Artificial Intelligence and Social Computing, Vol. 72, 2023, 189–197
https://doi.org/10.54941/ahfe1003287

"Human Swarms" of Novice Sports Fans Beat Professional Handicappers When Forecasting NFL Football Games

Hans Schumann[1], Louis Rosenberg[2], and Gregg Willcox[3]

[1]Unanimous AI, San Francisco, CA, USA
[2]Unanimous AI, Pismo Beach, CA, USA
[3]Unanimous AI, Seattle, WA, USA

ABSTRACT

The biological phenomenon of Swarm Intelligence (SI) enables social species to converge on group decisions by interacting in real-time systems. Studied in schools of fish, bee swarms, and bird flocks, biologists have shown for decades that SI can greatly amplify group intelligence in natural systems. Artificial Swarm Intelligence (ASI) is a computer-mediated technique developed in 2015 to enable networked human groups to form real-time systems that can deliberate and converge on decisions, predictions, estimations, and prioritizations. A unique combination of real-time HCI methods and AI algorithms, ASI technology (also called "Human Swarming" or "Swarm AI") has been shown in many studies to amplify group intelligence in forecasting tasks, often enabling small groups of non-professionals to exceed expert level performance. In the current study, small groups of approximately 24 amateur sports fans used an online platform called Swarm to collaboratively make weekly predictions (against the spread) of every football game in four consecutive NFL seasons (2019 - 2022) for a total of 1027 forecasted games. Approximately 5 games per week (as forecast by the human swarm) were identified as "predictable" using statistical heuristics. Performance was compared against the Vegas betting markets and measured against accepted performance benchmarks for professional handicappers. It is well known that professional bettors rarely achieve more than 55% accuracy against the Vegas spread and that top experts in the world rarely exceed 58% accuracy. In this study the amateur sports fans achieved 62.5% accuracy against the spread when connected as real-time "swarms." A statistical analysis of this result (across 4 NFL seasons) found that swarms outperformed the 55% accuracy benchmark for human experts with significance ($p = 0.002$). These results confirmed for the first time that groups of amateurs, when connected in real-time using ASI, can consistently generate forecasts that exceeded expert level performance with a high degree of statistical certainty.

Keywords: Swarm intelligence, Human swarms, Artificial swarm intelligence, Collective intelligence, Wisdom of crowds, Hyperswarms, AI

INTRODUCTION

It is widely accepted in the field of Collective Intelligence (CI) that the combined knowledge, wisdom, and insights of human groups will generally exceed that of its most skilled or informed members. A variety of methods have been explored over the last hundred years for harnessing groups to drive more precise forecasts, predictions, and decisions (Boland, 1989; De Condorcet, 1785; Malone 2019; Larrick and Soll, 2006). Artificial Swarm Intelligence (ASI) is a relatively new technique that was first proposed by Rosenberg in 2015 as a real-time alternative to traditional collective intelligence methods and is modeled on biological swarms (Rosenberg 2015; Rosenberg 2016). A variety of subsequent studies have shown that ASI can significantly amplify the decision-making accuracy of networked human groups (Askay et al., 2019; Metcalf et al., 2019; Willcox et al., 2019).

ASI systems operate by connecting distributed teams of networked users in real-time, forming closed-loop systems moderated by algorithms inspired by biological swarms. Unlike votes, polls, or surveys, which treat participants as separable datapoints for post-hoc statistical aggregation, the "swarming" process treats each individual as an active member of a synchronous system, enabling the full group to converge on solutions as a unified intelligence. This is achieved using a unique combination of human-computer interaction (HCI) methods and intelligence algorithms (Rosenberg et al., 2017; Rosenberg and Willcox, 2020). A snapshot of a real-time "human swarm" is shown in Fig. 1 below as it would be seen simultaneously by all participants.

