Amplifying Prediction Accuracy using Swarm A.I.

Louis Rosenberg
Unanimous A.I.
San Francisco, CA, USA
Louis@Unanimous.ai

Niccolo Pescetelli
University of Oxford
Clarendon, UK
niccolo.pescetelli@chch.ox.ac.uk

***Abstract*— In the natural world, many species amplify the accuracy of their decision-making abilities by working together real-time closed-loop systems that converge on optimal solutions in synchrony. Known as Swarm Intelligence (SI), the process has been deeply studied in schools of fish, flocks of birds, and swarms of bees. The present study looks at the ability of human groups to make decisions as an Artificial Swarm Intelligence (ASI) by forming similar real-time closed-loop systems online. More specifically, the present study tasked groups of sports fans with predicting English Premier League matches over a period of five weeks by working together in real-time swarm-based systems. Results showed that individuals who averaged 55% accuracy when working alone were able to amplify their accuracy to 72% by forming real-time swarms. This corresponds to 131% amplification in predictive accuracy across five consecutive weeks (50 games).**

Keywords— Swarm Intelligence, Artificial Swarm Intelligence, Collective Intelligence, Human Swarming, Artificial Intelligence.

I. INTRODUCTION

Artificial Swarm Intelligence (ASI) strives to amplify the collective wisdom of human groups by connecting participants online into real-time closed-loop systems that are modeled after biological swarms. Prior studies have shown that such "human swarms" can produce significantly more accurate predictions than traditional methods for tapping the collective intelligence of groups, such as votes, polls, surveys, and markets. For example, one recent study tested the ability of human swarms to forecast the outcome of College Bowl football games (in the U.S.) against the Las Vegas spread. A swarm was comprised of 75 amateur football fans was tasked with predicting each of 10 college bowl games. As individuals, the participants averaged 50% correct (i.e. coin flip accuracy). But, when working together as a real-time swarm, those same participants achieved 70% accuracy against the spread. Not only is this a significant accuracy increase, it also enabled the 75 amateur football fans to out-predict the football experts at ESPN [1].

While prior studies have documented the ability of Artificial Swarm Intelligence to amplify the predictive ability of online groups in singular events, no long-term study has been performed to assess consistency of swarm-based predictions over time. To address this, the present study tasked human swarms with predicting all of the scheduled English Premier League (EPL) matches over a period of five weeks in 2016. The objective was to assess whether or not a statistically significant amplification of human intelligence could be measured when comparing individual prediction accuracy to swarm accuracy. In addition, swarm performance over the five week period was compared to the predictions made by the Sports Analytics Machine (SAM), a super-computer built by the University of Salford to predict English Premier League games using rigorous mathematical models [2]. Because SAM results are published weekly by the BBC to reflect an "expert" assessment of weekly matches, this allowed for comparison of professional level predictions with novice-based human swarms.

II. SWARMS AS INTELLIGENT SYSTEMS

The decision-making processes in honeybee swarms have been observed to be remarkably similar to the decision-making processes in neurological brains [3,4]. 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. When one sub-population exceeds a threshold level of support, the corresponding alternative is chosen. In honeybees, this enables optimal decisions over 80% of the time [5,6,7]. It is this amplification of intelligence that Artificial Swarm Intelligence aims to enable among distributed networked humans.

