Artificial Swarm Intelligence

Louis Rosenberg, PhD & Gregg Willcox
Unanimous AI

I. Introduction

The technology of Artificial Swarm Intelligence (ASI) has been shown to substantially amplify the collective insights and significantly increase predictive accuracy of human groups. It works by connecting teams of networked users into real-time systems moderated by AI algorithms. Sometimes referred to as *"human swarms"* or *"hive minds,"* these systems function very differently than traditional methods for harnessing the wisdom of human groups. Unlike votes, polls, surveys, or prediction markets, which treat each participant as a passive source of data for statistical aggregation, "swarming" treats each person as an active member of a real-time control system, enabling the full population to think together in synchrony and converge on optimized solutions as a unified amplified intelligence [1-8].

Inspired by nature, the algorithms used by ASI systems are modeled on the biological principles of Swarm Intelligence, the phenomenon that enables flocks of birds, schools of fish, and colonies of bees to maximize their collective intellect [9]. In this way, ASI systems enable networked human teams to quickly solve problems and reach decisions by converging on solutions that optimize their combined knowledge, wisdom, insights, and opinions. And because "human swarms" are interactive systems in which the participants act and react, continually adjusting their conviction as the group converges on solutions, much smaller populations are required to achieve statistically significant results than polls, surveys, or markets.

Fig 1*. Biological Swarms as Intelligent Systems*

Over the last five years, major progress has been made in the field of Artificial Swarm Intelligence, resulting in many studies that strongly validate the ability of ASI systems to amplify the intelligence of networked human groups. In 2015, the first study was published demonstrating that networked teams can produce collaborative forecasts by working together as swarm-based systems, achieving substantially higher accuracy than traditional crowd-based methods [2]. Over the years since, dozens of additional studies have been conducted, including:

In 2018, Stanford University published a study showing that small groups of radiologists, when connected by real-time swarming algorithms, diagnosed chest X-rays with 33% fewer errors than standard method [3,4]. Researchers at Boeing published a study showing that small teams of military pilots, when working in "human swarms," could generate qualitative insights about the design of cockpits with higher effectiveness than current methods [5]. Researchers at California Polytechnic published a study showing that networked business teams could increase their subjective decision-making accuracy by over 25% by swarming [6]. Researchers at Oxford University and Unanimous AI showed that small groups of financial traders, when forecasting market key indicators (Oil, Gold, and S&P), amplified their accuracy by over 25% by forming swarms [7]. And in 2019, researchers at Unanimous AI published a study showing that the behaviors of swarms could be post-processed by a dense neural network to amplify the predictive precision. In this study, groups of average sports fans were tasked with predicting the outcome of 238 basketball games in the NBA. The group significantly outperformed the Vegas odds-market, delivering a 24% ROI across the full set of games forecast. The post-processed forecasts more than doubled this ROI, achieving a 56% ROI over the full set of games [8].

II. Data-Points to Data-Processors

Research shows that human groups can significantly amplify their intelligence through swarming, outperforming individual experts and traditional crowd-based methods. But why? It's largely because the participants within real-time swarms serve a very different function than the "respondents" within votes, polls, surveys, and prediction markets. In traditional crowd-based instruments, respondents are simply that – a source of discrete responses that are captured as isolated data points and combined statistically with data from other respondents. While such methods are often said to *tap into the wisdom of crowds*, the "crowd" is a statistical metaphor for data aggregation. Even prediction markets are not truly interactive, as each transaction is between just one "buyer" and one "seller," executed in sequence over time. Such methods do not enable a population to interact together as an emergent intelligence [10-13].

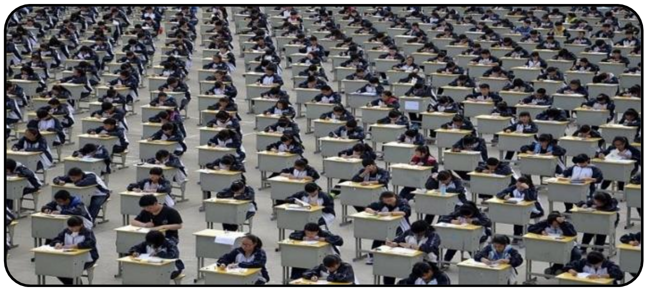

Fig 2. *Crowds as a metaphor for Data Aggregation*

When using ASI on the other hand, human participants are not treated as passive ***data-points***, but as active ***data-processors***, empowered to act, react and interact with the full population of other users. By thinking together in real-time systems, swarming groups interactively explore the decision-space and converge on solutions that maximize their combined knowledge, wisdom, insights, and opinions. Thus, while "a crowd" is just a statistical metaphor, "a swarm" is a true emergent system, powered by AI to amplify group intelligence. This enables any team, from small groups of financial traders to large engineering teams, to quickly and accurately answer questions, make predictions, reach decisions, prioritize options, and generate insights. Simply put, an ASI turns any team into a unified amplified intelligence.

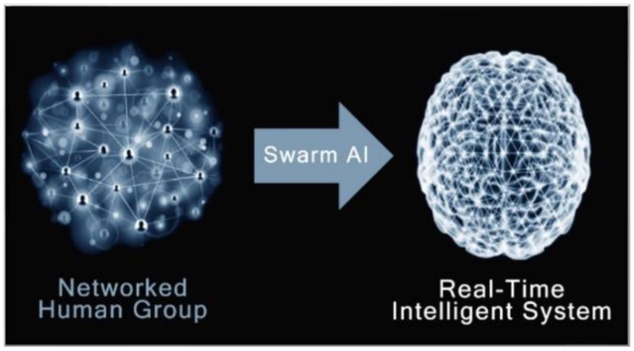

Fig 3*. Human Swarms as real-time Intelligent Systems*

III. From Biology to Technology

Swarm Intelligence is a biological phenomenon that spans a wide range of species, from schooling fish and flocking birds, to slime molds and bee swarms. The product of millions of years of evolution, biological swarming enables groups of organisms to collaboratively solve problems with accuracy and efficiency that is beyond the capability of the individual members. In this way, Swarm Intelligence is nature's method for tapping the insights and instincts of groups, creating a "super-organism" that is significantly smarter together than the individuals alone.

The most studied swarm in nature is that of honeybees, as their remarkable decision-making abilities have been researched since the 1950s and have been shown to be surprisingly similar to decision-making in neurological brains [14-15]. Brains and swarms both employ large populations of simple excitable units (i.e., bees and neurons) that work in parallel to (a) integrate noisy information, (b) weigh competing alternatives and (c) converge on optimized solutions in synchrony. In both brains and swarms, outcomes are arrived at through a real-time competition among sub-populations of excitable units until a dominant solution emerges. When one sub-population exceeds a threshold level of support, the corresponding alternative is decided upon. In honeybees, this enables populations of simple "scout bees" to collect information about their environment and converge on optimal decisions when searching for a new home site, finding the ideal solution over 80% of the time [15-17].

