Stampeding Robots

Philip Feldman

2018-09-18

Abstract

Machines are designed to communicate quickly and efficiently. Humans are not. Humans and other animals [~~,~~] ~~appear to~~ have evolved social structures that function best with approximately seven contacts at any given moment [5]. This combination of small and large scale information can produce emergent behaviors such as flocking birds or fashion trends. Technology changes this dynamic, by allowing all individuals in a population to be connected at the speed of light. A dense, tightly connected population can behave like a single individual. In animals, this happens in constrained areas like slot canyons, where stampedes can easily be triggered. But ~~our~~ machines, ~~ever more and ever~~ better connected don't need these kinds of conditions to stampede. ~~We've designed that in.~~ The very techniques used to design best-of breed solutions may place large numbers of people at risk from dangerous mass behaviors among extremely homogeneous machines. In this paper we explore some scenarios, and argue that the presence of diversity is the only broadly effective approach to defend against lethal unintended consequences at scale. [4]

1 Introduction

~~Consider a self-driving car scenario. The time is~~ the near future. It's summer in California, and once more it is a land of record heat and wildfires. The self driving car is taking over. Transportation as a service has turned out to be cheaper than anticipated, and hardly anyone drives any more. Traffic jams are generally a thing of the past, as sophisticated routing algorithms keep traffic running quickly on the highways or route around the infrequent problem. ~~And the these cars are *fast*~~. On their designated lanes they travel in tight, aerodynamic packs at ~~inhuman~~ speeds. Commute times that used to take hours now take minutes. "Manual" cars, trapped in their garages by ever increasing insurance start to become collector's items.

This summer, a large fire is heading towards greater Los Angeles. Fed by hurricane-force winds it progresses with unanticipated speed, covering 10 miles in 30 minutes and endangering the 405 near Bel-Air. The police block off the highway, and ~~although a few cars are stuck at the roadblock, the~~ rest of the traffic quickly routes around the obstruction.

Through burning neighborhoods. Many of the routes are also blocked, but there are many paths and A* [3] is relentless in finding optimal routes. A way through is quickly found. Thousands of cars optimally converge on the way through, a road running ~~along a ridge into the heart of the fire.~~

Although these cars are state-of-the-art and handle normal circumstances, 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. ~~Just a little brighter than usual, but within acceptable ranges.~~

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

Thousands of cars head into the flames.

One by one, the vehicles closest to the fire have their antenna burn off disappear from network. Shortly after that, the car itself begins to burn. It drifts off the road and rolls down the hill into a gully. Sometimes the passengers escape. Most of the time they don't. ~~The cars themselves~~ add fuel to the fire as batteries overheat and explode, ~~and fiberglass shells ignite.~~

To the routing system and to the vehicle's sensors, the way stays open. Cars continue to flood through the opening until the pile of burnt-out hulks finally starts to push up onto the road. Undaunted, the routing algorithm continues to do its work, but by now the the only routs that are open are far enough away from the inferno that traffic stops killing people.

It takes ~~a few~~ weeks to figure out what happened and issue a patch, and once more, thousands of identical cars are back on the roads. They are just as vulnerable to the next unforeseen problem. Is there anything that can be done to change this calculus so that every potential problem has to be foreseen?

2 Previous Work

Our best design practices a are about building for scale. Mass production, Software as a Service, reusable libraries, genetic sequencers. But with such systems resilience has to be deliberately planned for. We design for graceful degradation, to fail safely. We deisgn out the unpredictable, the random. As I write this, I'm flying home from Europe. I *want* these systems keeping me in the air to perform flawlessly and uneventfully.

But I think that the systems we have been designing can really be thought of as collections of one or more identical items, made to operate as individual components or combined as subcomponents designed into a smoothly operating whole. We don't think about designing populations and ecosystems, where emergence and surprise are the bedrock of resilience. I'd like to discuss in this paper why these types of designs are different, and why it matters, well beyond the scenario in the previous section.

Sorcerers apprentice more than the Golum. Something about ecology? The vulnerability of monocultures is well known [2]

3 Simulation Results

We built a model that is based on two ideas: 1) that human navigation through *belief space* (a subset of information space that contains items associated with opinions) is analogous to animal motion through physical space, and 2) that the *digital inadvertent social information* provided by humans interacting with the belief environment can be used to characterize the underlying belief space. To explore these concepts in depth, we built a standalone simulator (Fig.1), based on the Reynolds model [?], that represents belief space as a hypercube composed of labeled cells.