In recent years, ASI technology has been used to support a diverse array of real-world decision-making applications, from simple itemized selections (as shown above) to more complex estimations, predictions, prioritizations, or diagnoses. For example, a study of ASI technology conducted at Stanford University School of Medicine in 2018 showed that small groups of radiologists, when using real-time swarming algorithms, could diagnose chest

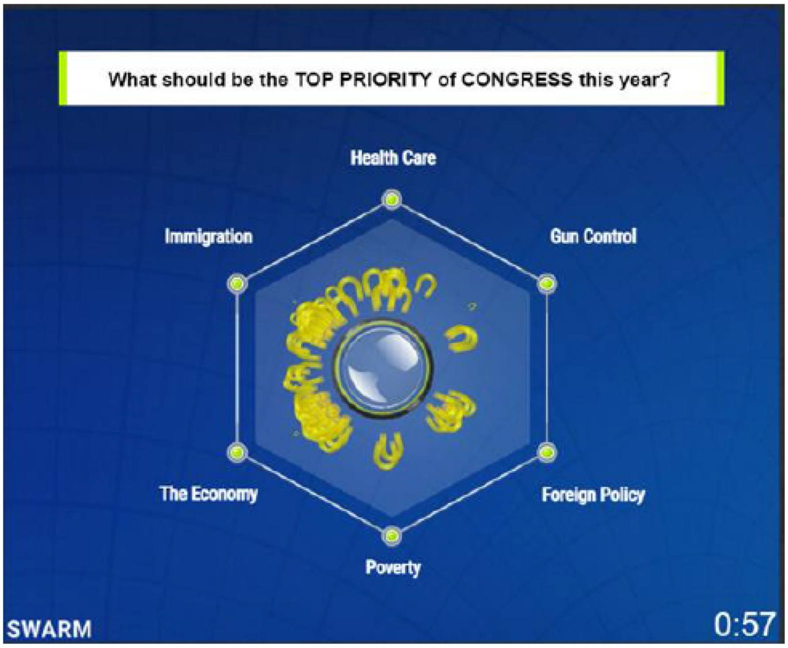

Figure 1: Snapshot of a real-time human swarm in which 150 networked participants work together as a dynamic system to collaboratively answer a question.

X-rays with 33% fewer errors than standard methods (Halabi et al., 2018; Rosenberg et al., 2018; Shah et al., 2022). Researchers at Boeing and the U.S. Army published a study in 2018 showing that small teams of military pilots, when using ASI technology, could generate subjective insights about the design of cockpits with higher effectiveness and usefulness than standard methods (Befort et al., 2018). Researchers at California Polytechnic published a study showing that networked business teams could increase the accuracy of subjective judgement by over 25% when deliberating as real-time swarms (Askay et al., 2019; Metcalf et al., 2019). Researchers at Unanimous AI, Oxford University, and MIT showed that small groups of financial traders, when forecasting the price of oil, gold, and stocks, increased their accuracy by over 25% when using ASI (Rosenberg et. al., 2021; Schumann et al., 2019; Willcox et al., 2019).

While the power of swarm-based systems to amplify group intelligence has been validated across many disciplines, an open question is the ability of ASI technologies to amplify the accuracy of networked groups of amateurs and enable them to perform collectively at levels that meet or exceed individual experts. In this study we compare groups of approximately 24 novice sports fans, connected using an ASI software platform called Swarm (from Unanimous AI) to predict every NFL football game during four consecutive seasons (2019 - 2022). All games were predicted against the spread (ATS). Such bets offer near 50% odds that the favored team wins by X points or more. For example, if Tennessee is favored by 2.5 points over Houston, then Tennessee -2.5 and Houston +2.5 will both be offered as bets with roughly equal odds. The bets will pay out if Tennessee wins by 3 points or more, or if Houston loses by less than 2.5 points or wins the game. Due to the even odds on each outcome, a novice bettor with no expertise is expected to be 50% accurate over the course of a full season.