The similarity between brains and swarms becomes even more apparent when comparing models that represent each. For example, the decision-making process in primate brains is often represented as a mutually inhibitory "leaky integrator" that aggregates incoming data from competing neural populations and gradually attenuate support over time [18]. This model indicates that (a) support signals must be continually maintained over the decision period or lose influence and (b) decisions are reached when a threshold level of activation is exceeded. A common framework for representing primate decisions is the Usher-McClelland model in Figure 4 below.

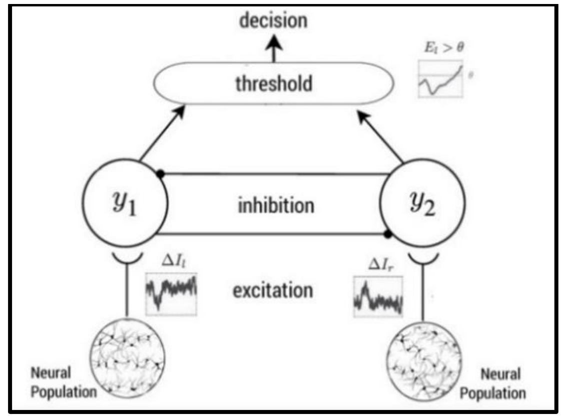

Fig. 4. *Usher-McClelland model of neurological decision-making*

This neurological decision model can be directly compared to swarm-based decision models represented in Figure 5 below. As shown, swarm-based decisions follow a similar process to neurological brains, aggregating input from sub-populations of participants through mutual excitation and inhibition, until a threshold level of support is exceeded. In fact, bee swarms have been shown to employ the same "leaky integrator" model in the vibrations they generate with their bodies that neuron employ with activation signals. When viewed in this context, it becomes apparent that, while biological brains are systems of neurons structured such that intelligence emerges, biological swarms are systems of brains structured such that **amplified intelligence** emerges. Simply put, a swarm is a "brain of brains."

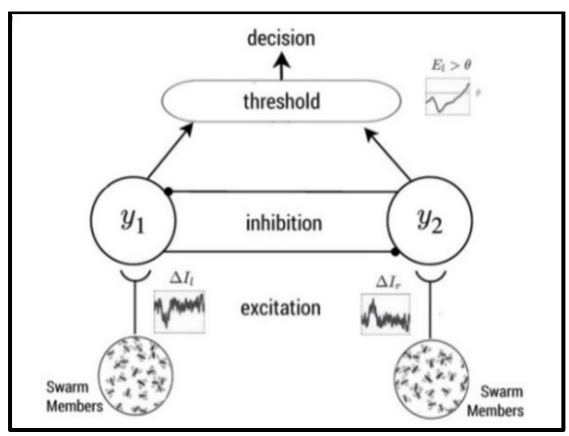

Fig. 5. *Mutually inhibitory decision model in bee swarms*

If birds, bees and fish can amplify their intelligence by forming "super-organisms," it seems natural that human teams could benefit by combining insights in similar ways. This has been the motivation for researchers to develop ASI technology as a set of algorithms and interfaces that turns networked teams into intelligent systems modeled after natural swarms.

IV. ENABLING "HUMAN SWARMS"

Unlike many other social species, humans have not evolved the natural ability to form closed-loop swarms that converge in synchrony on optimized solution. That's because we lack the subtle connections that other organisms use to establish high speed feedback-loops among members. Schooling fish detect vibrations in the water around them. Flocking birds detect subtle motions propagating through the population. Swarming bees use complex body vibrations called a "waggle dance." To enable swarming by networked human groups, specialized technology is required in lieu of these natural abilities.

To address this need, the **Swarm®** platform was developed and deployed by Unanimous AI. It enables networked groups to think together as real-time systems, connecting from anywhere in the world using standard web browsers. Modeled largely on the decision-making process of honeybee swarms, the Swarm platform (and underlying Swarm AI technology) empowers online teams to perform the biologically inspired steps of (a) integrating noisy information, (b) weighing competing alternatives, and (c) converging on optimized solutions as real-time closed-loop systems. A screen shot from the Swarm platform is shown below in Figure 6.

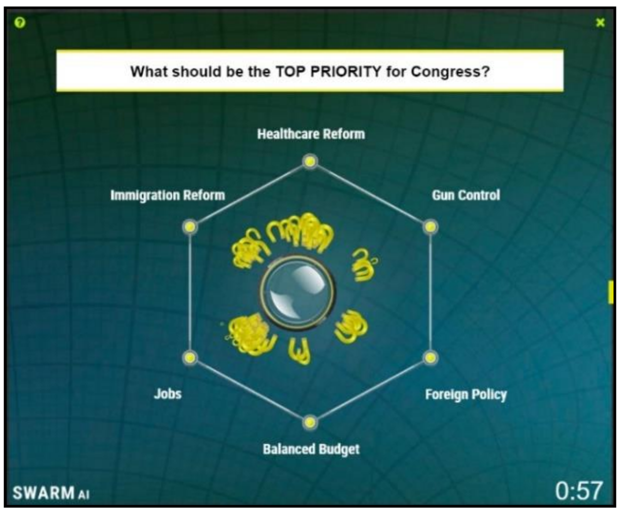

Fig. 6. *A human swarm answering a question in real-time (view replay: swarm.ai/r/1624)*

As shown above, swarms converges on answers by working together to move a **graphical puck**, positioning it to select a preferred solution from among a set of alternatives. Each participant provides individual input by manipulating a graphical magnet with a mouse or touchscreen. By positioning their magnet with respect to the moving puck, participants "apply forces" on the system, expressing and imparting their personal intent upon the swarm as a whole. The input from each user is not a discrete vote, but a stream of vectors that varies freely over time. Because the full population of users can adjust their intent continuously in real-time, the swarm moves, not based on the input of any individual member, but based on the dynamics of the full system. This enables a complex physical negotiation among all members at once, empowering the group to explore the decision-space together and converge on the most agreeable solution in synchrony.

It is important to note that participants do not only vary the direction of their intent, but also modulate the magnitude of their intent by adjusting the distance between their magnet and the puck. Because the puck is in continuous motion across the decision-space, users need to continually move their magnet so that it stays close to the puck's outer rim. This emulates the *"leaky integrator model"* employed by biological systems, requiring each participant's support signal be maintained over time or lose influence. In other words, participants must continuously engage the puck throughout the decision process, repeatedly evaluating and re-evaluating their intent as they convey their updated contribution. If they stop adjusting their magnet with respect to the changing position of the puck, the distance grows and their applied sentiment wanes.

Thus, like bees vibrating their bodies to express sentiment in a biological swarm, or neurons firing activation signals to express conviction levels within a biological neural-network, the participants in an artificial swarm must continuously update and express their changing preferences during the decision process, or lose their influence over the collective outcome. In addition, biologically inspired AI algorithms monitor the behaviors of all swarm members in real-time, inferring their implied conviction based upon their relative movement over time. This reveals a range of behavioral characteristics within the swarm population and weights their contributions accordingly.

The swarming algorithms that moderate the Swarm platform enable similar amplification effects achieved by neurological brains and biological swarms. Specifically, AI algorithms track the behavior of networked participants in real-time, monitoring how users modulate their sentiment every 250 milliseconds in response to others. In this way, swarms don't ask participants to merely "**report**" their views, as polls, surveys, and focus groups do. Instead, swarms inspire participants to "**behave**" as part of an interactive system, tracking their changing sentiments. This difference between "reporting" and "behaving" is significant to the power of swarm-based systems.

