Generative Allegories of Oppression and Emancipation: Reflecting with Computational Social Models

Ricardo Sosa, PhD

Auckland University of Technology, Auckland 1010 Aotearoa New Zealand

Monash University, Victoria 3145, Australia

Abstract: This paper presents a computational approach to growing *artificial societies* (agent-based simulations) as an explicit, accessible, and systematic tool to visualise and generate insights and new questions about Paulo Freire's concepts of oppression and emancipation. These models do not make claims of validity or prediction, instead their value is to structure our thinking and support our understanding. Here, I use computational social simulations as generative allegories to reflect upon the role of designers in participatory, co-design, and social design contexts. The paper shows how Freirean ideas can help reframe design as a pedagogical craft based on dialogue and collective inquiry.

Keywords: thought experiments, creative research methods, multi-agent simulations, emergence, change agents.

1. Introduction

An intense rain, a campfire, a poem, or a song can be, of course, deeply enjoyable on their own, but at the right moment and with the right mindset, they can also provide *metaphorical* lines of thinking. Some iconic examples of generative metaphors include Plato's Allegory of the Cave, the Garden of Eden, and the Mesoamerican underworld of El Mictlán. They convey important ideas about who we are, how we got here and even how we ought to live. A story resonates with us because we can draw from our lived experience, we understand or can vividly imagine how it *feels* to be in dark caves or in lush gardens. We can also observe raindrops moving on a window or flames dancing and use those observations to build analogies with what we know from previous experience and what we imagine for the future. The shapes of trees, for instance, help us think about genealogy, evolution, and other concepts that *branch out* and *grow* over time.

Models created with computational code can also share this capacity to direct our gaze, thoughts, and sensibilities. Such algorithmic allegories can be valuable to help us visualise, grasp, and ask questions about ideas and principles that our lived experiences teach us. Like poems and raindrops, these models have intrinsic aesthetic value, but they can also help us think better, more clearly. By the time you finish reading this paper, you should be able to judge whether the in-silico models shown here can help you and others think more sharply and ask new questions about Paulo Freire's foundational ideas on oppression and emancipation -and the role of designers in societal change.

These in-silico models do not model *realistic* scenarios, yet they are *real* in that they can help us grasp reality better and thus transform it. Some computational simulations do make predictive claims, however successfully, but not the ones presented here. The best way to think of artificial societies (of computational agents) is akin how we refer to groups of insects as *societies* -a helpful metaphor, yet we instinctively understand that these are very different phenomena. Likewise, *artificial societies* are metaphorical, they aim to illustrate *some* principles and behaviours of interest to help us think and feel about rather complex ideas such as Freirean *oppression* and *emancipation*.

2. Background: PotO models

This section recaps *oppression* and *emancipation* from the *Pedagogy of the Oppressed* (Freire, 2000, potO) (Freire, 2000). Any situation in which a person *objectively exploits* or hinders someone's *pursuit of self-affirmation* is one of oppression. Thus, oppression interferes with people's vocation *to be more fully human,* and oppression acts are those where an elite denies the majority their agency and capacity for self-determination. This *dehumanisation* ends up creating a distortion that affects everyone, including the oppressors. Hence, Freire declares that "the great humanistic and historical task" (Freire, 2000, p. 44) of emancipation is for the oppressed to liberate themselves *as well as their oppressors*. He also sharply observes the "tragic dilemma" (Freire, 2000, p. 48) of oppressors becoming "sub-oppressors" (Freire, 2000, p. 45) by identifying with, admiring, and adhering to the oppressor. In such alienation state, oppression has become *their model of humanity*. In this work we refer to this as **PotO System I**: a model of oppression by an elite who exerts dominance and creates rifts among members of a society.

Freire refers to certain members of the oppressor class who may cease to be *indifferent spectators* or *heirs of exploitation*, yet they can carry in this process "the marks of their origin" (Freire, 2000, p. 60) as long as they continue to lack confidence in the creative power of people, in their abilities "to think, to want, and to know" (Freire, 2000, p. 60). These *converts* may genuinely aspire to transform the unjust order, but often because of their background they continue to believe that *they* must be the "executors of the transformation" (Freire, 2000, p. 60). Even when well-intended, this lack of trust continues to remove the agency of the oppressed in a system where "naming the world is the task of an elite" (Freire, 2000, p. 90). In this model of false emancipation, the oppressors continue to treat people as *things*, while those in power retain the right to be *human*. We define this as **PotO System II**: a model of false revolutions where attempts are made by leaders at liberating the oppressed without their creative and reflective participation.

The journey to Freirean dialogic emancipation starts with a *confrontation* that moves the oppressed to change how they perceive the world. Through a constant "expulsion of myths"[1] (Freire, 2000, p. 55), the structures of oppression are transformed leading to the disappearance of a dominant class. Authentic leadership in this model is based on dialogue, a constant re-examination of ideas, and a *pedagogical* nature of change. Dialogue starts with "the present, existential, concrete situation, reflecting the aspirations of the people" (Freire, 2000, p. 95). The fear of freedom and the lack of trust in others and the self is replaced with autonomy and shared responsibility by roles I paraphrase as the *leading-being-led* and the *led-being-leaders* permanently co-creating and re-creating the world. This becomes **PotO System III** here, a model of permanent liberation imbued with "a profound trust in people and their creative power" (Freire, 2000, p. 75).

Having defined these PotO models, two questions are pertinent: a) What do these Freirean concepts entail for designers?, and b) How may computational simulations assist us to understand, visualise, and reflect upon these concepts?

First the question of what these ideas mean to me as a designer and a human being. My first exposure to PotO was as an undergraduate student, next to readings by Jean Baudrillard, Karl Marx, and Enrique Dussel. Two decades later in 2014 I returned to PotO and was blown away by its currency and applicability to design in at least three levels: first, with the feeling that notions of 'human-centred' and 'users' may carry *the marks of the origin* of design as a corporate profession (Sosa et al., 2021) . Having worked in participatory and community projects in Mexico, South-East Asia, and Aotearoa New Zealand, I have witnessed these *marks* in how designers occupy these spaces and how mainstream design education perpetuates the myth of designers as *change agents* (Sosa,

[1] Freire identifies myths of a free society, free market, personal success, and equality of all individuals.