After predicting the full set of 12 to 16 NFL games (ATS) each week using a real-time "human swarm," a simple set of statistical heuristics were used to select approximately 4 to 8 games each week deemed "most likely to beat the spread." This was done to mimic the behavior of expert handicappers, as the betting skill of a human expert is not their ability to predict the outcome of all sporting events but rather to assess which subset of weekly matchups are most likely to beat the published spread in major betting markets. This task was chosen because it could be compared against well-known benchmarks for professional sports bettors, who generally achieve a predictive accuracy of 55% against the spread when forecasting NFL football games, while the very best forecasters achieve a predictive accuracy of around 58% on the same games (Fish, 2015).

METHOD

Starting in 2019 and running through the completion of the 2022 season, the Swarm software platform was used with human participants to forecast NFL games against the spread on each Thursday before the weekend of games. Roughly 24 human participants (sourced using Amazon Mechanical Turk), who self-identified as NFL fans, were gathered in real-time for each swarm session. To motivate performance, participants were paid a nominal fee for their time and given small bonuses for accurate forecasts. For each of the 17–18 weeks in the season, the following process was used to collect human swarming data on each of the games:

1) Separate the 12 to 16 weekly games into three groupings based on the size of the spread: a low, medium, and high spread grouping.
2) Forecast the winner ATS for each game using Swarm. An example is shown in Figure 2.
3) After forecasting each game alone, ask the participants to collectively rank the ATS favorites within each grouping from the most to least likely to cover the spread using a process of elimination. An example of successive questions is shown in Figure 3.

The methodology above first generated a swarm-based forecast for every one of the 12 to 16 games each week. The method then asked the participants to rank the games (by grouping) on how likely it is that the favorite will cover the spread. The support for each question was calculated as the amount of "pull" imparted by users towards the chosen answer over the course of the full deliberation.

To select games to bet on using this data, four heuristics were followed:

1) For the ATS pick with the highest collective support, bet against the swarm's pick. This generally shows public overconfidence in that team.
2) For all ATS underdogs picked in which the swarm's support for the underdog was less than 67.5%, bet on the underdog ATS.
3) For low-spread and medium-spread groupings, bet on the top-ranked ATS favorite if the game is not already covered by Heuristic #1.
4) For low-spread and medium-spread groupings, bet against lowest-ranked favorite, taking the underdog ATS, if game is not already covered by Heuristic #1.

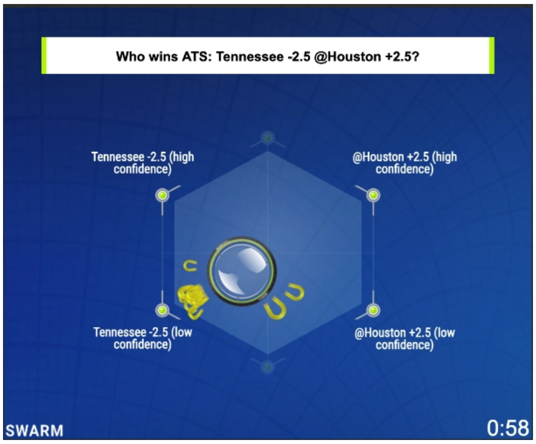

Figure 2: Snapshot of a real-time human swarm in which the swarm considers the winner of Tennessee and Houston against the spread.