V. BEHAVING VS REPORTING

From conducting polls and surveys, to interviews and focus groups, it's common practice to ask participants to **self-report** their opinions, forecasts, and sentiments. Unfortunately, many studies have shown that individuals are highly unreliable when tasked with self-reporting their feelings [19,20]. Compounding this problem, participants generally express their views as numerical values on linear scales. Studies have shown that people are nonlinear-thinkers and that participants have different nonlinearities in the internal rating scales they employ [21-23] This means the underlying data used by traditional sampling methods can be highly distorted, tracking numerical values that appear similar on the surface, but mean different things to different respondents.

Swarming addresses this problem by not relying on how participants report, instead processing how they **behave** when engaging real-time systems that connect all members. This means for every participant, a large set of time-varying behavioral data is collected that reflects his or her authentic intentions, opinions, and/or beliefs in the context of all other participants, enabling them to converge not just on a common solution, but quantify it on a common numerical scale

VI. ASSESSING SWARMS

When collecting isolated data points via traditional polls and surveys, researchers perform statistical tests to determine if there is a central tendency in the data and if that tendency is statistically significant. In a control system, the outcome is not determined based on statistical inference across a sample of independent data points but based on whether the networked system of human data-processors is able to (a) **converge** upon a unified solution together, or (b) **diverge**, such that no solution can be reached. When answers are reached by the swarm through convergence, performance metrics are captured that indicate the degree of conviction within the system, giving valuable insights into the strength of the results.

Specifically, the speed of convergence and the degree of alignment among the swarm participants during a response is used to compute a useful performance metric known as Brainpower. Brainpower is a numerical value (0.0 to 1.0) that reflects the sentiment strength in an answer converged upon by a swarm. The higher the number, the stronger the sentiment. If the question is a forecast, the Brainpower is an indication of confidence in the forecast. A Brainpower of 0.0 is the case where the swarm is so conflicted that it could not reach an answer within the allotted time. Typically, the maximum allotted time for an artificial swarm to converge is 60 seconds, as beyond that time there is such low conviction that it's better to reframe the question and ask again. In general, swarms converge within the allotted time with a Brainpower between 0.65 and 0.95.

In this way, the output from an ASI system includes not only the final solution converged upon by the swarm, but also includes Brainpower, a metric that allows for relative comparison across a series of questions and/or across a variety of unique swarms, indicating the comparative conviction supporting the final sentiments. Because an ASI is a form of intelligence and not a form of statistical aggregation, it is best to think of the Brainpower as a measure of how confident the emergent swarm intelligence is in the solution reached. Just like an individual can make a decision and have mixed feelings, so can a swarm intelligence, as represented by the Brainpower score.

For AB-comparison questions, an even more accurate metric of a swarm's confidence can be calculated, called Conviction. It uses a dense Neural Network to process the deliberations and can accurately and rigorously quantify the sentiment strength expressed in each solution. By correlating the behavioral data of thousands of previous swarms with known sentiments, the network can quantify the Conviction of any swarm with a degree of statistical certainty, based not just on the answer reached, but on the complex behaviors that led to the answer. Read more about Conviction Analysis at: http://unanimous.ai/conviction.

As one example of the value of this more rigorous approach, Unanimous AI was asked to assess the perceived trustworthiness of major news sources in the US. To address this, a Swarm AI system was assembled that connected 50 voting age Americans, controlling for political affiliation to ensure that the number of Democrats, Republicans, and Independents approximately matched the national average.

The New York Times was chosen as the reference news source, and eleven other news sources were compared to the New York Times in terms of perceived trustworthiness. An example question was formatted as follows, with the Swarm AI system asked to compare the New York Times to CNN, shown in figure 7 below:

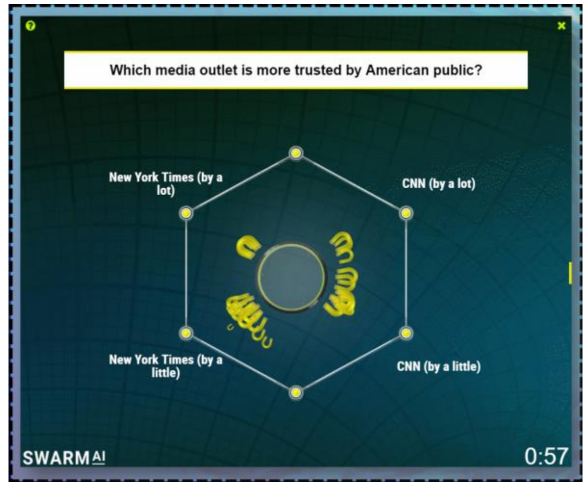

Fig. 7. *Swarm AI system in the process of comparing CNN to the New York Times (view replay: swarm.ai/r/366831)*

The output of this process is a set of comparisons between the New York Times and each of the other media items. Some sources were assessed as more trustworthy than the Times, while others were assessed as less trustworthy. And for each, a conviction index was generated by the Behavioral Neural Network, with confidence interval, enabling the comparison boxplot shown in figure 8 below.

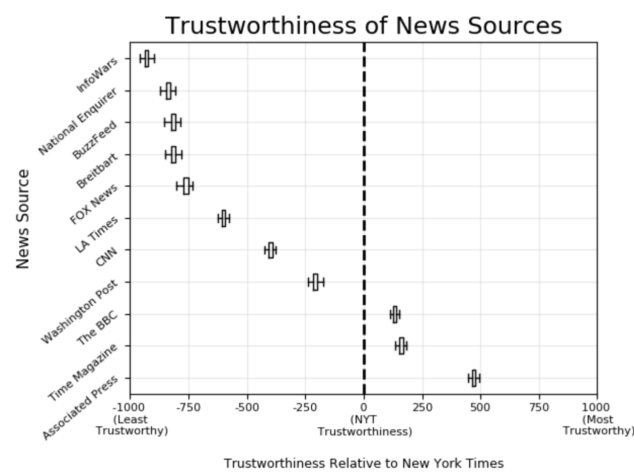

Fig. 8. *Trustworthiness of News Sources Relative to New York Times, as measured by Conviction Analysis.*

In this way, swarms can quickly generate assessments with a rigorous Conviction Index values assigned to each item. While the above example involves comparing the trustworthiness of media items, a similar process can be followed across a wide range of applications, from financial forecasting to business prioritizations.

Following a similar process, Conviction Analysis can enable the comparison of not only one swarm answering multiple different questions, but also the comparison of the Conviction of different populations responding to the same question. This allows the comparison of sentiment of multiple demographic groups: as one example, do Republicans or Democrats perceive CNN as more trustworthy?

Of course, having mixed feelings when making a decision often reflects a complex set of internal tradeoffs and cannot be fully represented in a single metric such as the Brainpower or Conviction values. For example, if asked to choose among 6 flavors of ice-cream and predict the one that will sell best among a target audience, individuals may have strong positive feelings for some options, strong negative feelings for other options, and relative uncertainty or ambivalence about the options between. To gain deeper insights into which options the swarm is conflicted among, we may look at the time varying data itself. To support this, a technique called **Faction Analysis** was developed which enables the decision process to be presented visually for rapid assessment of the group decision dynamics.