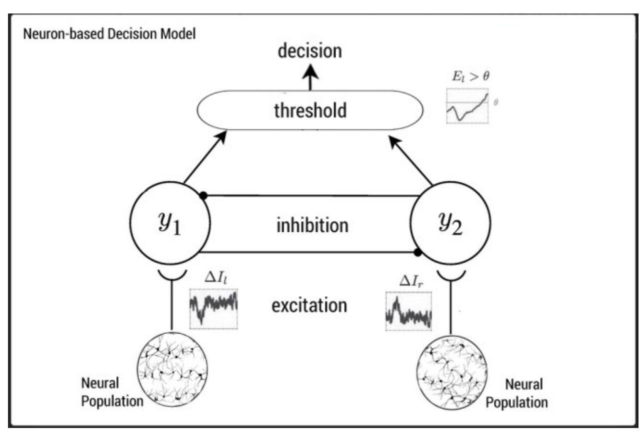

Fig. 1. Usher-McClelland model of neurological decision-making

The similarity between neurological intelligence and swarm intelligence becomes even more apparent when comparing decision-making models that represent each. For example, the decision-making process in primate brains is often modeled as mutually inhibitory leaky integrators that aggregate incoming evidence from competing neural populations. A common framework is the Usher-McClelland model [8] represented in Figure 1 above. This can be directly compared to swarm-based decision models, like the honey-bee model in Figure 2 below. As shown, these swarm-based decisions follows a very similar process, aggregating incoming evidence from sub-populations of swarm members through mutual excitation and inhibition.

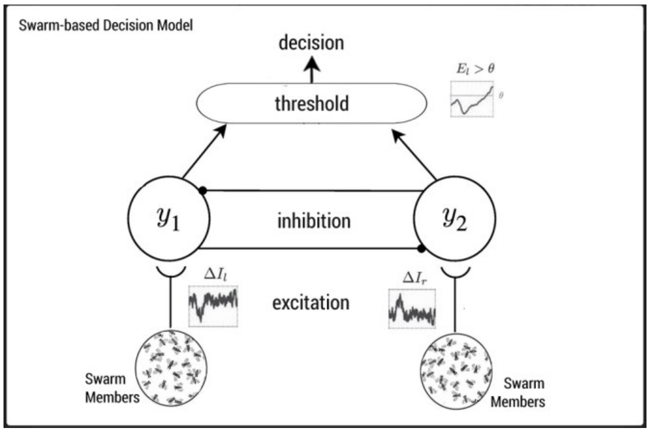

Fig. 2. Mutually inhibitory decision-making model in bee swarms

III. ENABLING "HUMAN SWARMS"

Unlike many other social species, humans have not evolved the natural ability to form a closed-loop Swarm Intelligence. That's because we lack the subtle connections that other organisms use to establish tight-knit feedback-loops among members. Schooling fish detect vibrations in the water around them. Flocking birds detect motions propagating through the group. Swarming bees use complex body vibrations called a "Waggle Dance". Thus to enable a real-time Artificial Swarm Intelligence among groups of networked humans, specialized technology is required to close the loop among members.

To address this need, an online platform called **UNU** was developed in 2015 to allow distributed groups of users to login from anywhere around the world and participate in a closed loop swarming process [9]. Modeled after the decision-making of honeybee swarms, the UNU platform 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, while also allowing all participants to perceive and react to the changing system in real-time, thereby closing a feedback loop around the full population of participants.

As shown in Figure 3, participants in the UNU platform answer questions by collectively moving a graphical puck to select among a set of alternatives. Each participant provides input by manipulating a graphical magnet with a mouse or touchscreen. By positioning their magnet, users impart their personal intent 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 population of users can adjust their intent at every time-step (200 ms), the puck moves, not based on the input of any individual, but based on the dynamics of the full system. This enables real-time physical negotiation among all members, empowering the group to collectively explore the decision-space and converge on the most agreeable solution in synchrony.

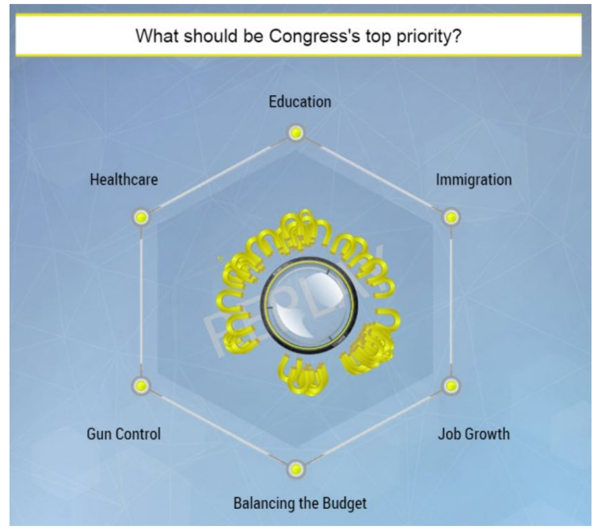

Fig. 3. A human swarm answering a question in real-time

It is important to note that participants do not simply vary the direction of their input, but also modulate the magnitude of their input by adjusting the distance between the magnet and the puck. Because the puck is in continuous motion across the decision-space, in order to apply force users need to continually move their magnet so that it stays close to the puck's rim. This is significant, for it requires participants to be engaged continuously during the decision process, evaluating and re-evaluating their contribution. If they stop adjusting their magnet to the changing position of puck, the distance grows and their applied force wanes. Thus, like bees vibrating their bodies to express sentiment in a biological swarm or neurons firing activation signals to express sentiment in a neural-network, the participants in an artificial swarm must continuously express their changing preferences during the decision process, or lose their influence over the collective outcome.