2022). Second, the so-called creative industries, big tech, and the media profit by defining the world around us, how we communicate with each other, and how we think about others and ourselves. As designers we participate in this system that reduces the majority to passive consumers. Lastly, as an educator and a parent I see how dialogue is constantly dissuaded and prevented in classrooms and homes, and how dialogic pedagogies could transform not only how we learn but how we organise our societies via legitimate authority (Chomsky, 2013).

Freire is explicit about how science and technology are used as "powerful instruments for their purpose: the maintenance of the oppressive order" (Freire, 2000, p. 60). Design has been equally instrumental to persuade people and create an appealing way of living and consuming that has been profoundly exploitative of humans and Nature, threatening the survival of the most vulnerable (Amir, 2004; Martin Juez, 2002). Freire defines *true creativity* in a climate of emancipation as one where everyone cultivates their creative enthusiasm, have hope, trust, and engage in experimentation to determine their futures. I believe that design has some potential to help this happen (Escobar, 2018; Martin Juez, 2002). The work presented here offers a complementary approach: one that generatively uses computational simulations as allegories to scrutinise our thinking and to inspire our collective imagination about the future roles of design for emancipation.

This leads us to the relevance of computer code. The PotO models above are fundamentally systemic: they illustrate a set of individual behaviours, interactions, a context, and feedback causation that yields systems shaped by their aggregate properties: System I is a model of oppression, System II is one of false revolutions, and System III is one of *a constant fight* for dialogic emancipation. The behaviours and interactions in these systems give way to their macro-level outcomes, and causation dynamics are persuasively sketched and exemplified by Freire. The opportunity here is to use code as an alternative and powerful way to describe, analyse, and *play* with these systems beyond words, however persuasive.

"Generative Allegories" are based upon an algorithmic approach to model, grow, and experiment with ideas of oppression and emancipation. This allows to observe their mechanisms and to consider what are the type of factors that can cause transitions between them, and what the role of designers in societal changes may be. They are *generative* in two senses: first, they "use a set of elements and rules for interaction between them to produce a pattern or an outcome" (Costopoulos, 2015) (Freire, 2000, p. 266), in our case patterns of Freirean oppression. Second, they aim to generate new understandings and new questions. I therefore choose a starting point for this work that avoids a research question to be tested via these models, and rather aim to explore if and how this modelling can help us reason about and feel about the possibilities and the risks for designers in social change.

3. A Model of Oppression and Emancipation

Agent-based simulation is a type of computer modelling where large populations of individuals are defined in working code based on theoretical and/or empirical bases and where the intention is to *grow* a macro structure or behaviour of interest in order to inspect, explore, and understand the reference system (Epstein, 1999; Gilbert, 2019). There are several research approaches to simulation for the study of social phenomena (Gilbert, 2005) including of a qualitative nature (Yang & Gilbert, 2008), and to examine questions including those related to innovation (Watts & Gilbert, 2014). Unfortunately, some very interesting models have made claims that are unsupported, such as cellular automata said to capture 'the dissemination of culture' (Axelrod, 1997). These days, a wide range of modelling approaches exist, each with strengths and weaknesses that make them suitable for a variety of inquiry goals. As a researcher, I have been interested in this sort of modelling for twenty years (Gero & Sosa, 2002; Sosa & Gero, 2005, 2008).

Here, I avoid more complex models that aim for empirical validation and veridicality (Carley, 2002), and instead aim for a type of models where agents with simple behaviours can start to reveal how we think and on the implications of what we think about. When it comes to developing agent simulations about ideas like oppression and emancipation, a modelling approach could aim to include a large number of evidence-based data gathered through census, survey, and interview data. Such models would seek explanatory and even predictive powers to assess the likelihood of a social movement in specific countries at a specific period. *"What is likely to happen in Afghanistan/Ukraine when the US/Russia invasion ends?"* and *"What may the anti-mandate and anti-vax convoys achieve in Ottawa and Wellington?"* are the type of interesting questions that such models can address. But that is *not* the type of models discussed here.

Instead, I am aiming to model more generally the *possible* systems where phenomena related to the notion of oppression can be observed. These models follow an *abstract-generalist* form of simulation (Costopoulos, 2015) and their value comes down to how they act as *intuition pumps*, or tools-to-think-with that allow us to show "reliable and convincing" (Dennett, 2013, p. 197) intuitions about complex ideas, or alternatively to "focus attention on what is wrong with [our] presuppositions" (Dennett, 2013, p. 197). Unlike rhetorical devices, computer models have important features defined in Table 1 including explicitness, accessibility, stochasticity, emergence, analogical thinking, experimentation, and maximum parsimony.

Table 1. Properties of Computational Social Simulations

Explicitness	Their definition requires clear, explicit, comprehensive, and full specification into code that compiles
Accessibility	Results are visible including those that are measurable. At every step of the simulation we have access at the mechanisms at play being possible to reconstruct every scenario in detail
Stochasticity	Randomness is used to model external factors, but this needs to be fully and explicitly justified. The models are run a large number of instances (cases) to characterise them. A (Freire, 2000, pseudo)random number generator is used to inspect causality
Emergence	The system behaviour cannot be determined from complete knowledge of its rules and initial state. Emergence has multiple sources including bottom-up, downward, and second-order (Gilbert, 2002). We may be able to define and understand the system rules, but their interaction across levels of agency and over long periods make it extremely hard to grasp unaided by computers
Analogical thinking	Many algorithmic ways are possible to grow a macro or social behaviour of interest, therefore no claims of validity are relevant between "the map and the territory" here. It can be tempting to exaggerate the connection between these artificial societies and human societies, but a tempered view is required to know where these systems assist and when they obstruct our thinking
Experimentation	These models work rather as the thought experiments used in theoretical fields except that these models are indeed experimental. They go beyond stories like Schrödinger's cat in a box, a falling elevator, Flatland, or Zeno's paradoxes because they help translate "What if…?" questions into code, run it over simulated time, record the results, and formulate explanations of how the system behaves, and repeat
Maximum parsimony	For every model element, role, and behaviour we adopt the simplest alternative possible and follow an iterative bottom-up course where additions are made only when we can fully account for how the model works, and why