VII. FACTION ANALYSIS

If we think of a human swarm as an **intelligent system** that considers a query, explores a set of competing alternatives, and converges on an optimal response, Faction Analysis can be thought of a "brain scan" that shows how the alternatives were debated within the swarm. To provide insight into the complex decision process of a swarming system, a variety of graphing techniques have been developed. Figure 9 below is a snap-shot of a Swarm AI system (comprised of 87 Republican voters) predicting the 2016 US Presidential Primary.

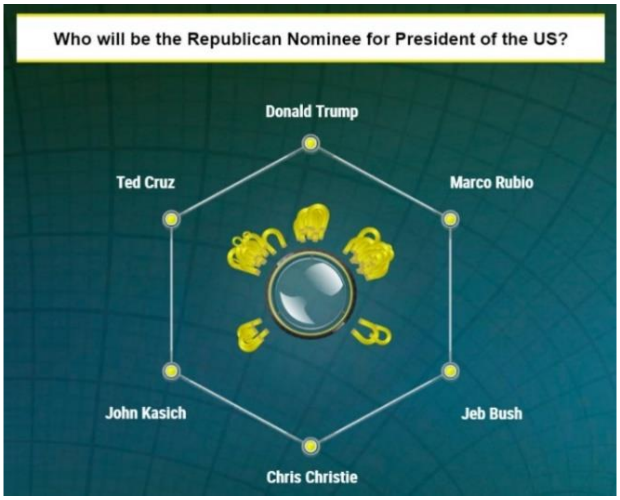

***Fig 9**. Snapshot of Swarm in Progress*
(view replay: swarm.ai/r/45367)

The decision process lasted 20 seconds for this Swarm AI system, during which it explored the decision-space and converged on Donald Trump as a final prediction. In this case, the Brainpower of the swarm was 0.72, reflecting moderate confidence in the forecast outcome. This is a useful result, and turned out to be a very accurate prediction. Still, a researcher may desire more insights into how that result was reached. To provide such insight, the **Faction Force Graph** of Figure 10 below was generated. The plot reflects the change in factional force imparted by participants in support of each candidate across the 20 second decision period.

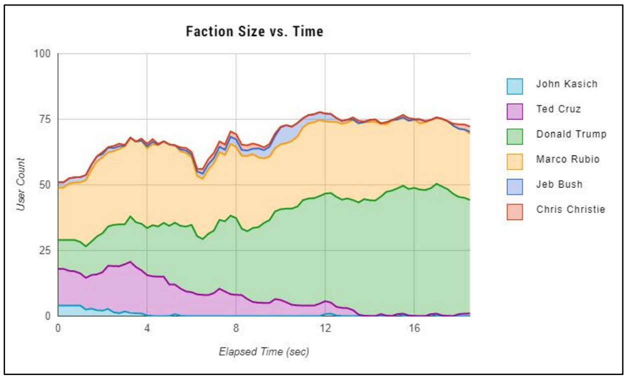

***Fig 10**. Faction Force vs Time Graph*

As shown in Figure 10 above, the Swarm AI system was torn between three principal outcomes – Donald Trump, Marco Rubio, and Ted Cruz. It's important to observe that the most popular option at time t=0 was Marco Rubio, with very low differentiation among the top 3 options. This reflects the initial response provided by participants in isolation, before real-time swarming proceeds. This initial time step is generally aligned with a typical survey and demonstrates the flawed answer a survey would reveal. But a swarm is an intelligent system, enabling participants to converge over 20 seconds on the answer they best agree upon. And in this case, that answer arrived upon proved to be the correct prediction of Donald Trump.

We can dig deeper into the deliberation process, looking not just at the factional support but also how and when the individual participants switched their opinions in order to converge on a solution the swarm could best agree upon. This is reflected in the **Faction Change Graph** shown in Figure 11.

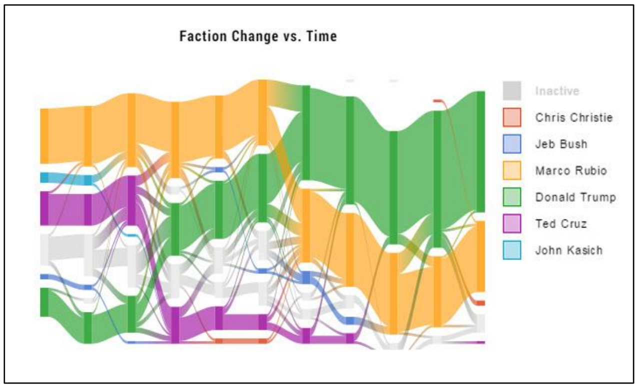

***Fig 11.** Faction Change vs Time*

As reflected in this graph, we can see that while predicting Trump was not the most popular initial choice, or even the largest plurality at the onset of swarming, the participants converged on Trump with high conviction, and that option gained steady support over time, capturing supporters who were conflicted over Cruz and Rubio. This is a powerful insight, because it reveals, for example, that Trump campaign staff could target Cruz and Rubio supports and expect defections, while the inverse would not be expected.

It’s interesting to note that CNN ran a Political Prediction Market to forecast the odds of each candidate becoming the nominee mid-stream in the primary process. They forecast a dead heat for the Republican nominee, with Rubio at 33% odds, Cruz at 30% odds, and Trump at 28% odds [24]. Looking at Figure 8 above, the CNN prediction is roughly what we would expect from a traditional polling system, corresponding to the initial sentiment distribution at t=0 seconds on the chart. The swarm-based prediction on the other hand, as represented by t=20 seconds on the chart, shows that with real-time feedback, the system converges on a far more insightful result –Trump.

VIII. Interpolation

Converging on precise numerical estimations is a common use case for business teams using Swarm, from predicting the odds of a business outcome to forecasting the change in value of a financial asset. While swarm-based forecasts have been shown to be significantly more accurate than other methods, researchers at Unanimous AI have recently developed a technique for post-processing swarm-based data that makes the output even more accurate. Known as “Force Density Interpolation,” this software feature processes the full range of behaviors captured during a real-time swarm and interpolates to compute a more precise final value.