IV. PREDICTION STUDY

To assess the predictive ability of human swarms over an extended period, a formal study was conducted over a five week period using groups of randomly selected human subjects from a pool of individuals who self-reported being enthusiasts of EPL football. Each weekly group consisted of 25 to 31 participants who engaged the experiment via online access to the UNU swarming platform. Each subject was paid $2.50 for their participation in each weekly session, which required them to make predictions for the outcome of all 10 English Premier League matches being played that week, first as individuals on a standard online survey, and then as part of a real-time Artificial Swarm Intelligence comprised of the full weekly group. In addition, the researchers compared results to the predictions made by SAM, a sports super-computer at the

University of Salford which uses ten years of data and sophisticated algorithms to predict EPL games.

Across the full five week period, predictions were made for a total of 50 games wherein the participants were required to forecast one of three outcomes for each game: (i) Team A wins the match, (ii) Team B wins the match, or (iii) the match ends in a tie. It is worth noting that tie games occur at a rate of approximately 25% in EPL matches, making it a significant outcome possibility. It is also worth noting that 94% of the swarm participants were American citizens for whom EPL is a foreign sport covered mostly by foreign media. This context is relevant when comparing performance of the human swarm to the performance of the SAM super-computer, which is a UK-based analytical system designed specifically to predict EPL outcomes. In other words, it allows us to test if groups of American fans, working together as artificial swarms, can produce comparable results to a rigorous computational model that is used by the BBC to forecast the UK's national sport.

In Figure 2 below, a snapshot of a human swarm comprised of 31 participants is shown in the process of predicting a match between Arsenal and Watford. As shown in the figure, the swarm is given five options to choose among, enabling the swarm to identify which of the two teams will win and whether the winning team will prevail by a single goal ("by 1"), by 2 or more goals ("by 2+"), or if the swarm believes the match is too close to call. In the example shown below, a large majority of participants have already shifted their pull towards Arsenal, and so the puck is currently heading in that direction.

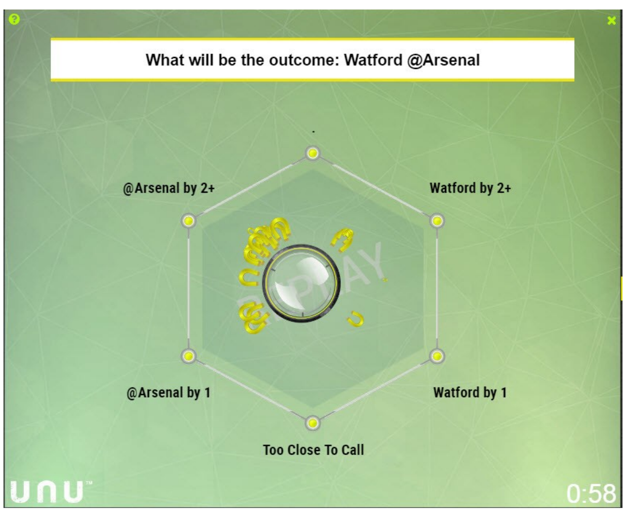

Fig. 4. A human swarm predicting an EPL match in real-time

If the swarm converges on an answer that indicates one of the two teams will win, that is selected as the predicted outcome for the given match. If, on the other hand, the swarm converges on "too close to call," the swarm is given a second question asking if the predicted outcome is most likely a tie. In the example shown in Figure 4 above, the artificial swarm demonstrated strong conviction that Arsenal would beat Watford by a wide margin. In Figure 5, a series of snapshots demonstrate how the swarm converged upon this final answer over time. It should be noted that all predictions made by the swarm were converged upon in under 60 seconds.