While basic programming skills are necessary to run and extend these models, they offer advantages over more formal, mathematical models. First, they are written in *high-level* code, the more English language-like code that non-technical people can learn. They are built using a platform for artists and designers, *processing.org*. Second, the models prioritise visual thinking, as they present on screen a graphical representation of every interaction at every step. Third, they allow for individual agent variation including from feedback of aggregate effects, something not amenable to mathematical formulae. These properties make these models 'low-floor, wide-wall, high-ceiling' (Resnick & Silverman, 2005): low-floor as they provide easy access for novices to get started, wide-wall in that any number of possible changes are possible since Java is an open access language with many libraries available, and high-ceiling in that the models have no defined limit to where they can be extended and modified. Now: how may we start accounting for Freirean oppression with this modelling approach?

3.1. System I: Clique formation controlled by elites

Modelling requires an iteration of *art* and *science*. I started by sketching initial ways to represent Freirean oppression in screen-based agent worlds. I considered a *SugarScape*-type model (Macal, 2020) but opted out of agent societies that maximise material resources in a competitive zero-sum game. I was aiming for a less prosaic metaphor and found one to represent the *dehumanising* of agents inspired by Reynolds' *boids*: algorithms that simulate the mesmerising navigation patterns of animal flocks/herds by applying three simple individual rules based on nearby individuals: align with neighbours, stay close to neighbours, and avoid crowding (Reynolds, 1987).

This approach looks at the interplay between individual and group agency where *movement* is determined via a process of decentralised and local coordination. These highly choreographed synchronic flocking patterns serve as a metaphor for oppression when agency is defined as the capacity to freely move but some agents take away this freedom and decide for others. I imagined such an elite *taking over* social groups and hindering the majority's pursuit of free movement, deciding for them, and controlling, quite literally, their *destiny* (destination). This is the PotO System I defined earlier and became the baseline here implemented as follows:

- An array stores all agents in a society (case). Each agent is initialised with:
    - a Gaussian-distributed *agency value* (mean 0.0 standard deviation 1.0),
    - an initial random *location* on the grid (uniform distribution),
    - an empty set of *neighbours*,
    - a random *colour* for display (uniform distribution),
    - an undefined (null) *preference* for a grid region,
    - a label to identify *elite* agents, and
    - an *influence* index (set to 0) to keep track of its influence over others
- At the societal level, a two-dimensional space (a grid) is created and populated with a ratio of agents, typically 10-30% so they can move during a simulation. Four grid regions are defined as North, East, South, West (N/E/S/W). Constant values are defined for the size of agents and the size of their *neighbourhood area* (the adjacent space to consider other agents as neighbours). A limit number of iterations is set for cases to run, and a modulo value is set to periodically record information from all agents, for analysis.
- In System I, agents with the highest *agency value* are selected at initialisation as *elite* agents. Up to three elite agents are chosen at random from this set of candidates. Each of the elite agent's preference for a grid region is set.

Figure 1 shows five screens of four cases from steps 0 to 10^5, but to get a feeling for the system, the reader is encouraged to watch the supplementary video "PotO System I"[2] running for 100,000

[2] PotO System I (01:41): https://youtu.be/roycAPHHEaI

iterations. Four cases are shown to reinforce the idea that our interest is in how a type of models work rather than in the peculiarities of any single case. Simulation data are recorded from 100 cases with unique random seeds running for 100,000 steps to characterise the model applying numerical analyses (beyond the scope here). What follows is a descriptive account of System I:

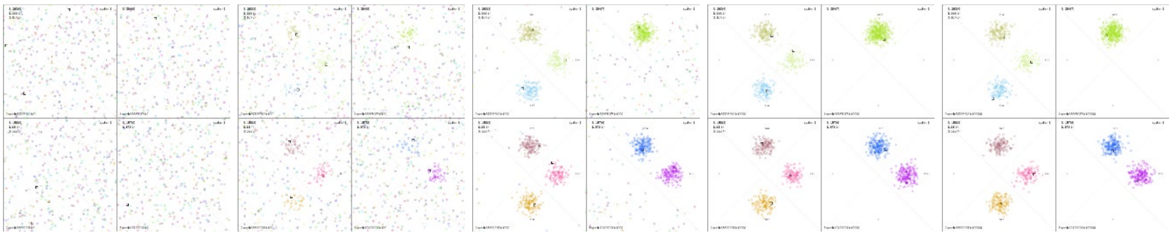

Figure 1. PotO System I: Four cases running over 10^5 steps

The simulation starts with agents of random colours positioned at random grid locations. Agents are displayed by circles and *elite* agents by squares. At every iteration step, all agents are activated: first, if an agent has no preferred grid region, it moves at random (roams). As it moves through the grid, an agent checks around and identifies its adjacent neighbours. When an agent has a neighbour and that neighbour has a preference for a grid region and that preference is different from its own, an agent has a 50% chance to *recruit* that neighbour, i.e., change their preferred region (and colour) to their own (Trigg et al., 2008, p. 55). When this happens, the agent gets an increment of one *influence* point. When an agent has a preferred grid region, it moves in that direction with a probability that is proportional to how far it is from said region (the furthest away the more likely it will tend to approach).