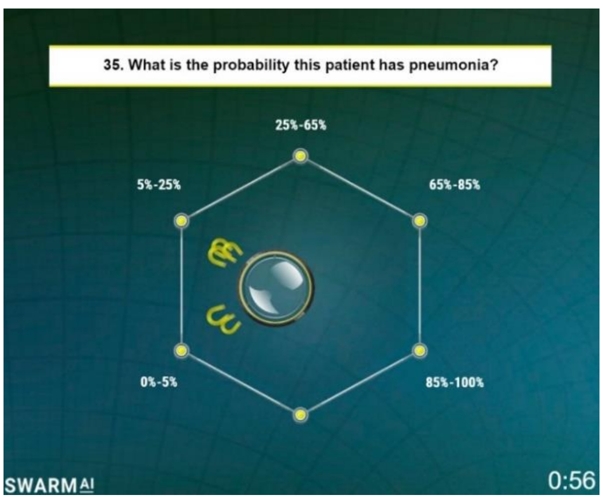

Fig. 12. *ASI in the process of Diagnosing Pneumonia (view replay: swarm.ai/r/371381)*

This process is best described by example. Let’s consider a recent study conducted by Stanford University [3,4] in which small groups of radiologists were tasked with diagnosing chest-X-rays for the presence of pneumonia. In this study, the swarms significantly outperformed the individual doctors, reducing errors by over 30%. Figure 12 below shows a screenshot of a swarm in the process of selecting the coarse range of probabilities for a chest X-ray that was displayed to all members. It’s important to note that this is a snapshot as the collaboratively controlled puck moves across the decision-space and converges upon an answer. The full process of deliberation, as moderated by the real-time swarm intelligence algorithms, generally takes between 15 and 60 seconds. In the example shown above, the swarm converged on an answer (5-25%) within 16 seconds. Note, the two groups used different layouts of probability bins.

While the final probability selected by the swarm is a good first estimate for the chosen probability of pneumonia, we can use the underlying data generated by the swarm as they converged upon their answer to refine this value. This is done using a weighted averaging process referred to as *Squared Impulse Interpolation*. This process, as outlined in Equations 1 and 2, calculates a weighted average of the probabilities in the swarm using the squared net “pull” towards each answer as weights. The pull is represented as the force (F) imparted by members of the swarm and the weight for each answer w_i is calculated as the squared impulse towards that answer (equation 1). The weighted average over the answer choice values v_i is then computed (equation 2). The answer choice values v_i are taken as the midpoint of each bin. For example, the bin “0%-5%” has a midpoint v_i of 2.5%.

$$w_i = \frac{F(i)^2}{\sum_{a \in Answers} F(a)^2} \quad (1)$$

$$Interpolation = \sum w_i v_i \quad (2)$$

This process can be visualized in Swarm by plotting the net vector force of each radiologist over the course of the swarm, as shown in Figure 13. In this Force Density Visualization, the puck’s trajectory is shown as a white dotted line, and the distribution of Force is plotted as a Gaussian Kernel Density heatmap. Notice that the swarm was split between the “5%-25%” and “0%-5%” bins, and more force was directed towards the 5%-25%. This aggregate behavior is reflected in the swarm’s Interpolated Diagnosis of 11.1%.

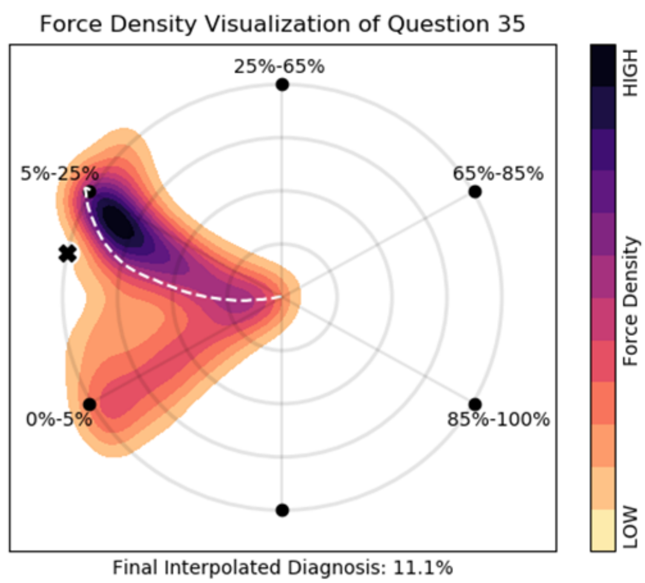

Fig. 13. *Force Density Interpolation*

IX. Scaled Rankings

Ranking a set of outcomes or prioritizing a set of alternatives is a common use case of Swarm AI systems. Swarm-based rankings have been shown to be significantly more representative of group priorities than voting [30]. While ordinal rankings (1st, 2nd, 3rd…) are useful, it's often also valuable to quantify the difference in ranking between two alternatives to understand the relative difference in sentiment between those alternatives.

To accomplish this, the behavioral data from swarms can be used to create Scaled Rankings that represent the relative sentiment of each answer choice in a prioritization or across a set of comparison questions. Scaled rankings are cardinal rankings that take values between 1 and n, where n is the number of possible outcomes (1, 1.8, 3.2 …. n).

To illustrate this more clearly, we can consider a case where Unanimous was challenged to rank the most likely winners of the FIFA World Cup. To do so, a team of football enthusiasts from around the world was assembled into a Swarm AI and generated an ordinal ranking from least likely (1) to most likely (16) to win the championship.

The behavior of the swarm in the ranking process was then used to generate Scaled Rankings, from 1 (least likely) to 16 (most likely), as shown below in Figure 14. The Scaled Ranking of the teams are shown on the x-axis, while the given Ordinal Ranking of the teams are shown on the y-axis. The data is significantly more meaningful when represented with Scaled Rankings, as the differences between each team's performance level is clearer. In fact, only in the Scaled Ranking representation do five distinct clusters emerge in team performance (highlighted in red): Moonshot winners are significantly lower ranked than the Likely Upsets, and so on, onto the two most likely winners.

In this example, Brainpower was used to generate the scaled rankings, but Conviction may be used instead on appropriate AB questions, similar to the News Trustworthiness Analysis in figure 8 above.

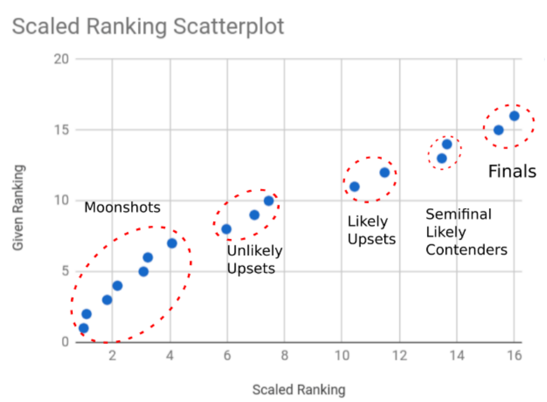

Fig. 14. *Scaled Ranking Scatterplot for 2018 FIFA Predictions, with clusters highlighted in red.*

X. Swarm Size

What is the right size of a swarm? In many ways, that is similar to asking *how many neurons* it takes to build an effective brain. Clearly it depends upon the task at hand and the features of the neurons themselves. And even with those caveats, science does not yet have an answer. For swarms, it's not any more definitive. After all, a biological swarm is essentially a "brain of brains" – a networks of high-level processors that work together as a unified system. As mentioned above, among the most studied example in nature are honeybee swarms, which have been researched since the 1950's. When working together as a system, optimized answers to complex problems emerge when bee swarms form among groups of 200 to 400 members. Thus, although bee colonies are 10,000 members in size, millions of years of evolution have produced swarming behaviors among just 2 to 4% of the overall colony population. This natural example gives a potential starting point for sizing human swarms.

