Swarm Intelligence and the Morality of the "Hive Mind"

LOUIS ROSENBERG & DAVID BALTAXE, Unanimous A.I.

1. INTRODUCTION

When designing artificially intelligent systems, researchers generally turn to Mother Nature for guidance. Not surprisingly, the first models explored were based on the framework most familiar to us humans – our own brains. Starting with Perceptrons of the 1950's [Rosenblatt, 1957] and continuing to the present, artificial Neural Networks have become the dominant biologically inspired paradigm for A.I. systems. Nature, however, is highly diverse. Billions of years of evolution, driven by natural selection, have produced at least one alternate method for assembling high-level intelligence and it's not neural – it's collective – enabling populations to think together in real-time synchrony.

Referred to as *Swarm Intelligence* (SI), nature shows us that by forming closed-loop systems among groups of independent agents, high-level intelligence can emerge that exceeds the capacity of the individual members. Researchers have explored this extensively for use among groups of networked robots and simulated agents [Beni et al., 1989], but only recently has swarming been applied to real-time networks of human participants [Rosenberg, 2015; Palmer et at., 2014; Eberhart et al., 2015].

Known as *Artificial Swarm Intelligence* (ASI), these systems enable human groups to work together in synchrony, forging unified systems that can answer questions, make predictions, and reach decisions by collectively exploring a decision-space and converging on preferred solutions. Prior studies have shown that by working together in real-time, human swarms can outperform individuals as well as outperform traditional methods for tapping the wisdom of groups such as polls, votes, and markets. For example, a recent study tasked a group of human subjects with predicting the top 15 awards of the 2015 Oscars. This was performed both by traditional poll and real-time swarm. Among 48 participants, the average individual achieved 6 correct predictions on the poll (40% success). When taking most popular prediction in the poll (across all 48 subjects), the group achieved 7 correct predictions (47% success), a modest increase. When working together as a real-time swarm, the group achieved 11 correct predictions (73% success) [Rosenberg, 2015]. This suggests that human swarming may be a superior method for tapping the wisdom of crowds.

Because Artificial Swarm Intelligence may be a viable path for building systems that exceed natural human intellect, we must ask whether the decisions reached by human swarms are likely be *moral* as compared to individual decisions, and compared to decisions generated through traditional collective intelligence methods. The prevalence of negative phrases such as "hive mind" and "mob mentality" suggests public angst over decisions reached collectively. The present study explores this formally, comparing decisions made by (i) individuals, (ii) group-wise polls, and (iii) real-time swarms, when each is confronted with a moral dilemma in a traditional "Tragedy of the Commons" format.

2. SWARMS AS INTELLIGENT SYSTEMS

Among A.I. researchers, the word "swarm" generally refers to groups of robots or simulated agents governed by simple localized rules. The most common of these systems are inspired by flocks of birds

and schools of fish and usually focus on enabling groups to navigate environments together. While such systems have many valuable applications, for example enabling robotic drones to fly in unison, the *human swarms* discussed herein are modeled less after the collective motions of flocks and schools, and more after the collective *decision-making* processes used by honeybees. This is because the impressive decision-making abilities of honeybee swarms provide a powerful natural proof of the potential for an emergent decentralized parallelized intelligence.

2.1   Natural Swarms and the Power of Decentralized Decision-making

As studied by Seeley et al., the underlying processes that govern decision-making in honeybee swarms are remarkably similar to the decision-making processes in neurological brains [Seeley, 2010; Seeley et al., 2012] Both employ large populations of simple excitable units (i.e., bees and neurons) that work in parallel to integrate noisy evidence, weigh competing alternatives, and converge on decisions in synchrony. In both, outcomes are arrived at through a real-time competition among sub-populations of excitable units, each sub-population vying for one of a plurality of alternate solutions. When one sub-population exceeds a threshold level of support, the corresponding alternative is chosen. The required threshold in both brains and swarms is not the unanimous support of the population, or even a simple majority, but a sufficient quorum of excitation. In honeybees, this helps to avoid deadlocks and leads swarms to optimal decisions most of the time [Seeley, 2010; Seeley et al., 2012].

To fully appreciate the decision-making ability of natural swarms, it's worth reviewing a well-studied example – the ability of bee swarms to select an optimal hive location. Every spring honeybees face a life-or-death decision to select a new home location for the colony. From hollow trees to abandoned sheds, the colony considers dozens of candidate sites over a 30 square mile area, evaluating each with respect to dozens of competing criteria. Does it have sufficient ventilation? Is it safe from predators? Is it large enough to store honey for winter? It's a complex problem with many tradeoffs and a misstep can mean death to the colony. Using body vibrations known as "waggle dances", hundreds of bees express preferences for competing sites based on numerous quality factors. Through a real-time negotiation, a decision is reached when a sufficient quorum emerges.

