

# Disrupting the Coming Robot Stampedes: Designing Resilient Information Ecologies

Philip Feldman<sup>1,2</sup>, Aaron Dant<sup>1</sup>, and Wayne Lutters<sup>2</sup>

<sup>1</sup> ASRC Federal, Columbia Maryland, USA

<sup>2</sup> University of Maryland, Baltimore County, Maryland, USA

**Abstract.** Machines are designed to communicate widely and efficiently. Humans, less so. We evolved social structures that function best as small subgroups interacting within larger populations [2] [15]. Technology changes this dynamic, by allowing all individuals to be connected at the speed of light. A dense, tightly connected population can behave like a single agent. In animals, this happens in constrained areas where stampedes can easily form. Machines do not need these kinds of conditions. The very techniques used to design best-of-breed solutions may increase the risk of dangerous mass behaviors among homogeneous machines. In this paper we argue that ecologically-based design principles such as the presence of diversity are a broadly effective strategy to defend against unintended consequences at scale. [12]

## 1 The bonfire of the vanity plates

The Southern California summer of 2024 is once again a scorcher. The record temperatures are mitigated by the cool air conditioning of the autonomous vehicles packing the freeways. They move fluidly, as traffic jams are generally a thing of the past. Sophisticated routing algorithms keep traffic flowing smoothly around any problems. On their designated lanes vehicles travel in tight, aerodynamic packs at inhuman speeds. Commute times that used to take hours now take minutes. The process is so efficient and trusted that passengers rarely even glance up from their screens.

A wildfire is rapidly descending on Los Angeles. Blown by gale-force winds, it covers 10 miles every 30 minutes, endangering the 405 near Bel-Air. Police shut down the highway, and the traffic quickly routes around the obstruction. Many of the alternate routes are also blocked, but the A\* algorithm [8] is relentless in finding new routes. A solution is quickly found. Unfortunately, it is through burning neighborhoods not yet identified on the satellite map updates. Thousands of cars optimally converge on a two lane road, twisting along a ridge into the fire.

Although these cars are state-of-the-art and handle many circumstances flawlessly, they have not been trained to recognize sheets of fire blowing across the roads, generating temperatures that melt metal. To their sensors, the way ahead is clear.

Thousands of guidance systems agree that this is the best route.

Thousands of cars charge into the flames.

As they enter the fire, antennas burn off, sensors malfunction and vehicles disappear from the network. Like stampeding buffalo, they swerve off the road and into the ravine. This happens *fast*, with between 3-4 cars leaving the road *per second* [1].

For the remaining cars fleeing the closed freeway, the way still appears open. Cars continue to flood through the gap until the pile of burning wrecks is large enough that the system finally recognizes the obstruction and searches for a new alternative, thankfully away from the inferno.

It takes weeks to figure out what happened and issue a patch to the millions of identical vehicles. They are back on the roads but just as vulnerable to the next unforeseen problem. Is there anything that can be done to change this calculus so that every potential does not have to be predicted?

## 2 Understanding ecologies of artificial systems

Best design practices are about building for scale. These systems can be thought of as collections of mass-produced items operating as individual components or combined into a predictably functioning whole. But when the whole is an ad-hoc network of millions of independent components, these practices will no longer be sufficient to deal with emergent and unpredictable problems. For answers, we can look to ecosystems, where emergence and surprise are the bedrock of resilience [9].

Our work focuses on the behavior of individual agents and groups in “belief spaces”, the subset of information space that is associated with opinion (e.g., design, politics, & pertinence). Our simulations explore mechanisms for how consensus develops, starting with high-dimensional concepts that are then simplified to produce alignment within groups [4] [5] [3]. The type of interaction that emerges is a function of agent similarity combined with connection density and “stiffness”. Think of a noodle. Dry, it behaves like a single entity because every element is stiffly connected to every other. Cooked, the connections relax, and it is possible to move one end of the noodle without affecting the other. Applied to communication networks, it is easy to see that a densely connected network of identical agents communicating at lightspeed will behave very differently from a pre-industrial society communicating through letters and personal contact. The former has the capability and predisposition to behave as a monolithic entity, while the latter will have to adopt very different strategies to achieve their goals [11].