Still, it would be helpful to have additional guidance as to the effective size of swarms and ground our thinking with respect to traditional polls and surveys. To address this issue, researchers at Unanimous A.I. and Oxford University [25] performed comparative studies of **polls vs swarms,** testing the accuracy of group decisions and predictions. For example, in one recent study, researchers compared a poll of 469 football fans with a swarm of 29 football fans in a challenge to predict 20 Prop Bets during the 2016 Super Bowl. Results revealed that the poll results, although based on 16 times the number of participants, was significantly less accurate (at **47% correct**) than the swarm (at **68% correct**). This represents a significant amplification of intelligence resulting from swarming.

Figure 15 below shows the swarm's performance in this particular study as compared to the statistical distribution of poll participants. The swarm's performance is represented by the red line on the graph. The x-axis represents how many questions the individuals answered correctly. The y-axis represents how many people correctly answered that particular number of questions from the 469-person sample.

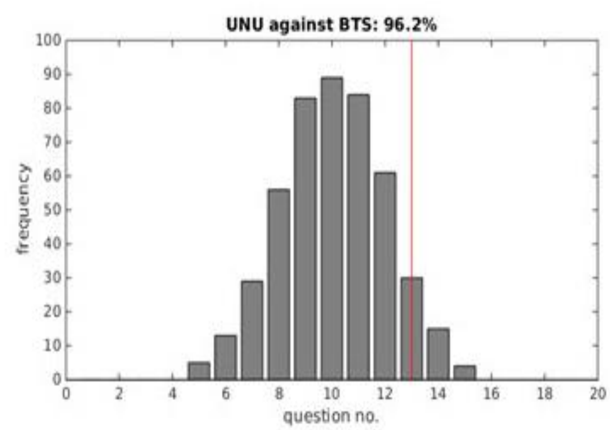

Fig 15. *Swarm Performance vs Group Members*

As indicated by the red line, the 29-person swarm outperformed the individuals in the much larger poll by 2 standard deviations (Z=1.99). In fact, the swarming process generated forecasts that outperformed 96.2% of the individual members, which suggests that swarming significantly amplified the intelligence of the individuals.

Similar results have been shown in swarms as small as 4 people [6], to swarms as large as 30 people [29], so swarms can amplify the accuracy of both small and large groups. These and other results support the view that swarming, with closed-loop feedback, is a far more efficient method for harnessing group insights than polling, even when polls target significantly larger populations.

XI. Swarm Repeatability

When conducting swarms, the question of 'sample size' comes up, meaning how many participants are needed in a swarm to get statistically significant results. The people who ask this question are generally familiar with conducting polls and surveys, where sample size determines the extensibility of the results to the population at large.

When comparing surveys and swarms, however, the idea of repeatability is often misleading: when a system is repeatable or "statistically significant" for a sample size, that does not mean that the answer produced is accurate, but rather that asking the question again would give the same results. In other words, reaching the wrong answer consistently is not the hallmark of an accurate system.

One example of the distinction between repeatability and accuracy can be seen in a recent study conducted by researchers at California Polytechnic and Unanimous AI, in which 283 individuals took a social sensitivity test first alone, and then as a swarm of 3-6 people (4 on average), for a total of 66 swarms [31].

The repeatability and accuracy of the Surveys and Swarms on one question of this test are compared in figure 16. The repeatability of each survey method on this question was calculated as the frequency that a random resample of responses would reach the most popular answer. The repeatability of the swarm was calculated as the frequency with which swarms chose the most popular answer. In addition, the accuracy and repeatability of an Aggregation of Swarms was evaluated as the result of a 'vote' between three swarms' final answers.

The large survey of 283 users was far more repeatable (98.6%) than the small swarms (62%). However, because most individuals answered incorrectly, a survey of 283 was correct only 1.4% of the time. Meanwhile, the swarms, comprised of the same individuals that voted mostly for the wrong answer, achieved a 62% accuracy on this question.

Clearly the repeatability of a system is primarily a function of the sample size of that system, which does not necessarily indicate accuracy. By connecting individuals in real-time Swarms, groups converge more frequently on the correct answer, even if the correct answer is chosen by the minority of constituents when surveyed.

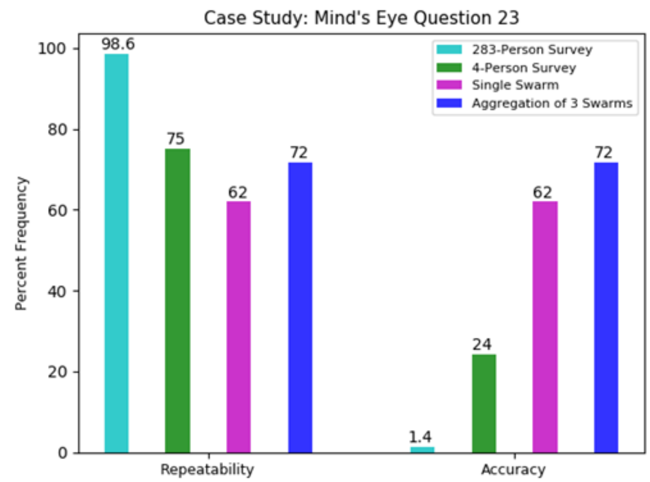

Fig 16. *Repeatability and Accuracy comparison of Swarms vs Surveys on Question 23 of the Mind in the Eyes test.*

Other than this singular question, we find that over the whole social sensitivity test, the swarm was slightly more repeatable than similarly-sized surveys, while also being significantly more accurate. As shown in Figures 17 and 18, an aggregation of 5 swarms of 4 users each yields similarly repeatable results to a survey of the same size (20 people), while at the same time providing more accurate results than a 283-person survey, more than 10 times the sample size of the swarming population.

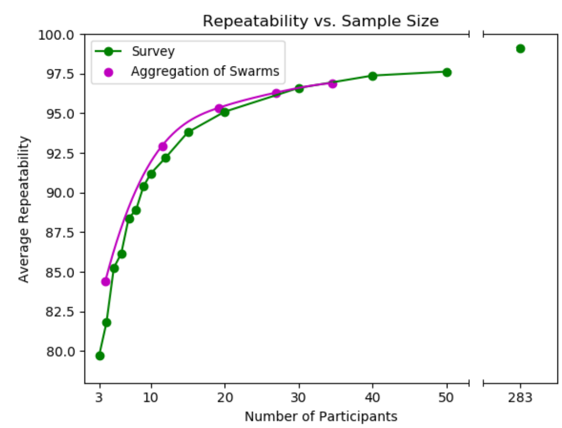

Fig 17. *Repeatability of Mind in the Eyes Swarms vs Surveys*

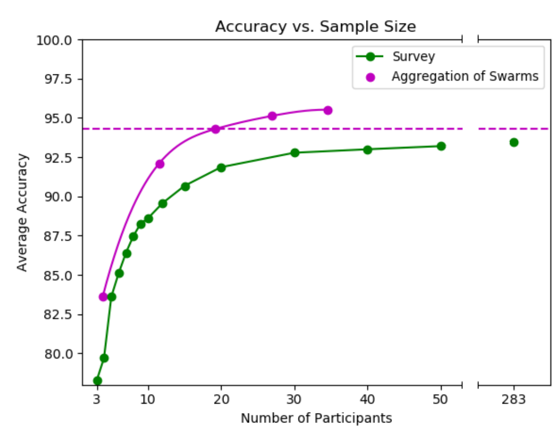

Fig 18. *Accuracy of Mind in the Eyes Swarms vs Surveys*

We can also study repeatability of human swarms in a different way, by testing whether human swarms can consistently outperform individuals. To investigate this, a study was recently conducted by researchers at Unanimous A.I. and Oxford University, testing the ability of a swarm to predict English Premier League (EPL) football matches over a period of 5 consecutive weeks.