Remarkably, the bees arrive at optimal decisions 80% of the time [Seeley, 2010]. Thus, although individual bees lack the capacity to make a decision this complex and nuanced, when hundreds of scout bees pool their knowledge and experience, they evoke a collective intelligence that is not only able to reach a decision, it finds an optimal solution. Thus by working together as a unified system, the colony amplifies its intelligence beyond the capacity of individual members. It is this emergent amplification of intelligence that *human swarming* aims to enable among groups of networked people.

2.2   Enabling Human Swarms

To evoke a Swarm Intelligence among groups of networked users, an online platform called **UNU** was developed. Modeled after the decision-making of natural swarms, UNU allows groups of independent actors to work in parallel to (a) integrate noisy evidence, (b) weigh competing alternatives, and (c) converge on final decisions in synchrony. Because humans can't *waggle dance* like honeybees, a novel interface had to be developed to allow participants to convey their individual intent with respect to a set of alternatives. In addition, the interface had to be crafted to allow users to perceive and react to the changing system in real-time, thereby closing a feedback loop around the full population.

As shown in Figure 1 below, participants in the UNU platform answer questions by collectively moving a graphical puck to select among a set of alternatives. The puck is modeled as a physical system with a defined mass, damping and friction. Participants provide input by manipulating a graphical magnet with a mouse or touchscreen. By positioning their magnet, users impart their personal intent as a force vector on the puck. The input from each user is not a discrete vote, but a stream of vectors that varies freely over time. Because the full set of users can adjust their intent at every time-step, the puck moves, not based on the input of any individual, but based on the dynamics of the full system. This results in a real-time *physical negotiation* among the members of the swarm, the group collectively exploring the decision-space and converging on the most agreeable answer.
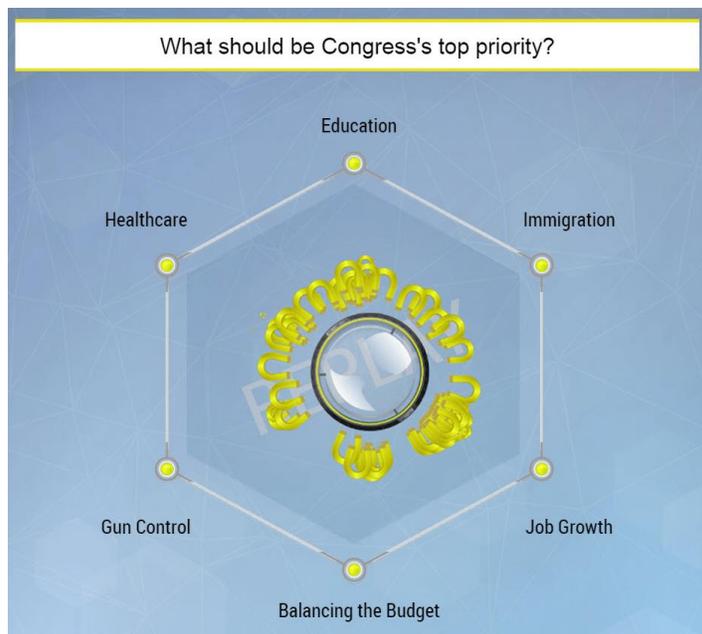

Fig. 1. Shown is a snapshot of a human swarm in the process of answering a question in real-time. Each magnet represents a live participant networked to the UNU platform from around the world. This shows 110 users working as a unified system.

We must note that users can only see *their own magnet* during the decision**,** and not the magnets of others users. Thus, although they can view the puck's motion in real time, which represents the emerging will of the full swarm, they are not influenced by the specific breakdown of support across the options. This limits social biasing. For example, if the puck slows due to an emerging deadlock, the participants must evaluate their own willingness to shift support to alternate solutions without knowing the specific distribution of support that caused the deadlock.

In Figure 1 above, an example question is shown as it would appear simultaneously on the screens of all networked participants. In this trial, a swarm of 110 users was asked to grapple with a politically charged question: *"What should be Congress's top priority?"* Users are given a *3,2,1* countdown to coordinate the start of the session. The swarm then springs into action, working in synchrony to guide the puck to a preferred answer.