As the simulation continues, more agents start to gravitate towards one of the grid regions (N/E/S/W) controlled by an elite agent (cases can have from one to four elite-controlled regions). As time progresses, cohesive groups emerge of the same colour. This is seen in Figure 2 in four cases: at step 10^3 the proto-cliques have formed and by 10^5 elites have captured every agent into one of its cliques. Sooner or later very few or no agents remain free from the control of an elite. This pattern resembles the *type of* systems where a few commanding individuals in authority shepherd others around and the population looks almost like 'zombies' devoid of a voice. This use of movement is a simple yet illustrative metaphor for the gist of oppression: if agents were free, they would be able to move where they chose or would have a say as members of a group. But agents in System I are in a situation of oppression and move to where they are told, and once cliques form, entropy is reduced to the point of a "death-affirming climate of oppression" (Freire, 2000, p. 68).

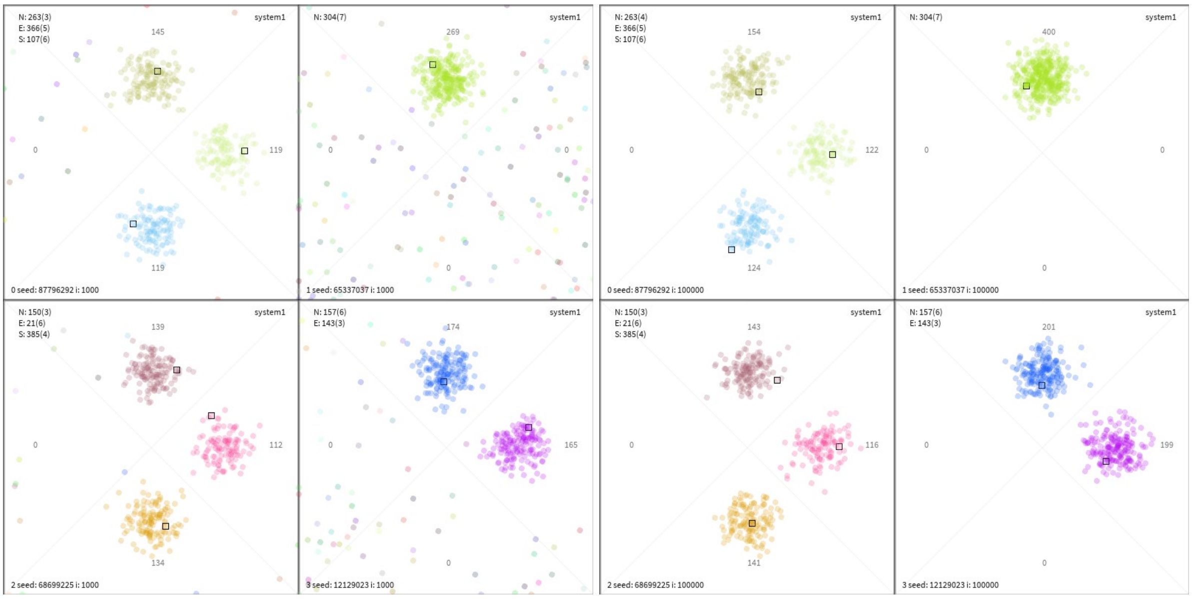

Figure 2. PotO System I: Clique formation at steps 10^3 and 10^5 in one, two, and three grid regions

This description is succinct here due to the word limit, but the modelling process is highly creative, iterative, and generative. The rationale behind some of these variables, parameters, and algorithms comes from trying alternatives, running the simulation over a few or many cases to notice patterns (visually and numerically). This process that takes weeks and allows the modeller to develop a *feeling* for what the agent simulation is capturing vis-à-vis the appropriate and sufficient baseline dynamics that reflect the concepts of interest while striving for extreme parsimony. The models are therefore *debugged* in two senses: they need to be *compiling* and *compelling*. The code is publicly available to assess the former; the editors, reviewers, and readers of this journal will help evaluate the latter.

At first, elite agents are fully responsible for recruiting agents under their control, but soon after they start to benefit from a growing number of non-elite agents who exert influence over their neighbours and thus amplify their power. This alludes to, albeit crudely, Freire's concept of sub-oppressors, alienated agents who adhere to and help maintain the status quo of oppressive elites.

Are these the *right* variables and parameters to model System I? There clearly are other modelling options, but two things are important: first, most of these mechanisms are not particularly determinant of the outcomes: the randomness in the recruitment algorithm can be removed, the neighbourhood size value or the density of agents on the grid can be changed, and the shape of the random distribution can change without affecting *the type* of outcome observed here, at most they change the speed and the sharpness of the results. System I has low sensitivity to many of these parameters and conditions making our claims illustrative, not predictive (Bertolotti et al., 2020).

When watching the video "PotO System I" it can be tempting to *anthropomorphise* these agent societies. Observing how cliques form and how they seem to "fight" other cliques to take over followers can be seen as intentional, and one can even identify "strategies" that the elites put in action with their associates. When an agent triggers a group-level transformation, it could be tempting to refer to them as "change agents". However, we must remember that these agents have no concept of groups much less of *social change*, they are simply following the same set of deterministic rules and responding to their environment in a non-changing way from the initial to the final step. It is the *aggregate* effect of their localised interactions which generates the macro structures and events that we (but not them) observe. Still, for such basic agent societies it is remarkable how they can enable observers to reason about authoritarian human groups which function by depriving individuals of autonomy and are headed by a minority who decides for them. Over time, feedback mechanisms cause these elite agents to accumulate more influence and to further extend their control over others. Yet, these dynamics can also account for the demise of old and the appearance of new cliques.