			INDIVIDUALS	SWARM
	# Games	Group Size	%Correct	%Correct
Week 1	10	28	42%	60%
Week 2	10	31	60%	80%
Week 3	10	31	58%	80%
Week 4	10	25	59%	60%
Week 5	10	31	55%	80%
MEAN		29	**55%**	**72%**
StDev		2.7	7%	11%

Table 1. *Summary of prediction results over 5 weeks.*

For each of the 5 weeks of the study, predictions were made for the full slate of 10 matches played. This means the swarm and individuals predicted, in total, the outcome of 50 professional soccer matches. These results are shown in Table 1 below, revealing that the swarm outperformed the individuals, week after week, with an average amplification of intelligence, across the full five weeks, equal to 131%.

To assess statistical significance, swarm performance was compared to the performance expected by chance from a matching population using a bootstrap approach as follows: each week, researchers took a random sample of 10 individuals who participated in that week's trial and took the first individual's prediction for the first match, the second individual's prediction for the second match and so on until ten predictions from the ten randomly selected individuals were generated. Researchers then averaged the accuracy of these predictions Repeating the procedure (i.e., random selection of ten individuals and response assignment) 10,000 times, researchers then computed the average distribution of correct answers for that week.

Distributions are shown in Figure 19 below. The mean of the distribution represents the average number of correct picks that should be expected by chance, by matching forecasters population. As shown, the swarms are well above the mean as compared to individual predictions. Researchers then computed the distance of the swarm performance for each week from that week's mean in the form of a z-score distance and computed the value of the cumulative density function of a normal distribution with that mean and standard deviation. The value indicates the probability of obtaining the score by chance.

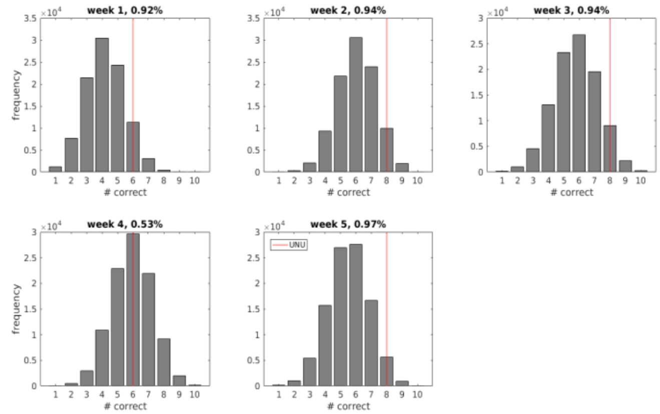

Fig. 19*. Individual vs Swarm predictions, assessed weekly.*

To aggregate the results from the five weeks into one, researchers compared the overall number of hits (i.e. successful predictions) made by the swarm in the 5 weeks and the number of hits made by the average individual (rounded to the closest integer). We then used a two-proportion z-test, with the null hypothesis that the two hit rates are the same. A z-statistic was obtained using the following formula:

$$z=(pIND - pSWARM) / \ sqrt(p*(1-p)*(2/50))$$

where pIND is the hit rate of the average individual, pSWARM is the hit rate of the swarm and p is the total sum of hits made by both the average individual and the swarm and divided by the total number of predictions (i.e., 100). The results show that the average individual was significantly worse than the unified swarm intelligence (z=-1.78, p=.03). The aggregated results can be shown in a single profile, as depicted in Figure 20. The red line indicates the superior performance of the swarm as compared to individuals.

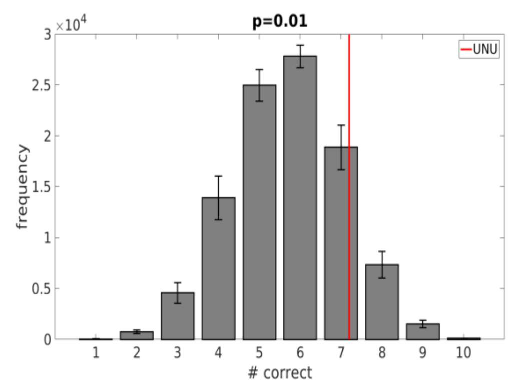

Fig. 20*. Individual vs Swarm predictions aggregated over 5 weeks*

XII. POSITIONAL BIAS

When using human swarms to select among a set of spatially arranged options, a question sometimes arises about whether the spatial placement of the options influences the outcome. The answer is no - not at a significant level in the vast majority of situations. The reason is twofold:

First, a swarm is not a vote or poll, but an interactive system that explores the set of options and converges on the most agreeable solution. The path taken by the swarm will vary based on spatial layout of the options, but the result converged upon should be independent of layout in most cases. This is true of both of human and natural swarms. In fact, this effect has been studied in honeybees and is believed to be one of the primary evolutionary benefits of swarming. For example, when selecting a new home site, bees explore a 30-square mile area and consider dozens of potential options, some of which are discovered long before others. Biologists have shown that the swarming process enables bees to convergence on the optimal solution, independent of the order of discovery [16]. The evolutionary need is obvious: if honeybees converged on home sites that were not optimal simply because they were considered first, their chances for survival would be greatly diminished. Fortunately, a swarm intelligence is an adaptive system that is robust in combating ordering effects.

It's worth noting that we humans have a lot to gain by using synchronous swarms, as many common methods for asynchronous polling are not only deeply susceptible to biasing based on the order in which options are evaluated, their results become meaningless because of it. For example, online forums such as Reddit allow popular content to rise and fall with sequential up-voting and down-voting. Similarly, online *prediction markets* allow commoditized content to rise and fall with sequential buys and sells. Research studies show that sequential polling can greatly distort outcomes by introducing social biasing effects (often referred to as *herding* or *snowballing*). One well-known study [12] found that a single up-vote, when inserted first into an online sequential polling system like Reddit, influenced the final decision of the group by more than 25%. Similarly, prediction markets suffer from momentum effects, price bubbles, risk-aversion biases, and over-corrections as a consequence of asynchrony [27].

Second, the underlying routines that govern human swarms include unique algorithms designed specifically to minimize the cases where positional layout could impact outcome. Specifically, every 250 milliseconds, the algorithms assess whether or not the spatially arranged options are on the same side of the puck as each other, and thus could influence each other constructively, or if they are opposite sides of the puck, and therefore can influence each other destructively. In this way, neighboring factions do not aggregate support as a result of their close proximity, and distant factions do not diminish support as a result of their opposing locations. Such mathematics ensure that the puck can explore the decision-space without significant positional biases and enable the natural swarming process to proceed spatially independent.

