrast this with the boolean and mutually exclusive racial or gender categories of mainstream RPGs and roguelikes.

Another character might have `identifiesAs Black/5` and `Latinx/7`, but `experiencesOppressionAs Black/7` and `Latinx/2`, suggesting that they are more often seen as Black by the rest of the world. A similar multivalent identity connects a character to different gender concepts. It will perhaps always feel reductionist to distill identity concepts like this in code, but we propose that this semantic network-like approach is complex enough to capture a high degree of nuance while being still manageable for authoring.

Intents

Once you define a schema in Ensemble, you can derive high-level character desires (or intents, or volitions) from them, which will direct your volition rule and action authoring. Intents are what appear on the right-hand side of volition rules. They are predicates specifying the schema category, type, and direction to push the corresponding

state value. In the schema, the property `actionable` under a category indicates whether or not characters should be able to form volitions to change their values. For example, `emotionalEnergy` is the first type under the first `actionable` category in Figure 1. This suggests that characters should be able to form a volition (using volition rules) to raise or lower their or another character's `emotionalEnergy`. For example, we have a simple, Sims-like volition rule of the form, 'if someone has less than maximum (10) emotionalEnergy, they tend to want to raise it (+1 to the volition raise emotionalEnergy).' For `emotionalEnergy`, it makes less sense for a character to form a volition to lower its value, but that is the second intent that may be directly derived from that schema concept.

Some of the high-level intents that characters can form, which are tightly coupled to the schema as illustrated in Figure 1, include but are not limited to: learn about an identity, learn about an issue, support person in group discussion, call out person for oppressive behavior, join group, leave group, raise involvement in group, forward issue, center a person in group, raise one's social power in a group, and distribute group power.

Now we can determine what actions characters can perform to fulfill those high-level desires.

Actions

Actions in an Ensemble domain are modeled as an arbitrarily connected action graph. The graph connects intent predicates to terminal actions. The simplest action graph is one where each intent appears in a single terminal action node, and there are no connections. In that case, intents correspond directly to actions, and action calculation given a volition is trivial. This doesn't take full advantage of Ensemble's action model, however. Adding nodes and connections allows you to specify conditions on action nodes, so action calculation involves traversing the graph, from an intent node to a terminal action. Thus, the same high-level character desire can be realized through different actions depending on context. Furthermore, the same action can be used to achieve different social goals.

Actions that characters can enact in the activist meeting domain are mostly speech acts or nonverbal communication cues. They include but are not limited to:

- Introduce yourself to someone
- Say you'll do something
- Quietly take responsibility for something
- Admit to a mistake
- Ask for someone's thoughts on a subject
- Stay silent (to give quieter people time to speak)
- Stay silent (because you are quiet)
- Make a point hesitantly or awkwardly
- Interrupt a speaker (because you strongly disagree with them)
- Interrupt a speaker (to amplify and clarify what they are saying)

- Talk over the other person when someone interrupts you

- Hold side conversation

Recalling that our focus is on group discussion in a meeting setting, there are special considerations involved in enabling this with Ensemble. Ensemble actions can involve any number of characters, but are typically between just one or two, that are cast into roles specified by the `first` and `second` action object properties. But actions in group discussions usually involve as many characters as there are in the group (rarely just two). Typically, there is one acting character, whose action is directed towards the rest of the group, as in the case of someone speaking in a meeting. In those cases, the abstract `group` "character" is used as the action target. Yet a speaking character may alternatively direct their words at one or two audience members in particular, rather than scanning the entire group, which plays a role in group participation patterns and relations.

To simulate semi-real-time group discussion that allows for dynamics like mid-utterance interruption, we recalculate volitions and actions (calling corresponding Ensemble functions) for all characters between and during other characters' actions. A few volition rules give relatively high values to volitions resulting in "stay silent" actions when someone else is speaking (or `hasTheFloor`), but tense sequences of events can lead interruption-causing intents to outweigh silence. When no one `hasTheFloor`, the otherwise high-valued stay silent action becomes more comparable in score to other actions and their corresponding intents. We choose the top-scoring intent and action among all characters given the state of the world, so the character that most wants to speak does so next, and the cycle of mid-utterance reasoning by the other characters begins again.

Besides interruption, this design approach allows characters who don't currently have the floor to act and communicate non-verbally while another character speaks. They may want to show signs of listening intently to a character who they want to support, or leave for an opportune bathroom break when an overzealous colleague has dominated the meeting for a while, etc. We anticipate this pattern being useful in the design of many other Ensemble social simulations which seek to model character interaction at a finer level of granularity and with greater dynamism in character performance.

Modeled Phenomena

Once we know the space of actions characters have to pick from, we decide what factors might lead a character to want to do those actions; or what volition rules influence them. Volition rules usually take the most time and effort to author in an Ensemble domain. For this project, we first identified phenomena from source texts on feminist theory, activist organizing, and social movement studies that highlight how oppression and inequality can be recreated in activist context, despite the good intentions and socio-political awareness of the social actors involved. Some of the phenomena identified include activist aversion to hierarchy or formal leadership, leading to the "tyranny of structurelessness" (Freeman 1972), fluid and distributed responsibility, power-

from-within, social power and privilege, group norms and misunderstandings, 'splaining, elitism, white feminism, and decentered voices. As a brief example of one of these, consider the last phenomenon.