3.2. System II: Elite changes (Freire, 2000, plus ça change...)

As the implications of System I's variables, parameters, and outputs became clearer, I continued to sketch ideas for how to account for System II. Watching closely the video "PotO System I" reveals a very low but interesting possibility for "phase transitions" in some of the cases. A phase transition can happen once cliques of same-colour agents have formed and:

- a non-elite agent comes across a member of a different clique (unlikely)
- one of said agents influences the other and changes its preferences (fairly likely)
- this creates a "snowball effect" by which an increasing number of agents switch preferred grid regions causing the clique to disband (extremely unlikely)

To include the possibility of changes in elites and cliques in PotO System II, I add a mechanism by which non-elite agents can *overthrow* existing elite agents or they can *become* elite agents in a region that previously had none. The following algorithms are added to System I to build System II:

- a society-level measure of *inequality* is updated at every iteration step calculated by the sizes of cliques as a proportion of the total population
- at every iteration step, after a non-elite agent moves, if *inequality* is higher than a threshold value defined for a case (I use 50% but again, it could be 20% or 80% the only difference is the likelihood of these events), then:
    - if the region is ruled by an elite agent, and
        - if the non-elite agent's *influence* is equal or higher than the elite agent's *influence*, then replace it, half its *influence*, and record a 'revolution' event
        - else if the non-elite agent's *influence* is lower than the elite agent's *influence*, then leave things as are and record a 'failed coup' event
    - else if no elite agent controls this region, claim the elite status and record a 'new elite' event
- keep a tally of all events registering the timestamp, the agents involved and their *influence* indices, with the sizes of cliques and the *inequality* level at the time.

Figure 2 shows three screens of four simulation cases from steps 10^2, 10^4, and 10^6; the video "PotO System II"[3] gives a better idea of what happens over one million steps. At first sight, cliques seem to form and behave the same way as in System I, however, changes do take place over time among the elites: new elite agents emerge either by claiming a grid region or by replacing old elites. To some extent, these account for transformational societal changes but in ways that leave the fundamental structure of oppression unchanged. These approach the idea of *pseudo*-revolutions described by Freire (2000, p. 69) led by agents who oppress while trying to liberate others. Non-elite agents do have opportunities to become elite based on the influence they accrue at the "grassroots level", thus changing the way elite status is achieved. In System I elite agents are defined simply based on their programmed traits, whilst in System II non-elite agents can achieve elite status by "merit". Still, for the majority of agents in the population this is immaterial: the new master could originate from their

[3] PotO System II (16:41): https://youtu.be/8dDtROEKTqY

ranks (and be well or ill-intentioned) but the structural hierarchies of oppression from System I remain.

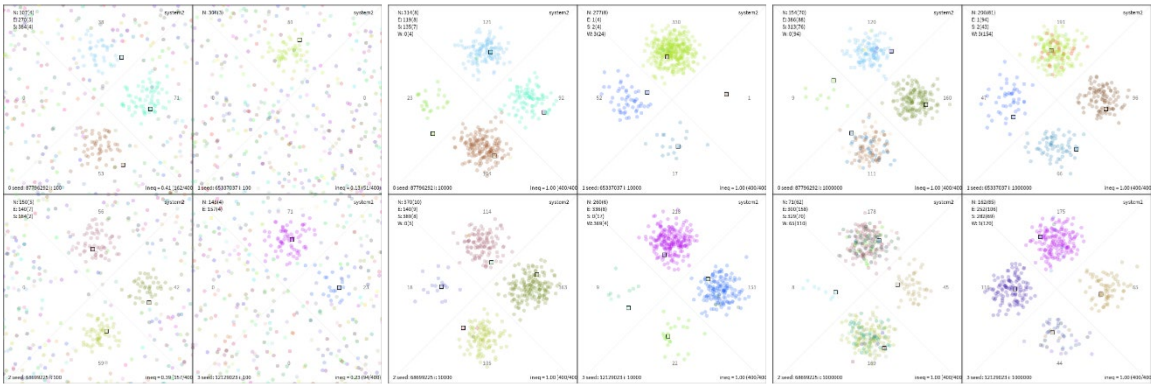

Figure 3. PotO System II: Four cases running over 10^6 steps

System II highlights a fundamental property of these societies, a dual system-level mechanism: the formation of cohesive groups needs to enable sporadic contact across groups to allow for the change cycles observed. The features that set this balance include the number and size of agents in the grid, the size of neighbourhoods, and the step size that sets agent mobility. Varying these parameters leads to the growth of tighter more long-lasting cliques, or looser and short-lived ones. I use mid-range values to help visualise these dynamics without an interest to capture any *realistic* conditions. In other words, whether we think of millennia-lasting pharaonic eras or months-long cults, these models allow us to experiment with and think about the type of factors at work. Figure 4 shows how cliques that are well-formed by step 10^5 can sometimes change quite significantly by step 10^6: in every case at least one elite agent has been replaced and the effects are noticeable by the size and colour of cliques.

Nonetheless, agents in System II situations not only continue to be *alienated*, perhaps worse, now a deceiving rhetoric can be formed around the idea that *anyone* can “make it”: stories about the new elites can be presented as evidence of the potential and the legitimacy of personal success. Never mind that the “self-made” elite agents in System II are primarily a product of their circumstances, they have no extraordinary innate traits, and their likelihood of succeeding is akin to winning the lottery. It is remarkable that such simple models can illustrate rich ideas such as Freire’s model of humanity where the oppressed aspire to follow, imitate, and resemble the elite to the point where “success” is defined as occupying a position of oppressor. In System II who is an oppressor can change, but the situations of oppression go unchanged. Agents follow the same algorithm, after all. How may System II need to change to account for the labour of emancipation?