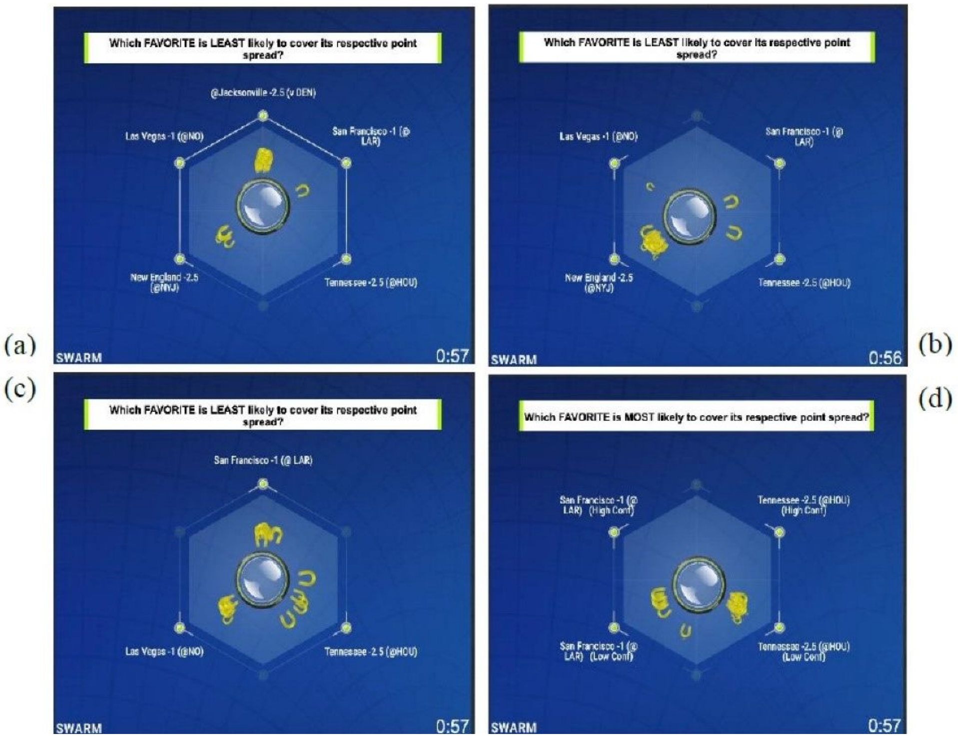

Figure 3: Series of snapshots of a real-time human swarm using a process of elimination to rank the least-to-most likely favorites to cover the spread from a single spread group. Process proceeds from (a) to (d), each time eliminating the answer selected by the swarm as the least likely choice. The final question (d) is reversed, instead asking which is most likely. The final ranking generated in this sequence, from the most to least likely, is: [Tennessee −2.5, San Francisco −1, Las Vegas −1, New England −2.5, Jacksonville −2.5].

The motivation behind these heuristics is as follows: Heuristic #1 indicates that the public is overconfident that a team will win ATS, meaning the spread line is probably set too high. In this case, it makes sense to bet against the Swarm's pick. Heuristic #2 is a rare case, as the swarm more often chooses the favorite to win ATS despite the even odds on both outcomes. In the case where the Swarm picks an underdog with mild support, the group is often carefully considering the game and reaching a conclusion that goes against their normal habit, which is often a profitable opportunity for betting. In cases where the group chooses the underdog with too much support, however, an element of groupthink may be at play, so underdog picks with support over 67.5% are avoided. Heuristics #3 and #4 build off the strength of rankings in Swarm by identifying bets that ranked very high and very low respectively, which often indicate that the participants collectively view the spread as too low or too high compared to other games in the same category.

A total of 1027 games were considered with Swarm, and 368 were selected using these heuristics (an average of 5.26 selected bets per week). The accuracy and profit results of these 368 bets will be discussed below.

RESULTS

The season-by-season accuracy and number of bets are shown in Table 1. In all seasons, the heuristics exceeded 60% accuracy over at least 80 picks. The 60% accuracy is higher than both a novice bettor's expected accuracy of 50% and a professional bettor's accuracy of 55% against the spread.

To test the statistical significance of the swarm-based heuristics, binomial tests were used, assuming two different null hypothesis probabilities; 50% for novice bettors and 55% for professional bettors. The two hypothesis tests testing the true accuracy, p, of the swarm-based heuristic bets are:

1) H_0: the heuristics have the same accuracy as novice bettors ($p = 0.50$)
   $H_{A:}$ the heuristics are more accurate than novice bettors ($p > 0.50$)
2) H_0: the heuristics have the same accuracy as professional bettors ($p = 0.55$)
   $H_{A:}$ the heuristics are more accurate than professional bettors ($p > 0.55$)

The binomial tests for the hypothesis tests described above tested the likelihood that the bets made with Swarm were more accurate than random chance would predict given a baseline level of accuracy (50% or 55%). Because multiple hypothesis tests were performed on the same set of data, a conservative Bonferroni adjustment was used to adjust the level of significance needed to conclude statistical significance. With 10 tests, four seasons for each baseline accuracy level plus two more for the overall significance across all seasons, the 0.05 alpha-level for significance decreased to 0.005.

