“Human Swarming” Amplifies Accuracy and ROI when Forecasting Financial Markets

Hans Schumann
Unanimous A.I.
San Francisco, CA, USA
hans@unanimous.ai

Gregg Willcox
Unanimous A.I.
San Francisco, CA, USA
gregg@unanimous.ai

Louis Rosenberg
Unanimous A.I.
San Luis Obispo, CA, USA
louis@unanimous.ai

Niccolo Pescetelli
Massachusetts Institute of Technology
Cambridge, MA, USA
niccolop@mit.edu

***Abstract*— Many social species amplify their decision-making accuracy by deliberating in real-time closed-loop systems. Known as Swarm Intelligence (SI), this natural process has been studied extensively in schools of fish, flocks of birds, and swarms of bees. The present research looks at human groups and tests their ability to make financial forecasts by working together in systems modeled after natural swarms. Specifically, groups of financial traders were tasked with forecasting the weekly trends of four common market indices (SPX, GLD, GDX, and Crude Oil) over a period of 19 consecutive weeks. Results showed that individual forecasters, who averaged 56.6% accuracy when predicting weekly trends on their own, amplified their accuracy to 77.0% when predicting together as real-time swarms. This reflects a 36% increase in forecasting accuracy and shows high statistical significance (p<0.001). Further, if investments had been made according to these swarm-based forecasts, the group would have netted a 13.3% return on investment (ROI) over the 19 weeks, compared to the individual’s 0.7% ROI. This suggests that enabling groups of traders to form real-time systems online, governed by swarm intelligence algorithms, has the potential to significantly increase the accuracy and ROI of financial forecasts.**

Keywords*— *Swarm Intelligence, Artificial Swarm Intelligence, Collective Intelligence, Wisdom of Crowds, Human Swarming, Artificial Intelligence, Financial Forecasting, Human Forecasting.

I. INTRODUCTION

Extensive prior research has shown that groups of human forecasters can outperform individual forecasters by aggregating estimations across groups using simple statistical methods [1-3]. Often referred to as the Wisdom of Crowds (WoC) or Collective Intelligence (CI), this phenomenon was first observed over a century ago and has been applied to many fields, from predicting financial markets to forecasting geopolitical events. The most common methods involve polling a population of individuals for self-reported estimations and then aggregating the collected input statistically as a simple or weighted mean [4].

In recent years, a new method has been developed that is not based on aggregating data from isolated individuals, but instead involves groups of forecasters working together as real-time systems, their interactions moderated by AI algorithms modeled on the natural principle of Swarm Intelligence.

Known as Artificial Swarm Intelligence (ASI) or simply “Human Swarming,” this method has been shown in numerous studies to significantly amplify the accuracy of forecasts generated by human group [5-11]. For example, in a recent study conducted at Stanford University School of Medicine, groups of radiologists were asked to forecast the probability that patients are positive for pneumonia based on a reviews of their chest x-rays. When forecasting together as a real-time swarm, diagnostic errors were reduced by over 30% [12].

While prior studies have shown ASI systems to significantly amplify the predictive accuracy of human groups across a range of tasks, from forecasting sporting events to predicting sales volumes of new products, the present study was conducted to assess whether swarm-based forecasts of financial markets can achieve similar improvements. To address this, a nineteen-week study was conducted that tasked groups of financial traders with making weekly forecasts regarding the change in price of four financial indices – the S&P 500 (SPX), the price of gold (GLD), the price of gold mining stocks (GDX), and the price of crude oil (CRUDE). The objective was to assess whether a significant improvement would be measured when comparing individual forecasts to swarm-based predictions. Swarm performance was also compared with traditional “Wisdom of Crowd” aggregation methods. In this way, the present study compared three forecasting methods – as Individuals, Crowds, and Swarms.

II. SWARMS VS CROWDS

In crowd-based forecasting methods, participants provide input in isolation, usually via polling, for statistical aggregation. In swarm-based methods, groups of human participants forecast together in real-time systems modeled after biological swarms. The present study uses Swarm AI technology, which is modeled largely on the dynamic behaviors of honeybee swarms.

The decision-making process that governs honeybee swarms has been researched since the 1950s and has been shown at a high level to be quite similar to decision-making in neurological brains [13,14]. Both employ populations of simple excitable units (i.e., neurons and bees) that work in parallel to integrate noisy evidence, weigh competing alternatives, and converge on decisions in real-time. In both brains and swarms, outcomes are arrived at through 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 hundreds of scout bees to work in parallel, collecting information about their local environment, and then converge together on a single optimal decision, picking the best solution to complex multi-variable problems [15-17].

The similarity between "brains" and "swarms" becomes even more apparent when comparing decision-making models that represent each. The decision process in primate brains is often modeled as mutually inhibitory leaky integrators that aggregate incoming evidence from competing neural populations [18]. A common framework for primate decision is the Usher-McClelland model in Figure 1 below.