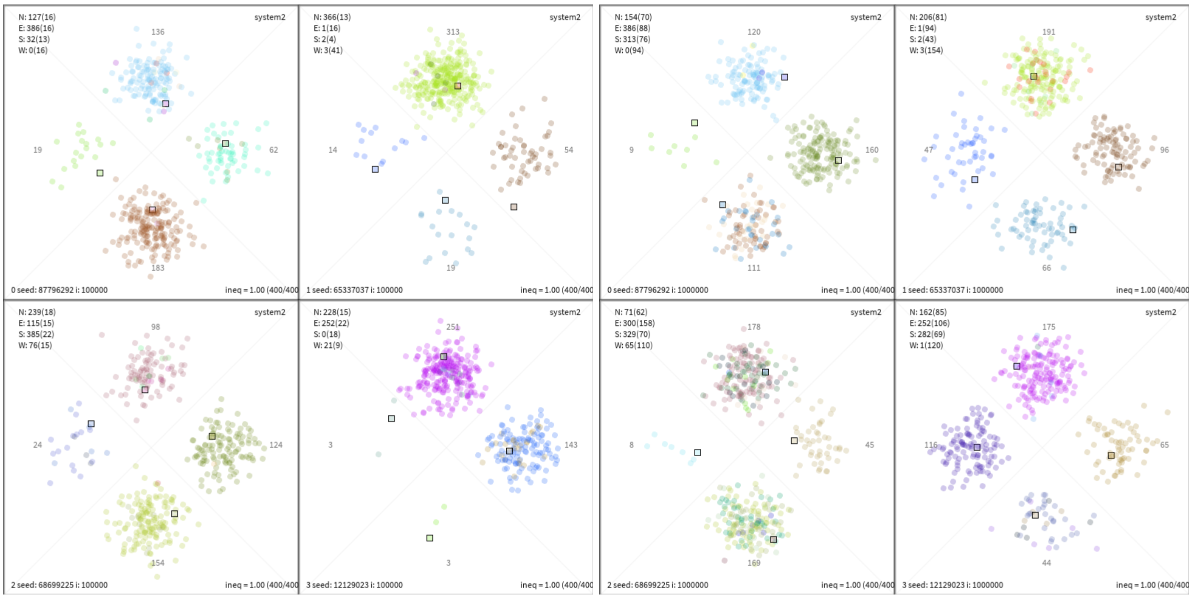

Figure 4. PotO System II: Elite changes in four cases at steps 10^5 and 10^6

3.3. System III: A permanent fight for emancipation

The goal in implementing System III is to illustrate some of the aspects described by Freire (2000, p. 80) as characteristic of emancipation. These include the *constant* expulsion of myths via an authentic dialogue of a pedagogical nature by which individuals work together to change the model. Such dialogue starts with the present situation and the aspirations of the people, who trust others and in partnership achieve autonomy. Notably, Freire describes this as a model of *permanent* liberation rather than an end state, and one where emancipation works for all including the oppressed and the oppressors. To implement System III, I introduce a new type of agent: *t-agents* dedicated to organise others based on their individual and collective preferences. The system builds upon Systems I and II, introducing the following additional features:

- at setup, one to four random non-elite agents are designated as *t-agents*
- all agents are now initialised with a working and a long-term preference for a grid region (N/E/S/W). The working value is susceptible to be influenced by others, primarily elite agents, while the long-term value remains unchanged
- *t-agents* roam the grid coordinating neighbours they encounter by asking for their long-term preferred grid regions. When neighbours prefer adjacent regions, they produce an interpersonal shared preference, which can lead to the formation of *t-cliques* in four new regions: North-East, South-East, South-West, and North-West.
- meanwhile, elite agents continue recruiting followers as in Systems I and II, but their powers are constrained in three ways: a) only elite agents can recruit agents while non-elite agents no longer engage in recruiting, b) to recruit the influence index of an elite agent needs to be greater than the neighbour's, and c) *t-agents* cannot be recruited by elites.
- the way the *influence* index is calculated changes: for elite agents it continues to be based on their recruitment of followers as in System II, but for all other agents it stops being an individual feature and becomes a group-based metric. At every step, the members of a *t-clique* all increase their influence as a function of the size of their *t-clique*.

Figure 5 shows three screens of four simulation cases from steps 10^2, 10^4, and 10^6; the supplementary video "PotO System III"[4] shows these cases running over one million steps. System III allows for the formation of elite-grown cliques in N/E/S/W regions as in System II, but now a different type of process leads to the formation of *t-cliques* in NE/SE/SW/NW regions. While cliques in Systems I and II are the (in)direct product of elite agents, *t-cliques* in System III are of a different nature: they are co-created by agents who combine their individual preferences, i.e., everyone has a say in their formation. Thus, *t-agents* enable the organisation of agents at the local level to form these bottom-up *t-cliques*.

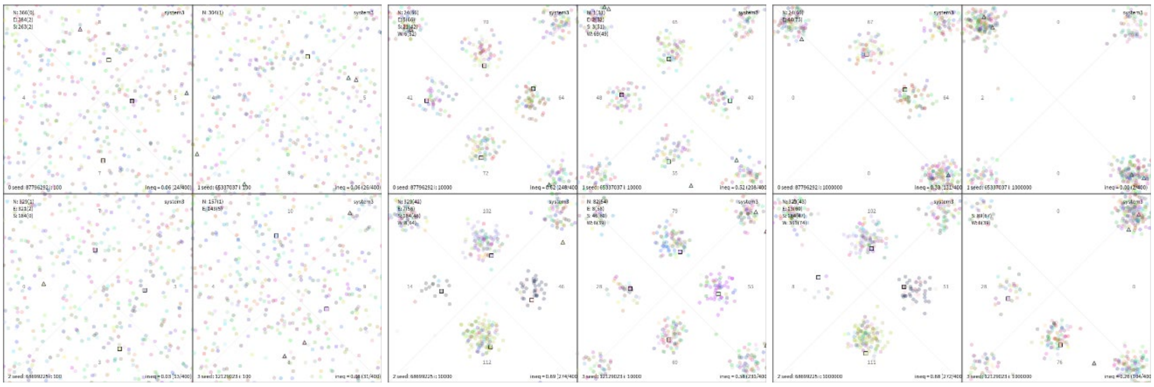

Figure 5. PotO System III: Four cases running over 10^6 steps

In System III elites can be replaced and new elites can emerge as in System II, but elites can now also disappear and their groups may even disband altogether. This happens when *t-agents* manage to form large enough *t-cliques* and encounter an elite agent that is compatible and joins the *t-clique*, losing the elite status. This is unlikely and only happens occasionally and only in some cases, while in other cases elites continue to dominate and control large groups thus avoiding being integrated by a *t-clique*'s collective influence. And even when an elite disappears, out of the group in that region a new elite agent often emerges to occupy the position of control over others.