As shown in Table l, at the 5% alpha-level, the swarm-based heuristics are statistically significantly better than both novice bettors and professional bettors for the results over all seasons. The probability of a novice bettor having an accuracy of 62.5% or more over 368 bets was less than 1 in 1000. For a professional bettor, it was still less than 1% ($p = 0.002$).

In addition to testing betting accuracy, Figure 4 and Table 2 below show the profits that a simulated bettor would have generated using the betting heuristics outlined above. Assuming the bettor placed $1,000 on each game over all four of the NFL seasons, the total profit would be over $75,000, with no seasons having less than $16,000 of profit. With 368 total bets achieving $77,718 in profit, the average return on investment per bet is +21.1%.

These results indicate that Swarms can more accurately forecast NFL games ATS than professional bettors—even when the swarms are composed of groups of amateurs. Further, by following the simple heuristics laid

Table 1. Season-by-season results of betting heuristics.

Season	Number of Bets	Accuracy	P-value against 50% novice	P-value against 55% professional bettor
2019	84	63.1%	0.011	0.083
2020	89	62.9%	0.001	0.081
2021	99	63.6%	0.004	0.051
2022	96	60.4%	0.026	0.168
All seasons	368	62.5%	< 0.001	0.002

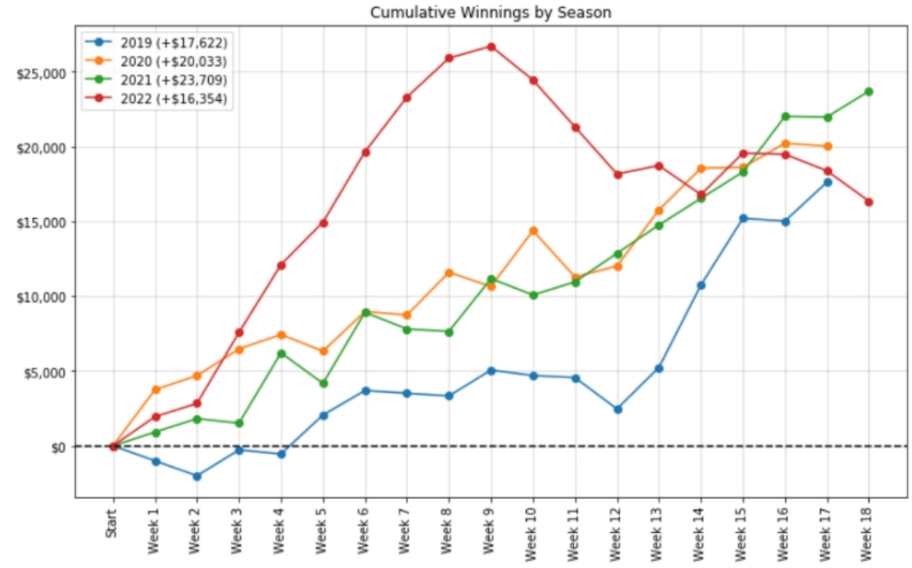

Figure 4: Weekly cumulative profit by season with $1,000 bets on each game.

Table 2. Season-by-season profits using $1,000 bets on each game.

Season	Number of Bets	Profit	Average ROI Per Bet
2019	84	+$17,622	+21.0%
2020	89	+$20,033	+22.5%
2021	99	+$23,709	+23.9%
2022	96	+$16,354	+17.0%
All seasons	368	+$77,718	+21.1%

out above, positive returns can be realized when betting on these games in real markets.