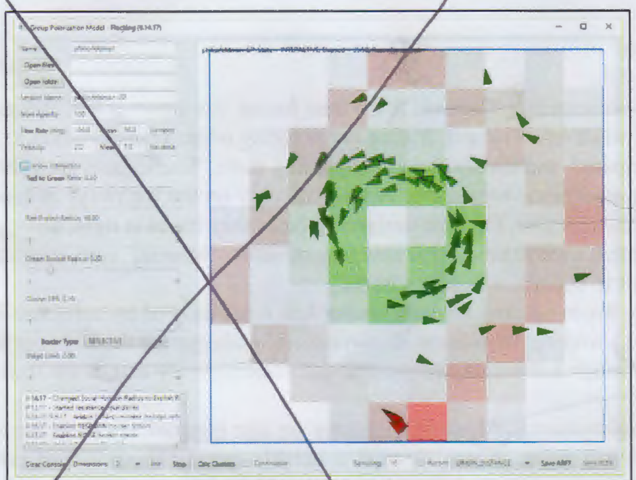

Figure 1: Our simulator

Stiff, densely connected networks vs loose networks connected with springs.
Diversity injection disrupts stampedes
Stampedes can be made
Diversity injection disrupts those too

This is a flat, enclosed, uniform and simple space. But it manifests in two levels - the negotiation on heading and velocity resembles opinion dynamics when using that frame of reverence, but the position of the agents in the space is critical, because that's where the complex behaviors emerge and also that's how the borders are encountered. The environment, mediated through the agent's position and their social influence horizon has a profound effect. What happens as becomes more heterogeneous? What kind of informational landscapes and the fitness landscape of the parameter space it corresponds with become more sophisticated?

4 Discussion

The results of the of the simulation are summed up in figure 2.

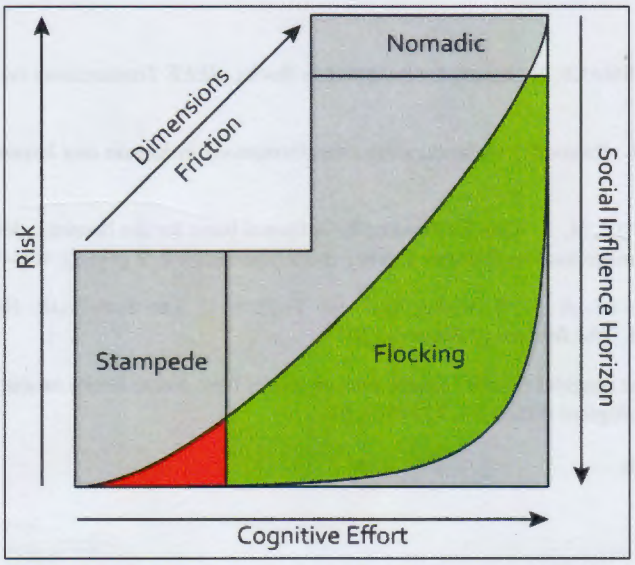

Figure 2: Continuum from Nomadic to Stampede behaviors

Stampede is a an attractor or sink. Under many conditions, it makes sense to just follow along with the group (examples). But at the same time, as the group becomes more homogenious, it produces nonconformists who leave to explore, and also outcasts. ~~The former may be better prepared then the latter to survive without a social safety net, but chromosomes don't care about that.~~ Nomads find new places as the original population transitions from flock to stampede to oblivion, and a new population emerges from the diaspora to repeat the pattern.

This large middle ground between individual nomads unaffected by others and the monolithic behavior of a stampede is a space rarely touched by deliberate design. The systems that we do design operate at either end of this spectrum, not at the middle. We design single standalone systems (Think Curiosity and the Large Hadron Collider), or mass produce interchangeable items (Your automobile, or your cell phone). In the last few decades, starting with operating systems, we've started to design platforms as well which start to fill in the middle space (more like a game with rules as compared with the interconnected, highly functional complexity of a tropical rain forest).

Ecosystems work for reasons. Monocultures are brittle and require constand intervention. Many small loosly coupled patches of multiresolution environments make it difficult for one element to fail in a way that endangers the entire system.

Designing large-scale systems as we do currently is the philosophy of factory farms. Efficient, but risky. This is not just the case for autonomous populations, it is the case for human design as well. As we connect ourselves ever more tightly and densely together, our communication patterns start to resemble that of the systems we build (unsurprising, kind of Worfian [sp?] hypothesis for tech). We split off into tight clusters of ever-more identically-thinking units, ever more prone (and vulnerable) to stampede.

How to address this corner that we've painted ourselves into is a long-term project. What does it mean to design socio-cultural systems for humans, robots, and mixtures of the two? I think that there are three paths that need to be taken

1. Short term fixes (do no harm - randomization, etc)
2. Socio-cultural User Interfaces (Lists, stories and maps + other?)
3. More research into how we make decisions as digital communities. Not only what creates and how to disrupt dangerous radicalization but also how to recognize emergent, novel thinking and support and nurture it.

5 Implications for Design

Diversity, even if it's far from optimal. Revisit of the fire scenario with Uber Murder Cars

How do we include this in design?