Figure 6 shows a case (top-right) that by step 10^6 has had all its elites extinct and all cliques practically disbanded (only one with two non-elite agents survives in the West region). Nearly all agents now are part of *t-cliques* with no elite in control. Agents in *t-cliques belong* there because they have negotiated with similar and compatible agent groups building shared mixed preferences. In System III thus, agents in situations of oppression can not only break free from an oppressor elite, but they can change the model thus liberating the elites as well.

In most System III cases, however, elite-grown cliques (N/E/S/W) co-exist with bottom-up *t-cliques* (NE/SE/SW/NW) over long periods. These cases show a type of *dynamic equilibrium* that alludes to Freire's characterisation of emancipation as a permanent journey (Freire, 2000, process of becoming) rather than a destination (a state of being). Most of these agent societies can continue in this flux forever, so they cannot be described as either oppressed or liberated, rather they are constituted by situations of oppression and situations of emancipation that vary in strength over time. This dynamic balance means that some *epochs* can be characterised by oppression and in others emancipation is stronger. And, given the right/wrong circumstances, a seemingly emancipated population is likely to regress into a highly oppressive system and only occasionally eradicate the elites. Until, of course, externalities to this model intervene to impose a new elite.

[4] PotO System III (16:41): https://youtu.be/TGWFNuHwZOw

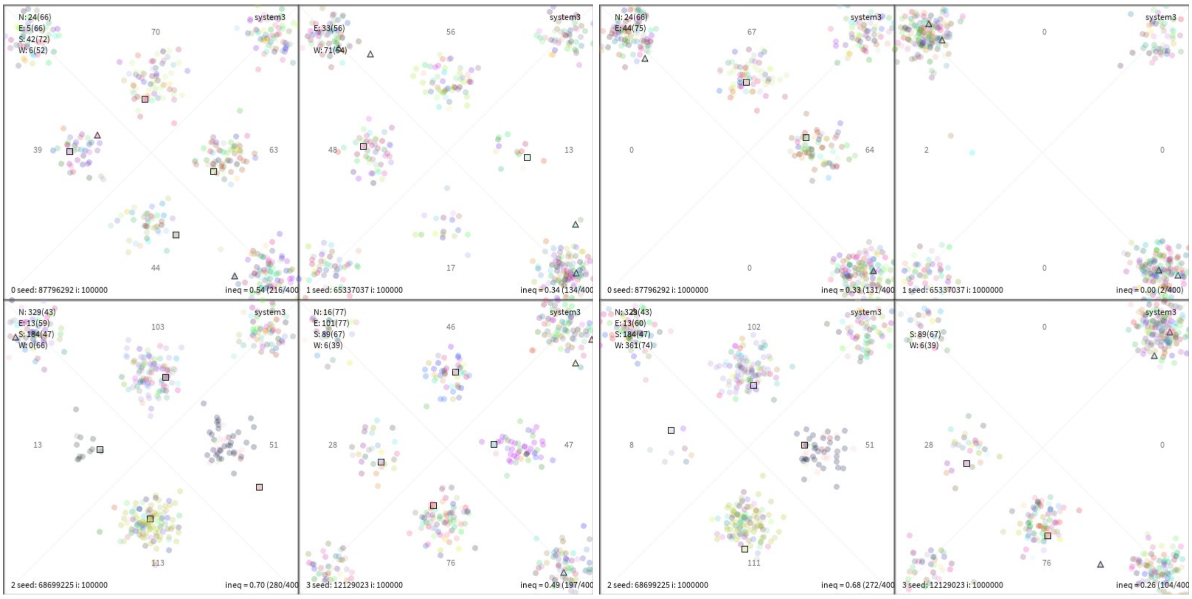

Figure 6. PotO System III: Cliques, t-cliques, and disappearance of elites at steps 10^5 and 10^6

An interesting feature of System III is the role *grassroots* organisation. The third type of agents introduced (*t-agents*) do not seek to control others and are not susceptible to be controlled. Instead, they organise agents at the local level, one interaction at a time. They do not seek to replace elites by exploiting the same mechanisms of oppression (resembling what Freire calls *libertarian propaganda*), but rather help agents organise into groups where each individual preference counts. The result is the decentralised and aggregate formation of groups defined by characteristics that no single individual has, i.e., those with North preferences and those with East preferences coalesce in groups at the North-East region. And since no single agent has this preference individually, these groups are co-defined and not susceptible to be controlled by any single agent. To a limited extent, this begins to illustrate Freirean principles such as trusting the agency of the people and the pedagogical, co-intentional character of liberation.

4. Conclusions

The goal of the simulations presented here is to visualise and generate insights and questions related to Freirean oppression and emancipation. The models are not presented as *evidence* of anything, they do not resolve a hypothesis. In this sense, they are assessed akin philosophical arguments or as creative artefacts in artistic practice-led methodologies. They are not assessed for validity or predictive accuracy; instead, their evaluation looks first at their internal coherence and correctness which is carried out by analysing and running the source code. Second, by their usefulness to refer to theoretical principles and inspire new insights, questions, and ideas about said theories. I close the paper focusing on what *t-agents* in these societies suggest about the work of designers working in participatory, co-design, and social design contexts.

How can designers act as *t-agents* of their societies? Our work shows that such organising role at the local level *can* cause aggregate effects at the system level and may even lead to structural changes in the model. One would be thus tempted to call *t-agents* “change agents”, but these models show that while their role is vital, they do not control the direction of change in any meaningful way. Their role is to elicit the agency of others. It would therefore be more accurate to call *every* agent in the population a *change agent* rather than those facilitating the process. Moreover, depending on the conditions, these change initiatives may have rather limited effects in time and space. This is an important lesson inferred from these artificial societies: it is *possible* that behind success and failure of systemic transformations lies the exact same mechanisms and the difference is due to situational factors and cumulative effects such as the metaphorical snowball, cascade, and butterfly effects. To reiterate: I draw inferences of *what is possible* (not likely) and these models aid our imaginations on *how* these processes may play out.