The decision process is generally a complex negotiation, with individuals shifting their support numerous times to break deadlocks or defend against options they disfavor. When a user pulls towards one option in the answer set, a component of their force also acts to impede the motion of the puck towards competing options. In this way, users don't only add support a preferred solution when

pulling towards it, but also suppress solutions they don't prefer. This enables the dual process seen in natural swarms and neurological brains wherein individual agents are enabled to both excite and inhibit, thereby reducing the chances of a deadlock.

If a group happens to be in substantial agreement at the very start of the question, the puck moves smoothly to the preferred answer. But, if two or more competing options have significant support, the swarm negotiates as a unified system. Most users begin by pulling towards the option they prefer most, then shift to alternate choices if the puck starts moving towards an option they dislike. With all users making these changes in parallel, the swarm explores the decision space and converges on an answer that often optimizes group satisfaction.

It's important to note that users don't just vary the direction of their input, but also the magnitude by adjusting the distance between the magnet and the puck. Because the puck is in motion, to apply full force users need to continually move their magnet so that it stays close to the puck's rim. This is significant, for it requires all users to be engaged during the decision process. If they stop adjusting their magnet to the changing position of puck, the distance grows and their applied force wanes. Thus, like bees executing a waggle dance or neurons firing activation signals, the users in an artificial swarm must continuously express their changing preferences during the decision process or lose their influence over the outcome.

Post testing interviews with participants suggest that users with high levels of conviction in favor of a particular outcome are more vigilant in maintaining maximum force on the moving puck. Conversely, users who have lower conviction are less vigilant. In this way, the swarming interface allows the population to convey varying levels of conviction in real-time synchrony. We believe this helps the swarms converge on solutions that optimize the overall satisfaction of the group.

Observations and post-testing interviews also reveal that human swarming yields consistent outcomes across varying spatial placement of answer options. For example, if two highly favored options are placed on opposite sides of the puck's starting position, the swarm will fall into an early deadlock as it grapples between them. Conversely, if the two highly favored options are placed on the same side of the puck's starting position, the swarm will not fall into an early deadlock, but instead move the puck towards those two highly favored options. Still, a deadlock will emerge as the puck approaches midpoint between the two favored options. In this way, the decision space can have alternate layouts, but the swarm arrives at the same outcome. A similar robustness has been observed in honeybee swarms, which are known to decide upon optimal nesting locations regardless of the order in which sites are discovered and reported by scout bees [Seeley, 2010].

Referring again to Figure 1 the default layout of answers is a set of six options in a hexagon pattern. This configuration was chosen because according to social-science research, people are efficient decision-makers when presented with up to six options, but suffer from increasing "choice-overload" inefficiencies when confronted with larger sets [Scheibehenne, 2010]. To enable swarms to consider larger sets of answers, the system employs an iterative approach, presenting users with a series of six-option subsets of the full answer pool, then pitting the winner of each subset against each other. The system also allows swarms to select values on a continuous numerical scale. This enables swarms to collectively decide upon quantities, prices, percentages, odds and other numerical values.

An example of a question using a numerical scale is shown in Figure 3. In this trial, a swarm of human subjects was asked to decide upon the fair price of a movie ticket on a scale from $0 to $25. When using this framework, the puck starts at the midpoint of the scale and can be moved smoothly in either direction. In these types of collective decisions, the swarm generally overshoots the final answer, then reverses the direction of the puck, oscillating in narrower and narrower bands as all the

users adjust their pulls in parallel. An answer is chosen from the continuous range when the puck settles upon value for more than a threshold amount of time (e.g., 3 seconds).
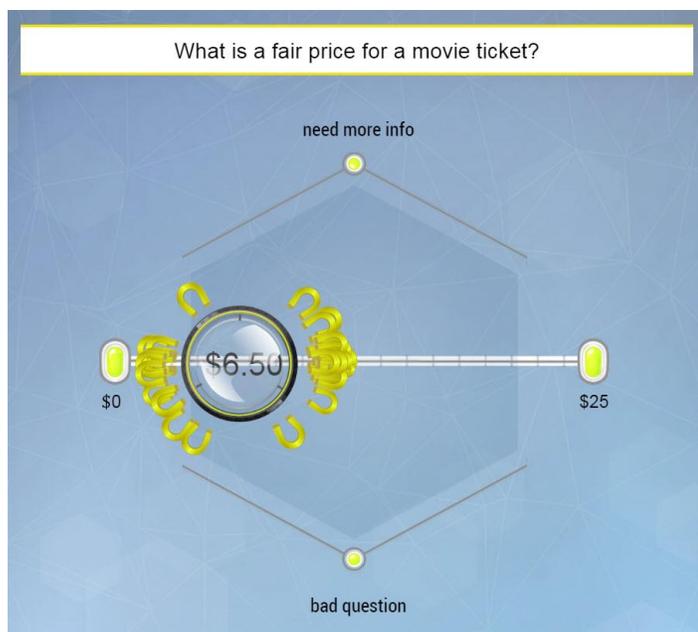