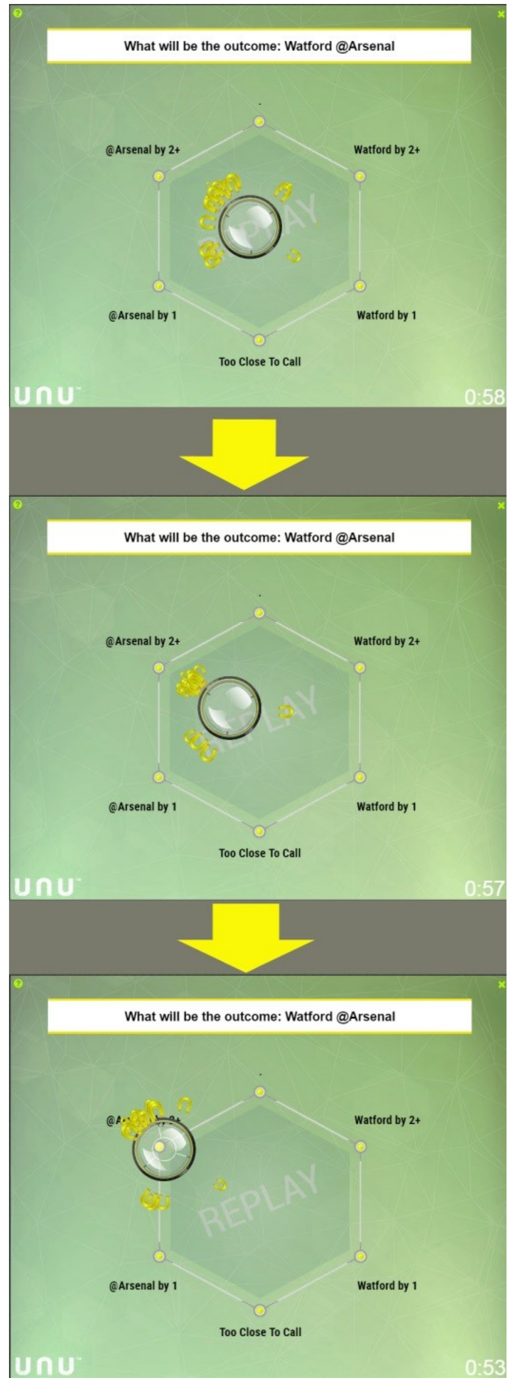

Fig. 5. A time-series of swarm converging on a final prediction

V. RESULTS

For each of the five weeks of the testing period, predictions were made for the full slate of 10 matches that were played by English Premier League teams. For each set of 10 matches, a group of participants provided their individual predictions via a private online survey. The group also logged into the UNU platform for real-time swarming and made predictions by working together as a unified swarm. In addition, data was collected from the BBC indicating the predictions made by the SAM super-computer for the same games. After the games were played, the results were scored by computing the number of correct predictions and the percentage of correct predictions for each test case. For individuals, the average values were computed across the 25 to 31 participants in each group. These results are shown in Table 1 below:

			INDIVIDUALS	SWARM	SAM
	# Games	Group Size	%Correct	%Correct	%Correct
Week 1	10	28	42%	60%	40%
Week 2	10	31	60%	80%	70%
Week 3	10	31	58%	80%	90%
Week 4	10	25	59%	60%	50%
Week 5	10	31	55%	80%	70%
MEAN		29	**55%**	**72%**	**64%**
StDev		2.7	7%	11%	19%

Table 1. Summary of prediction results over 5 weeks.

Assessing the raw results, we see that the swarm had the best performance of the three experimental cases tested, achieving 72% accuracy when predicting English Premier League games. This was significantly more accurate than the same individuals, when predicting independently, as they averaged only 55% accuracy across each group. And finally, the analytical super-computer, SAM, achieved a result in the middle of these two cases, generating 64% accuracy.