These simulations do strengthen Freire's warning that organisers "run the risk of falling into a type of generosity as malefic as that of the oppressors" (Freire, 2000, p. 60). These models illustrate how *herding* people can be easier and more efficient, thus it can be tempting for those who think that they know what is good for the people, to resort to "methods of dehumanization" *for their own good* (Freire, 2000, p. 67). My years working with, educating, and studying designers tell me that we do often resort to toolkits and methods that carry the "prejudices and deformations" of the corporate origins of design where designers are hired by clients to figure out how to sell new products to consumers. And then designers become celebrities or gain the admiration of their peers based on how well their products sell. The agent models shown here help expel these *myths* of design: the "creative class" and the status of "celebrities" (mostly white male) who "change the world" with their visionary talent.

At the moment, my *thinking-feeling* informed by countless hours coding, sketching, talking about, writing, and even dreaming of these agent societies is that designers would greatly benefit from an understanding of our profession as one of a *pedagogical* nature. I used the name *t-agents* to illustrate their (admittedly basic) teaching role working alongside others in dialogic ways to organise re-invention. The third model (System III) can illustrate how powerful this role can be to challenge the oppression by hegemonic elites. I suspect that our teachers at UAM included Pedagogy of the Oppressed with a similar purpose to bring Freirean ideas of dialogic pedagogies, generative and hinged themes, and conscientização into design. But in the 54 years since its first edition, these efforts are still limited (Kina & Gonçalves, 2018). For example, how do the curricula of design schools prepare young professionals to view their craft not as one of personal creativity but as one of bringing out the agency of others having a profound trust in *their* creative powers?

Freire refers to "co-intentional education" (Freire, 2000, p. 69) in ways that suggest how neither the people nor designers need be deprived of their agency: in a committed involvement, both can become permanent re-creators of knowledge and their worlds. When every voice counts, and diverse ideas are valid and welcome, there can be certain *contradictions* between organisers and the people, and of course, among the people too. A practice of design as *problem-posing* bases itself on creativity and stimulates true reflective action. Only such approach to design can reclaim and retain the "primordial right" (Freire, 2000, p. 88) of everyone to speak their word.

Limitations of our models include the minimalist behaviour of agents which can and should be extended. For example, to represent ways in which everyday situations can be coded by some agents and decoded in multiple ways by others depending on their own experience. Future work will aim to include the principle of "thematic fans": codifications that are not overly explicit nor overly enigmatic and allow for multiple decodifications, just as one would expect readers will reinterpret and extend the computational allegories presented here.

References

Amir, S. (2004). Rethinking design policy in the third world. *Design Issues, 20*(4), 68-75.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution, 41*(2), 203-226.

Bertolotti, F., Locoro, A., & Mari, L. (2020, 2020//). Sensitivity to Initial Conditions in Agent-Based Models. Multi-Agent Systems and Agreement Technologies, Cham.

Carley, K. M. (2002). Simulating society: The tension between transparency and veridicality. Proceedings of Agents,

Chomsky, N. (2013). *On Anarchism*. The New Press.

Costopoulos, A. (2015). How did sugarscape become a whole society model? In *Agent-based modeling and simulation in archaeology* (pp. 259-269). Springer.

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. WW Norton & Company.

Epstein, J. (1999). Agent-Based Computational Models and Generative Social Science. *Complexity, 4*(5), 41-60.

Escobar, A. (2018). *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press. https://www.google.com/books?id=M0NRDwAAQBAJ

Freire, P. (2000). *Pedagogy of the Oppressed*. Continuum.

Gero, J. S., & Sosa, R. (2002). *Creative design situations: Artificial creativity in communities of design agents* Computer-Aided Architecture Design Research in Asia CAADRIA, Malaysia.

Gilbert, G. N. (2005). *Simulation for the social scientist* (2nd ed. ed.). Open University Press.

Gilbert, N. (2002). *Varieties of emergence* Agent 2002 Conference: Social agents: ecology, exchange, and evolution, Chicago, Il.

Gilbert, N. (2019). *Agent-based models* (Vol. 153). Sage Publications.

Kina, V. J., & Gonçalves, A. (2018). The continuing relevance of Paulo Freire: Celebrating 50 years since the publication of Pedagogy of the Oppressed. *Critical and Radical Social Work, 6*(3), 363-376.

Macal, C. M. (2020). Agent-Based Modeling and Artificial Life. In M. Sotomayor, D. Pérez-Castrillo, & F. Castiglione (Eds.), *Complex Social and Behavioral Systems : Game Theory and Agent-Based Models* (pp. 725-745). Springer US. https://doi.org/10.1007/978-1-0716-0368-0_7

Martin Juez, F. (2002). *Contribuciones para una antropología del diseño*. GEDISA.

Resnick, M., & Silverman, B. (2005). Some reflections on designing construction kits for kids. Proceedings of the 2005 Conference on Interaction Design and Children,

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. Proceedings of the 14th annual conference on Computer graphics and interactive techniques,

Sosa, R. (2022). I am a Creative Loop: Towards Integrative Studios in Design and Creative Technologies. *Revista GEMInIS, 13*(3), 71-81.

Sosa, R., & Gero, J. S. (2005). A computational study of creativity in design: The role of society. *AI EDAM, 19*(4), 229-244. https://doi.org/https://doi.org/10.1017/S089006040505016X

Sosa, R., & Gero, J. S. (2008, 8//). Social structures that promote change in a complex world: The complementary roles of strangers and acquaintances in innovation. *Futures, 40*(6), 577-585. https://doi.org/https://doi.org/10.1016/j.futures.2007.11.006

Sosa, R., Mejia, G. M., & Adamson, J. (2021). Fluid worldviews: Designing within the common good. Design as Common Good/Framing Design through Pluralism and Social Values, Switzerland.

Trigg, A. B., Bertie, A. J., & Himmelweit, S. F. (2008). *Modelling Bourdieu: An extension of the Axelrod cultural diffusion model* (Open Discussion Papers, Issue.

Watts, C., & Gilbert, N. (2014). *Simulating innovation: Computer-based tools for rethinking innovation*. Edward Elgar Publishing.

Yang, L., & Gilbert, N. (2008). Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in complex systems, 11*(02), 175-185.