Fig. 3. Shown is a snapshot of a human swarm in the process of answering a question on a continuous scale. Each magnet represents a live participant, the full population working together in real-time.

3. **SWARMING AND MORALITY**

We humans are rational thinkers with an inherent sense of morality that guides us towards the greater good. This hold true across all levels of society and yet collectively, we often make decisions that lead to self-destructive outcomes such as war, pollution, poverty, inequality and climate change. But how can immoral decisions emerge from a society comprised almost entirely of moral individuals? Philosophers have been pondering this for ages. Nietzsche lamented – *"Madness is rare in individuals - but in groups, political parties, nations, and eras it's the rule."* Renowned American theologian, Reinhold Niebuhr was even more blunt, expressing –*"the group is more arrogant, hypocritical, self-centered, and more ruthless in the pursuit of its ends than the individual."* Is there something about human groups that cause us to behave so differently together than we do alone?

Social scientists often cite the "Tragedy of the Commons" when exploring group morality. First postulated by the Victorian economist William Foster Lloyd in 1833, the premise is that individuals, who act both morally and rationally on a local scale, are prone to producing immoral results when operating on a group scale [Lloyd, 1833]. He pointed to herdsman running cattle on open pastures. As an individual rancher, it's rational and moral to maximize the size of your herd. But, if all herdsmen follow this individual morality, the shared pasture gets overrun and is ruined for all. Thus individual morality is not always aligned with the common good.

Ecologist Garett Hardin brought TOC dilemmas to modern relevance in a 1968 when he linked this to population growth in the journal Science [Hardin, 1968]. He pointed out that at the most local level, it's a basic Human Right for parents to decide the number of children to have. And for much of the

world's inhabitants, a large brood is rational, optimizing long-term survival of the family. On a global level, however, if all families behave under that same local morality, overpopulation will likely result, putting most of the world's families in a weaker position for survival. With global population having doubled from 3.5 B to 7.1 B over the decades since Hardin published his famous paper, it's safe to say that we've not learned how to overcome TOC dilemmas.

A clever example of how easy it is for people to fall victim to TOC pitfalls was recently performed at University of Maryland by Dylan Selterman. He posed an extra-credit challenge to his Social Psychology class, allowing each of his students to indicate by secret ballot how many points of extra credit they wanted on their exam – 2 points or 6 points. The only twist was that if more than 10% of the class asked for 6 points, nobody would get any bonus. Clearly, it was in the best interest for everyone in the class to individually ask for 2 points, but that's not what happened. Far too many students asked for 6 points and nobody received extra credit [Selterman, 2015].

So, how do we handle social dilemmas in which the short-term interests of individuals are at odds with the long-term interests of the group? To date, one of the more successful path has been the use of democratic governance in which groups make decisions collectively through direct or representative polling of the population. The presumption is that by revealing the consequences of their collective actions to the overall population, democratic decisions will emerge that support the common good. The problem is, however, our current methods for polling groups through traditional votes and polls often fall victim to the "Tragedy of the Commons" pitfall [Cohen et. Al, 2006].

With human groups having a propensity to fail TOC style dilemmas, we must ponder what this means for systems that generate decisions based on input from large groups. Furthermore, we must ponder what this means for Artificial Swarm Intelligence systems that use synchronous input from networked groups to foster the emergence of a real-time intelligence. Does this suggest that the resulting intelligence will make decisions that are less moral than the individuals who comprise it?

4.   MORALITY TESTING

4.1   Morality Testing, Part I

To test the morality of human swarms, a set of pilot studies were conducted using the **UNU** platform. The tests tasked groups of networked participants with simple financial decisions framed as a Tragedy of the Commons dilemma. The first set of tests (Part I) were designed to compare the decisions made in response to the TOC dilemma under two experimental conditions – (i) test subjects were required to make their decisions through a standard online survey and (ii) test subjects were required to make their decision by participating a real-time swarm.