Decentered Voices A common pitfall of activist groups occurs when the loudest voices, the ones that lead and influence the collective effort, are not those most in need of that effort, and therefore not the best suited to design effective strategies. Because we are enmeshed in a society that devalues certain voices in favor of others, it can be hard for even the most socially conscious groups to continuously reject the habit, and prevent this default dynamic from rendering well-intentioned organizations ineffective at best, and actively oppressive at worst.

This phenomenon can be summarized as: *Without intervention, it is easier for some populations or identities to speak and be heard than others.*

Some of the relevant influence rules guiding this dynamic include:

- People with social power tend to speak more.

- People tend to listen more attentively to people with social power.

- People tend to interrupt people more if they have less social power.

- New or inexperienced activists tend to talk less.

- People are less likely to speak if they are the only one of an identity in a group.

- People are more inclined to leave the group if they consistently feel unheard.

Conclusion and Future Work

Social simulation can help us understand social justice issues. We bring feminist theory, like the credo, "the personal is political," to bear on computational media, resulting in the natural connection to social simulation. We argue that writing social justice theories in code forces us to understand them in a different way. And interacting with computational models of them affords a different kind of audience understanding and engagement than otherwise possible, due to the unique affordances of simulation and computational media. To illustrate this position, we present an Ensemble-based, in-progress social simulation of group discussion in activist meetings. Through this minimal example, we also present two unusual design approaches for Ensemble-based simulation: modeling abstract concepts as characters, and tightening the core computation loop for semi-real-time group conversation and simultaneous character action.

Evaluation is the next step. We plan to run user studies with activists to test whether they can observe the phenomena we intended to capture or any we didn't intend to capture, whether the simulation surfaces practical insights for their organizing work, and whether the simulation helps them imagine places for computational media and social simulation in their activist work. We further hope to raise the question of the simulation aiding in the process of unlearning hegemonic worldviews and imagining liberatory alternatives with non-activist users.

Acknowledgements

References

Burke, L. L. 2012. *Dog Eat Dog*. Liwanag Press.

Burton, M., and Kagan, C. 2005. Liberation social psychology: learning from latin america. *Journal of community & applied social psychology* 15(1):63–78.

Chen, C.-I.; Dulani, J.; and Piepzna-Samarasinha, L. L. 2011. *The Revolution Starts at Home: Confronting Intimate Violence within Activist Communities*. South End Press Brooklyn, NY.

Della Porta, D., and Rucht, D. 2013. *Meeting democracy: power and deliberation in global justice movements*. Cambridge University Press.

Freeman, J. 1972. The tyranny of structurelessness. *Berkeley Journal of Sociology* 151–164.

Hanisch, C. 1969. The personal is political. *Radical Feminism: A Documentary Reader* 113–116.

Harrell, D. F. 2010. Toward a theory of critical computing: The case of social identity representation in digital media applications. *CTheory*.

Harrell, D. F. 2013. *Phantasmal media: an approach to imagination, computation, and expression*. MIT Press.

Hart, V., and Case, N. 2014. Parable of the polygons. *URL: ncase.me/polygons*.

hooks, b. 1992. *Black Looks: Race and Representation*. Women's studies (RUP).: Black studies. South End Press.

Khandaker-Kokoris, M. 2015. Thinking about people: Designing games for social simulation. Game Developers Conference. gdcvault.com/play/1022334/Thinking-About-People-Designing-Games.

Martín-Baró, I.; Aron, A.; and Corne, S. 1994. *Writings for a liberation psychology*. Harvard University Press.

McCoy, J.; Treanor, M.; Samuel, B.; Reed, A. A.; Mateas, M.; and Wardrip-Fruin, N. 2014. Social story worlds with comme il faut. *IEEE Transactions on Computational intelligence and AI in Games* 6(2):97–112.

McIntosh, P. 1988. White privilege: Unpacking the invisible knapsack.

Mitchell, D. O., and Boyd, A. 2013. *Beautiful Trouble: a toolbox for revolution*. OR Books.

Moane, G. 2010. *Gender and Colonialism: A psychological analysis of oppression and liberation*. Springer.

Price, J. M. 2015. *Prison and Social Death*. Rutgers University Press.

Rowland, R., and Klein, R. 1996. Radical feminism: History, politics, action. *Radically speaking: Feminism reclaimed* 9–36.

Samuel, B.; Reed, A.; Maddaloni, P.; Mateas, M.; and Wardrip-Fruin, N. 2015. The ensemble engine: Next-generation social physics. In *Proceedings of the Tenth International Conference on the Foundations of Digital Games (FDG 2015)*, 22–25.

Smucker, J. 2017. *Hegemony How-to: A Roadmap for Radicals*. AK Press.

Starhawk. 2011. *The Empowerment Manual: A Guide for Collaborative Groups*. New Society Publishers.

Wardrip-Fruin, N. 2009. *Expressive Processing: Digital Fictions, Computer Games, and Software Studies*. MIT press.