CONCLUSION

Artificial Swarm Intelligence (ASI) is a powerful real-time method for amplifying the knowledge, wisdom, insight, and intuition of human groups, enabling optimized solutions to quickly emerge as an interactive system. While ASI has been found to be highly effective across a wide range of applications from financial forecasting to medical diagnosis, it has been challenging to find an application in which groups of amateur forecasters working as "human swarms" could be tested against experts across large numbers of forecasted events in a statistically rigorous way. In this study such a comparison was performed and found that across four NFL football seasons (over 1000 football games), groups of amateur forecasters, when connected using ASI technology, could not only achieve expert level performance, but were able to achieve **62.5%** accuracy against the spread, which was significantly better than the typical professional bettor at 55% accuracy ($\mathbf{p = 0.002}$) and higher than even the best professional sports bettors achieve across a full season. For example, one of the top professional bettors of all time, Billy Walters, reportedly averaged 58% accuracy against the spread during the height of his career (Schumann et al., 2019). In this study we demonstrated that a group of 24 amateurs could significantly outperform this benchmark when mediated by ASI technology and supported by selection heuristics.

ACKNOWLEDGMENT

The authors would like to thank Chris Hornbostel, Pranav Bhagat, Patty Sullivan, and David Baltaxe for their efforts in running and administering the Swarm sessions that collected data for this study.

REFERENCES

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

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.

De Condorcet, Nicolas. Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Cambridge University Press, 1785.

Fish, Mike. (2015) "Meet the World's Most Successful Gambler." ESPN, ESPN Internet Ventures, 6 Feb. 2015, http://www.espn.com/espn/feature/story/_/id/12280555/how-billy-walters-became-sports-most-successful-controversial\hbox{-bettor}

Halabi, Safwan., et al. (2018) "Radiology SWARM: Novel Crowdsourcing Tool for CheXNet Algorithm Validation", SiiM Conference on Machine Intelligence in Medical Imaging, 2018.

Larrick, Richard P., and Soll, Jack B. "Intuitions about combining opinions: Mis-appreciation of the averaging principle." Management science 52.1 (2006): 111–127.

Malone, T. W. (2019) Superminds: The surprising power of people and computers thinking together, Amazon. Little, Brown and Company.

Metcalf L, Askay DA, Rosenberg LB. Keeping Humans in the Loop: Pooling Knowledge through Artificial Swarm Intelligence to Improve Business Decision Making. California Management Review. 2019;61(4): 84–109.

Rosenberg L., Willcox G. (2020) Artificial Swarm Intelligence. In: Bi Y., Bhatia R., Kapoor S. (eds) Intelligent Systems and Applications. IntelliSys 2019. Advances in Intelligent Systems and Computing, vol 1037. Springer, Cham. https://doi.org/10.1007/978-3-030-29516-5_79

Rosenberg, L, Willcox, G., Halabi, S., Lungren, M, Baltaxe, D. and Lyons, M. (2018) 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.

Rosenberg, L., Domnauer, C., Willcox, G., Schumann, H. (2021) "Hyperswarms: A New Methodology for Amplifying Group Intelligence" ACM Collective Intelligence Conference 2021, Copenhagen, Denmark – June 29, 2021.

Rosenberg, L., Pescetelli, N., Willcox, G.., (2017) "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.

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

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.

Schumann, H., Willcox, G., Rosenberg, L., & Pescetelli, N. (2019). "Human Swarming" Amplifies Accuracy and ROI when Forecasting Financial Markets. In 2019 IEEE International Conference on Humanized Computing and Communication (HCC). 2019 IEEE International Conference on Humanized Computing and Communication (HCC). IEEE. https://doi.org/10.1109/hcc46620.2019.00019

Shah, R., Astuto A. Nunes, B., Gleason, T. et al. Utilizing a Digital Swarm Intelligence Platform to Improve Consensus Among Radiologists and Exploring Its Applications. J Digit Imaging (2022). https://doi.org/10.1007/s10278-022-00662-3

Willcox G., Rosenberg L., Schumann H. (2020) Group Sales Forecasting, Polls vs. Swarms. In: Arai K., Bhatia R., Kapoor S. (eds) Proceedings of the Future Technologies Conference (FTC) 2019. FTC 2019. Advances in Intelligent Systems and Computing, vol 1069. Springer, Cham.