The test engaged 18 randomly selected online test subjects, each paid $1.00 for their participation. All were told they would get an additional bonus of $0.30 or $0.90. They simply had to indicate which bonus they wanted to receive. Of course there was a TOC catch – if more than 30% of the group asked for $0.90, then nobody would get anything. This means oversubscription of the $0.90 option would defeat their common interests, resulting in none of the 18 participants receiving any bonus. All test subjects were clearly informed of this potential outcome, ensuring they understood the scenario.

When working as a swarm, the participants were instructed to move the graphical puck to one of six targets, three of which represented the $0.30 bonus, three of which represented the $0.90 bonus. Test subjects were told that any of the participants pulling towards $0.30 at the end of the decision would get that bonus, while anyone pulling towards $0.90 would get that bonus. Thus, subjects were able to pursue their individual interests while helping to guide the puck together as a swarm. Figure 3

shows a screenshot of the test with all magnets displayed. That said, the subjects were only able to see their own magnet, and thus could not see the pull directions of others.

Fig 3. Shown is a human swarm of twenty users, each of them networked to UNU from distributed locations around the world. Every user is represented by a unique graphical magnet as they work in synchrony to choose a bonus.

4.2 Results, Part I

When providing their input through the traditional online survey, 67% of participants asked for a $0.90 award, well beyond the 30% threshold. Thus, nobody received a cash bonus, the group failing to achieve an outcome that supported their common interests. This result, while self-defeating, is typical of TOC dilemmas. The swarm, on the other hand, configured itself such that 24% of the total pull on the puck was towards $0.90, with 70% of the total pull towards $0.30, and 6% abstaining. It's important to repeat that participants could only see the puck and their own magnet, but not the magnets of other users. Still, the group, when functioning as a unified system connected by real-time feedback loops, avoided the TOC pitfall. In fact, the swarm converged on a solution that optimized the payout for the full group.

4.3 Morality Testing, Part II

The second set of tests was designed to extend the prior testing from individual decisions to team decisions. More specifically, the Part II tests were aimed at comparing TOC decisions made under two new conditions – (i) test subjects were required to make a team decision by majority vote, again using a standard online poll, and (ii) test subjects were required to make a team decision by forming a real-time swarm, the team converging on an answer together, in synchrony.

The test engaged 70 randomly selected online test subjects, each paid $1.00 for their participation. These participants were split into three separate teams (Team Orange, Team Yellow, and Team Purple). Each team was told that all its members would get an additional bonus of either $0.25 or $0.75. The team simply had to indicate which bonus it wanted to receive as a team. Of course there was a TOC catch – if more than one of the three teams asked for $0.75, then none of the teams would get bonuses for its members. This means oversubscription of the $0.75 option would defeat their common interests, resulting in none of the 70 participants receiving a bonus. All test subjects were clearly informed of this potential outcome and pitfalls.

When responding by survey, the results were blind such that subjects had no indication of what other participants had selected for their response or what other teams had selected by majority vote. When responding as a real-time swarm, each team made their decision as a unified system, but had no indication of what the other teams had selected.

4.4 Results, Part II

Looking first across all 70 individual test subjects, we find that 47 of the 70 respondents requested the larger bonus of $0.75. This equates to 67% of respondents requesting a bonus that they know would only be awarded if less than 33% of the test subjects made the same selection. Thus, when viewed as a pool of disconnected individuals, the subjects once again failed the TOC dilemma.

Of course, this round of testing was not focused on individual decisions, but team decisions. Processing the survey results, we determined the collective decision of each team by majority vote. Team Orange voted collectively for the large bonus (60% in favor). Team Yellow voted collectively for the large bonus (70% in favor). And Team Purple voted collectively for the large bonus (55% in favor). Thus, all three teams requested the large bonus. The rules stated, however, that if more than one team requests the large bonus, no teams get any bonus. Thus, when making their decisions by majority vote, the three teams failed the TOC dilemma.

Finally, we reviewed the decisions made by each of the teams when each functioned as its own unified system, its members collectively moving the graphical puck in real-time synchrony. When using this swarming method, Team Orange collectively asked for the small bonus. Team Yellow collectively asked for the small bonus. And team Purple, collectively asked for the large bonus. Despite the fact that each of the teams had no knowledge of what the other teams had decided when swarming, they collectively optimized the payout to the participants, with two thirds of the participants getting $0.25 and one third of the participants getting $0.75. In this way, the groups overcame the pitfall of the TOC dilemma and achieved a solution that supported the common good.