XIII. ANSWER OPTIONS

As currently implemented by the Swarm AI platform, questions can be provided either as a continuous range across a number line or as a hexagon of up to six discrete selections (as in Fig 9). Researchers new to swarming often ask about the use of the hexagon and its limitation of only supporting up to six simultaneous options. This is a deliberate limitation based on historic social-science research indicating that most human participants are inefficient decision-makers when presented with more than six simultaneous options. Such "choice overload" causes people to become overwhelmed by larger option sets and make poor decisions, and even lose interest in the decision process itself [28]. To enable swarms to consider larger sets of alternatives, the Swarm AI 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. This allows a final answer to emerge from a large set of options. Using this iterative process, methodologies can be designed for any number of choices, presenting the alternatives in a manner that avoids "choice overload."

XIV. FINAL GUIDIANCE

Artificial Swarm Intelligence is a unique and powerful method for tapping the knowledge, wisdom, intuition, and insights of human populations, enabling optimized solutions to rapidly emerge. While many are tempted to compare the process to traditional polls, surveys, and focus groups, the relationship is tenuous at best. Yes, all of these methods collect input from human participants, but polls, surveys, and focus groups treat people as "respondents" – i.e. as a source of isolated data points that are added to a growing dataset. Because such methods are statistical constructs, their validity is based entirely on simple statistical tests. And even then, a statistically significant poll does not mean that the poll is providing researchers with accurate insights – it just means that repeating the poll on a similar population will yield the same answers, accurate or not.

Artificial Swarm Intelligence systems, on the other hand, treat people as "participants" and task them with being active "data processors" rather than passive data points. This enables populations to form real-time systems that converge on optimal solutions. While polls, surveys, and focus groups can indicate which option among a set of options might be most popular to individuals, in isolation, they give little insight into which options the population would best agree upon "in the wild." Because most marketing activities are about influencing populations in real-world contexts, not polling individuals in isolation, using Artificial Swarm Intelligence to reveal how groups are most likely to converge as natural systems is a far more effective technique.

REFERENCES

[1] Rosenberg, Louis. Artificial Swarm Intelligence, a human-in-the-loop approach to A.I., Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI 2016), Phoenix, Arizona. Pg 4381-4382. AAAI Press, 2016.

[2] Rosenberg, Louis, "Human Swarms, a real-time method for collective intelligence." Proceedings of the European Conference

on Artificial Life 2015, pp. 658-659 ECAL 2015, York, UK. MIT Press 2015, ISBN 978-0-262-33027-5

[3] Halabi, Safwan., et. Al. “Radiology SWARM: Novel Crowdsourcing Tool for CheXNet Algorithm Validation”, SiiM Conference on Machine Intelligence in Medical Imaging, 2018.

[4] Rosenberg, L, Willcox, G., Halabi, S., Lungren, M, Baltaxe, D. and Lyons, M. "Artificial Swarm Intelligence employed to Amplify Diagnostic Accuracy in Radiology,” 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018.

[5] Befort, K., Baltaxe, D., Proffitt, C., & Durbin, D. (2018). “Artificial Swarm Intelligence Technology Enables Better Subjective Rating Judgment in Pilots Compared to Traditional Data Collection Methods.” Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62(1), 2033–2036.

[6] Askay, D., Metcalf, L., Rosenberg, L., Willcox, D., “Enhancing Group Social Perceptiveness through a Swarm-based Decision-Making Platform.” Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS-52), IEEE 2019.

[7] Rosenberg, Louis. Pescetelli, Nicollo, Willcox, Gregg., "Artificial Swarm Intelligence amplifies accuracy when predicting financial markets," 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York City, NY, 2017, pp. 58-62.

[8] Unanimous NBA “Dense Neural Network used to Amplify the Forecasting Accuracy of real-time Human Swarms.”

[9] Marshall, James. Bogacz, Rafal. Dornhaus, Anna. Planqué, Robert. Kovacs, Tim. Franks, Nigel. “On optimal decision-making in brains and social insect colonies.” Soc. Interface 2009.

[10] Galton F, (1907) Vox populi. Nature 75:7.

[11] Lorge I, Fox D, Davitz J, Brenner M (1958) A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. Psychol Bull 55:337–372.

[12] Lev Muchnik, Sinan Aral, Sean J. Taylor. Social Influence Bias: A Randomized Experiment. Science, 9 August 2013: Vol. 341 no. 6146 pp. 647-651

[13] Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. Proc Natl Acad Sci USA 108(22):9020–9025.

[14] Seeley T.D, Buhrman S.C 2001 “Nest-site selection in honey bees: how well do swarms implement the ‘best-of-N’ decision rule?” Behav. Ecol. Sociobiol. 49, 416–427

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

[16] Seeley, Thomas D. Honeybee Democracy. Princeton Univ. Press, 2010.

[17] 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.

[18] Usher, M. McClelland J.L 2001 “The time course of perceptual choice: the leaky, competing accumulator model.” Psychol. Rev. 108, 550–592

[19] Ungar L, Mellors B, Satopää V, Baron J, Tetlock P, Ramos J, et al. The Good Judgment Project: A Large Scale Test. AAAI Technical Report. 2012; FS-12-06.

[20] Björkman, M, Juslin, P., and Winman, A. (1993). Realism of confidence in sensory discrimination. Perception & Psychology, 55, 412-428.

[21] Ranjan, R. and Gneiting, T. (2010), Combining probability forecasts. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72: 71–91.

[22] Ariely, D., Au, W. T., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T. S., Zauberman, G. (2000). The Effects of Averaging Subjective Probability Estimates Between and Within Judges. Journal of Experimental Psychology: Applied, 6(2), 130-147.

[23] Ferrell, W. R. (1994). Discrete subjective probabilities and decision analysis: Elicitation, calibration and combination. In G. Wright & P. Ayton (Eds.), Subjective probability (pp. 411-451). England: Wiley.

[24] CNN Publication: http://www.cnn.com/specials/politics/political-prediction-market-debate-sweepstakes

[25] Rosenberg, Louis. Baltaxe, David and Pescetelli, Nicollo. "Crowds vs Swarms, a Comparison of Intelligence," 2016 Swarm/Human Blended Intelligence (SHBI), Cleveland, OH, 2016, pp. 1-4.

[26] 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. Behavioral Ecology and Sociobiology 54 (5) 511-520.

[27] Ottaviani, Marco, and Peter Norman Sørensen. "Aggregation of information and beliefs in prediction markets." London Conference on Information and Prediction Markets. 2007.

[28] 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.

[29] Rosenberg, Louis and Willcox, Gregg. "Artificial Swarm Intelligence cs Vegas Betting Markets," 2018 11th International Conference on Developments in eSystems Engineering (DeSE), Cambridge, 2018, pp. 155-159.

[30] L. Rosenberg and D. Baltaxe, "Setting group priorities — Swarms vs votes," 2016 Swarm/Human Blended Intelligence Workshop (SHBI), Cleveland, OH, 2016, pp. 1-4.

[31] "Repeatability and Accuracy of Swarms and Surveys." Unanimous AI. Accessed January 14, 2019. https://unanimous.ai/wp-content/uploads/2019/01/Repeatability-Document.pdf.

* **Swarm AI** is a registered trademark of Unanimous AI