The combination of lower-dimension belief space containing a closed, well connected network of similar individuals creates the conditions for stampede. The agents begin to behave as a block instead of a group. Social pressure is powerful, as group consensus can outweigh environmental factors [7]. In people, this manifests as runaway echo chambers. In machines, this can look like a stock market “flash crash”.

These network properties become a set of “knobs” that can be used to affect the formation of group behavior. Our research indicates that a small number of “diverse” agents, injected into the awareness of a stampeding population can have beneficial effects. They enlarge the belief space so that the stampeding agents can be nudged away from their current trajectory. They slow the stampede, since more computation is required to evaluate the belief environment. They can peel away less rigid members, reducing the impacts when the stampede hits an environmental limit.

Unfortunately, our simulations show that turning these knobs in the other direction can *create stampedes*. It has been shown that social media can be used to increase polarization along *emergent* poles [6]. Examples have been documented of troll farms

simultaneously organizing opposing groups [17]. In these cases, bad actors are changing the underlying network structure to more densely connect like-minded people and push them further in a direction that they are already headed.

Diversity injection into social media feeds has the potential to disrupt artificial stampede conditions. In human systems this could take the form of adding random, interesting, known good information to the top of a news feed. For intelligent machines, using legal mechanisms similar to antitrust would ensure a minimum number of independent systems are incorporated in large scale applications like transportation. By enforcing diversity, we begin to gain the resilience to unpredictable threats as found in ecosystems.

### 3 Evolved vs. designed systems

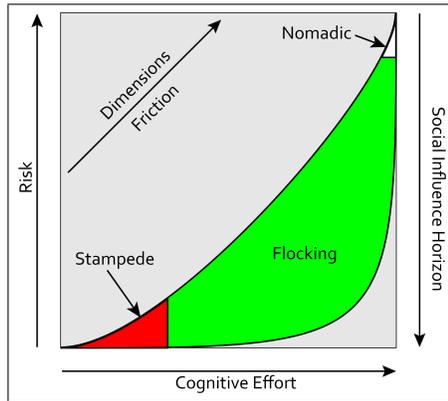


Fig. 1. Evolved systems

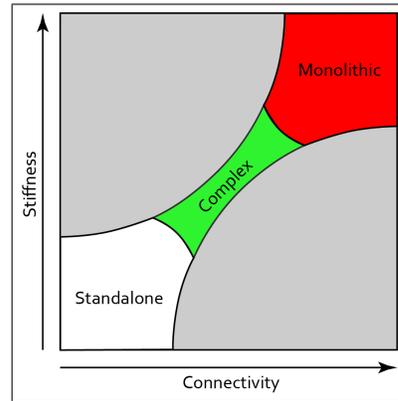


Fig. 2. Designed systems

Our work indicates that agents interact in one of three patterns (figure 1). These phases emerge in evolved systems, and they have an equivalent in designed systems. These phases can be induced by changing the social influence horizon (SIH) of the agents, ranging from “stiff” to “loose”. These phases are:

*Nomadic / Standalone* - Each agent is autonomous, and navigates belief space uninfluenced by other agents. Nomadic behaviors are risky for the individual but advantageous for the population, as they explore broader, possibly dangerous spaces. If they encounter something beneficial, they can bring it back to the population.

*Flocking / Complex* - Agents’ heading and velocity are affected by neighbors who are within their SIH. There is alignment between them, but the network is sparse and loose so that group orientations change over time. Flocking activity is a balance between environmental and social constraints.

*Stampede / Monolithic* - All members are stiffly connected. Alignment can become total and supports runaway conditions [13]. Such strongly aligned systems work well

when there are few environmental risks but have difficulty in adapting to new situations [7], and can literally run off cliffs [16].

The middle ground between standalone and monolithic behavior is rarely touched by our designs. The systems that we develop operate at either end of this spectrum (figure 2). At the standalone end is NASA's Curiosity rover. At the other is a fully automated robot factory. We are starting to explore a middle ground of resilient, self-organizing systems, such as High Frequency Trading which coordinates using markets [14], and Deep Neural Networks, where neuron "agents" interact using backpropagation to recognize patterns.

## 4 Engineering better information ecologies

Setting up ecologies is very different from designing systems. Ecologies require a substrate that also serves as some kind of fitness test [10]. In living systems, this is typically energy (food & light) and protection (defenses & group safety). The fitness test is survival and reproduction. In designed systems, the substrate is us. The fitness test is sales and contracts, as experienced by companies, their managers, designers, and customers. Which is why "adaptation" consists of patches, upgrades, and new models. When we are part of the system, it is adaptable. When we are not, it withers and fades.