5. DISCUSSION AND CONCLUSIONS

The results above support prior research that shows individuals acting alone, and teams acting by majority vote, are susceptible to the pitfalls of the TOC dilemma, producing outcomes that are at odds with the common good. What is encouraging, however, is that when working together as a real-time dynamic system (i.e. a "human swarm") the participants in the study did not fall victim to the pitfalls of the TOC dilemma, but instead optimized the payout for the full group. This suggests human swarming may be a viable technique for reaching decisions that are better aligned with the common interests of a group, as compared to poll-based methods for tapping collective intelligence. It further suggests that the emergent intelligence that arises from human swarms, may produce decisions that are more supportive of the common good than would come from the individual participants who comprise the swarm. This runs counter to the "hive mind" fears that permeate society.

So why do hives get a bad rap? It may stem from basic misconceptions about how natural swarms work. For example, many people assume that bees are "drones" take blind direction from an all-powerful queen. This is not the case. The queen has no known influence on colony decisions. Instead, honeybees make decisions by convening swarms of their most experienced members who negotiate and reach consensus, with few participants entrenching. This is arguably less "drone-like" than human elections wherein the majority of participants vote along entrenched party lines, the decisions being made by a small percentage of independents in the middle.
All in all, swarming is Mother Nature's brand of democracy, enabling groups to work together for the good of the population as a whole. The current study suggests that humans can benefit from

swarming when appropriate software is used to provide the real-time connections. Going forward, human swarming could point us to new methods for group decision-making across a wide range of applications, encouraging large groups to freely combine their individual knowledge, opinions, and interests in a manner that supports the common good.

Looking further out, online human swarms may be a path to super-intelligent systems. After all, a single honeybee lacks the intellectual capacity to even consider a complex problem like selecting a new home site for a colony, and yet swarms of bees have been shown to not only solve that multi-faceted problem, but find optimal solutions. If we humans could form similar swarms, we may be able to achieve similar boosts in intellect by thinking together in synchrony, solving problems that we currently, as individuals, can't tackle. This is not only an promising path to build smart systems, it is a path that keeps humans in the loop – possibly ensuring that any super-intelligence that emerges has our core values, morals, and interests at its core. From this perspective, Artificial Swarm Intelligence may be a safer approach to building A.I. systems than traditional methods.

REFERENCES

Cohen, Taya. Monotoya, Matthew, Insko, Chester. "Group Morality and Intergroup Relations' Society for Personality and Social Psychology", Vol 32 No 11, Nov 2006, pg 1559-1572.

Eberhart, Russell, Daniel Palmer, and Marc Kirschenbaum. "Beyond computational intelligence: blended intelligence." Swarm/Human Blended Intelligence Workshop (SHBI), 2015. IEEE, 2015.

Hardin, Garett. The Tragedy of the Commons. Science 13 Dec 1968: Vol. 162, Issue 3859, pp. 1243-1248

Lloyd, William. Two Lectures on the Checks to Population. Oxford, 1833.

Palmer, Daniel W., et al. "Emergent Diagnoses from a Collective of Radiologists: Algorithmic versus Social Consensus Strategies." Swarm Intelligence. Springer International Publishing, 2014. 222-229.

Rand, D. G., Arbesman, S. & Christakis, N. A. (2011) Dynamic social networks promote cooperation in experiments with humans. Proc. Natl Acad. Sci. USA 108, 19193–19198.

Rosenblatt, Frank, The Perceptron – a receiving and recognizing automoton. Report 85-460-1, Cornell Aeronautical Laboratory

Rosenberg, Louis, "Human Swarms, a real-time paradigm for collective intelligence." Collective Intelligence 2015, Santa Clara CA.

Rosenberg, Louis. "Human Swarms, a real-time method for collective intelligence." Proceedings of the European Conference on Artificial Life 2015, pp. 658-659

Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd. "Can there ever be too many options? A meta-analytic review of choice overload." Journal of Consumer Research 37.3 (2010): 409-425.

Selterman, Dylan: Why I give my students a 'tragedy of the commons' extra credit challenge The Washington Post, July 20, 2015.

Seeley, Thomas D., et al. "Stop signals provide cross inhibition in collective decision-making by honeybee swarms." Science 335.6064 (2012): 108-111.

Seeley, Thomas D. Honeybee Democracy. Princeton Univ. Press, 2010.

Seeley, Thomas D., Visscher, P. Kirk. Choosing a home: How the scouts in a honey bee swarm perceive the completion of their group decision making. Behavioural Ecology and Sociobiology 54 (5) 511-520.