Thus, at a first level of analysis we see that by working together as a swarm, individuals who averaged 55% accuracy when working alone were able to amplify their accuracy to 72% by forming real-time swarms and making the predictions together. This corresponds to 131% amplification in predictive accuracy across five consecutive weeks (50 games). This also corresponds to a performance level that not only matched, but slightly exceeded, an "expert source" of game predictions, the SAM super-computer used by the BBC to publish expert picks. Thus, by forming artificial swarms of approximately 30 individuals, groups of EPL fans (mostly American) were able to make game predictions at an expert level.

To assess statistical significance, we compared the swarm performance to the performance that would be expected by chance from a matching population using a bootstrap approach as follows: each week, we 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 we had ten predictions from the ten randomly selected individuals. We then averaged the accuracy of these predictions. We repeated the procedure (i.e. random selection of ten individuals and response assignment) 10000 times and computed the average distribution of correct answers for that week.

Distributions are shown in Figure 6 below. The mean of the distribution represents the average number of correct predictions that should be expected by chance, by a matching forecasters population. It can be seen that swarms are well above the mean as compared to individual predictions. We 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 of the swarm by chance.

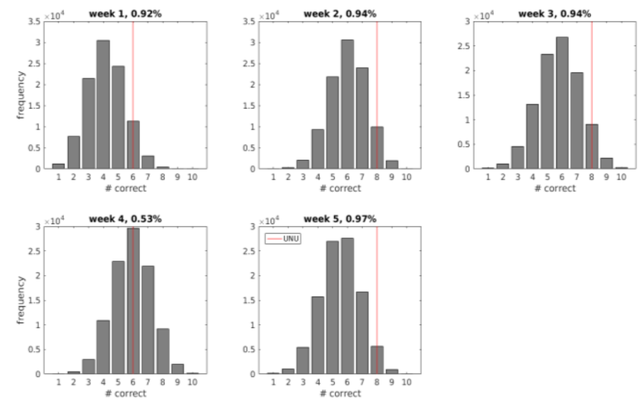

Fig. 6. Individual vs Swarm predictions, assessed weekly.

To aggregate the results from the five weeks into one, we compared the overall number of hits 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 result 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 7 below. The red line indicates the superior performance of the human swarm as compared to the individual forecasters.

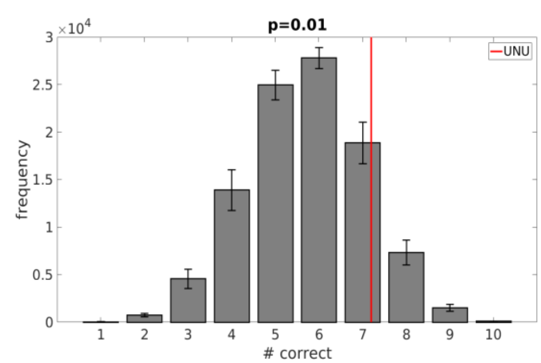

Fig. 7. Individual vs Swarm predictions aggregated for all five weeks.

VI. CONCLUSIONS

Can swarms of novice participants such as casual sports fans rival the predictive abilities of a respected expert source? The results presented herein suggest this may be the case. As demonstrated across five consecutive weeks of EPL match predictions, swarms of approximately 30 average sports fans were able to achieve competitive results to the SAM super-computer that the BBC employs for providing expert level predictions to the public. In fact, the 30 average sports fans, when working together as an Artificial Swarm Intelligence, out-predicted the SAM super-computer in four of the five weeks. Even more significant, by thinking together as a unified swarm intelligence, the groups of approximately 30 casual sports fans were able to significantly amplify their collective performance across all five weeks of EPL match predictions, boosting their overall prediction accuracy by 131% as compared to the average individual participant.

ACKNOWLEDGMENT

Thanks to David Baltaxe and Chris Hornbostel of Unanimous A.I. for their significant efforts in supporting the data collection for this research study. Thanks also to Unanimous A.I. for the use of the unu.ai platform throughout this five week experiment.

REFERENCES

[1] Rosenberg, Louis. Artificial Swarm Intelligence vs Human Experts, Neural Networks (IJCNN), 2016 International Joint Conference on. IEEE. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] McHale, Ian. Boshnakov, Georgi, and Kharrati, Tarak. A Bivariate Weibull Count Model for Forecasting Association Football Scores. Working Paper 2016.

[3] 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

[4] 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.

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

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

[7] 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.

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

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