Designing resilient systems, using the principles of ecology, is a capacity that we need to develop as we drift from our barely manageable current state towards more chaotic and unpredictable technological interactions. But there is some low hanging fruit. In particular, diversity injection. Let us revisit our firestorm example, but this time with an ecosystem of multiple different vehicles.

We are in the first group of vehicles on the same ridge road heading into the fire. In front of us, a handful of cars are already off the road, engulfed in flames. But the next car in line has frozen in the middle of the road. Its electronics are shorting out. The motor has seized, and the radio is blaring. This model is notorious for failures, and has been known to hit pedestrians. It is, in short, terrible.

But it is *different*. And it is *blocking the road*.

As before, but much earlier, the routing algorithm kicks in and cars fleeing the freeway are redirected to places that are *not* on fire. Resulting in hundreds of lives saved. All because there were a few, cheap, poorly made autonomous vehicles in the mix.

A population that is diverse has a greater chance of responding to unanticipated problems in unanticipated ways. The trade-off is lower efficiency, but that's a feature, not a bug. We need to start creating self-regulating systems that are inherently less likely to stampede out of control. Diversity injection is an accessible starting point. To begin this, *we ourselves* need diversity in the form of transdisciplinary research. For example, useful frameworks may emerge from the interplay of information science, ecology, and economics. This will enable new design languages for safe, self-regulating systems. Without which, we run the risk of being victims of our own complex systems as they drive us unaware into the wildfire.

## References

1. Caltrans: 2016 traffic volumes on california state highways (2016), [http://www.dot.ca.gov/trafficops/census/docs/2016\\_aadt\\_volumes.pdf](http://www.dot.ca.gov/trafficops/census/docs/2016_aadt_volumes.pdf)
2. Cucker, F., Smale, S.: Emergent behavior in flocks. *IEEE Transactions on automatic control* **52**(5), 852–862 (2007)
3. Feldman, P.: With maps and mobs: Searching for trustworthiness using belief spaces. In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. pp. 351–353. CHIIR '18, ACM, New York, NY, USA (2018)
4. Feldman, P., Dant, A., Lutters, W.: Simon’s anthill: Mapping and navigating belief spaces. In: *Proceedings of the 2018 Conference on Collective Intelligence* (2018), <http://arxiv.org/abs/1804.05409>
5. Feldman, P., Dant, A., Lutters, W.: This one simple trick disrupts digital communities. In: *Proceedings of the 2018 Conference on Self Adaptive and Self Organizing Systems (SASO)*. IEEE (2018), <http://arxiv.org/abs/1804.02251>
6. Giles, K.: *Handbook of russian information warfare*. NATO Defense College, Research Division (2016)
7. Grünbaum, D.: Schooling as a strategy for taxis in a noisy environment. *Evolutionary Ecology* **12**(5), 503–522 (1998)
8. Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* **4**(2), 100–107 (1968)
9. Holling, C.S., Gunderson, L.H.: Resilience and adaptive cycles. In: *Panarchy: Understanding Transformations in Human and Natural Systems*, 25–62 (2002)
10. Kaufman, S.: *At home in the universe: The search for the laws of self-organization and complexity*. New York: Oxford University Press (1995)
11. Kiesler, S., Sproull, L.: Group decision making and communication technology. *Organizational behavior and human decision processes* **52**(1), 96–123 (1992)
12. Kirilenko, A., Kyle, A.S., Samadi, M., Tuzun, T.: The flash crash: High-frequency trading in an electronic market. *The Journal of Finance* (2017)
13. Lande, R.: Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences* **78**(6), 3721–3725 (1981)
14. MacKenzie, D.: *A sociology of algorithms: High-frequency trading and the shaping of markets*. Preprint. School of Social and Political Science, University of Edinburgh (2014)
15. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* **63**(2), 81 (1956)
16. Patent, D.H.: *The buffalo and the Indians: A shared destiny*. Houghton Mifflin Harcourt (2006)
17. Stewart, L.G., Arif, A., Starbird, K.: Examining trolls and polarization with a retweet network. In: *Proc. ACM WSDM, Workshop on Misinformation and Misbehavior Mining on the Web* (2018)