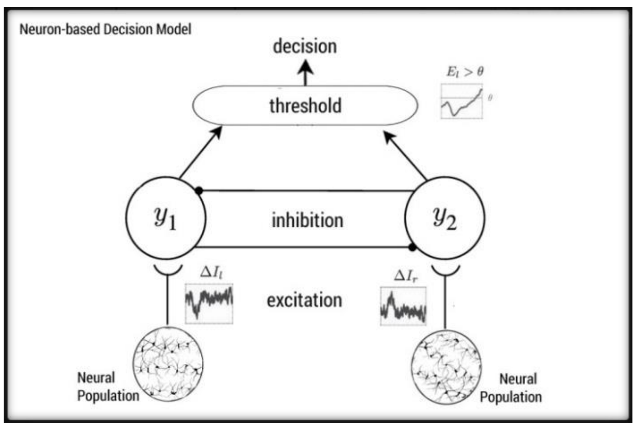

Fig. 1. Usher-McClelland model of neurological decision-making

This neurological decision model can be compared to swarm-based decision models, for example the honey-bee model represented in Figure 2. As shown below, swarm-based decisions follow a very similar process, aggregating input from sub-populations of swarm members through mutual excitation and inhibition, until a threshold is exceeded.

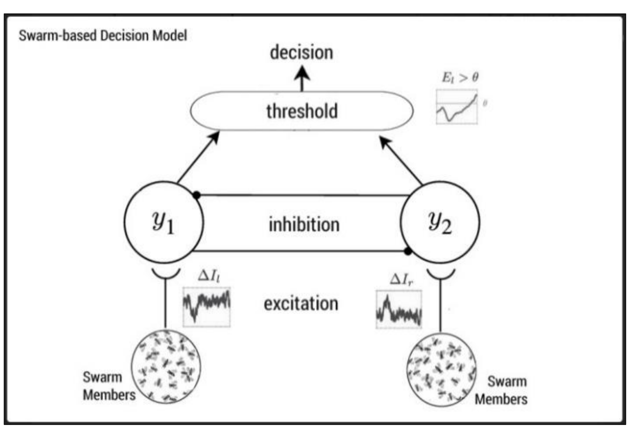

Fig. 2. Mutually inhibitory decision-making model in bee swarms

Thus, while brains and swarms are very different forms on intelligence, both are systems that enable optimized decisions to emerge from the interactions among collections of processing units. The goals of the present study are twofold – (i) to assess if groups of financial traders can form swarm-based systems that can "think together" as unified intelligence, and (ii) to compare accuracy of swarm-based forecasts with financial forecasts generated by individual members or by statistical groups aggregated using traditional Wisdom of Crowd techniques.

III. SWARMING SOFTWARE

In the natural world, swarming organisms establish real-time feedback-loops among group members. Swarming bees do this using complex body vibrations called a "waggle dance." To enable real-time swarming among groups of networked humans, Swarm AI technology was developed. It allows distributed groups of users to form closed-loop systems moderated by swarming algorithms [5-7]. Modeled on the decision-making process of honeybees, Swarm AI allows groups of distributed users to work in parallel to (a) integrate noisy evidence, (b) weigh competing alternatives, and (c) converge on 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, networked human groups can answer questions as a "swarming system" by collaboratively 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 with respect to the moving puck, participants impart their personal intent on the system as a whole. The input from each user is not a discrete vote, but a stream of real-time vectors that varies freely. Because all users can adjust their intent continuously in real-time, the swarm moves, not based on the input of any individual, but based on the dynamics of the full system. This enables a complex negotiation among all members at once, empowering the group to collectively explore the decision-space and converge on the most agreeable solution in synchrony.

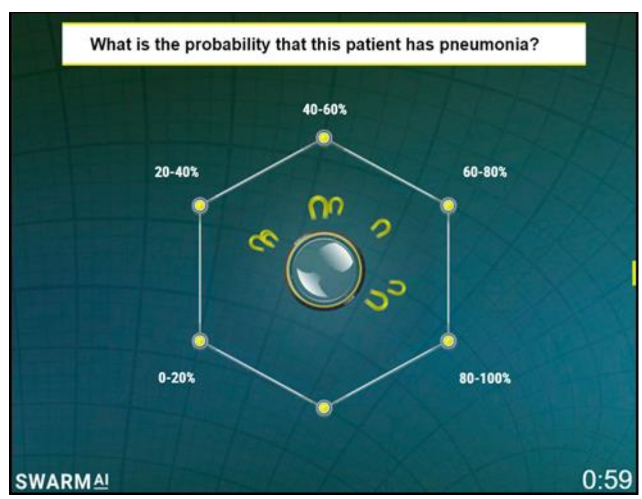

Fig. 3. A human swarm answering a question in real-time

It is important to note that participants freely modulate both the direction and 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 adjust their magnet so that it stays near the puck's outer rim. This is significant, for it requires participants to remain continuously engaged throughout the decision process, evaluating and re-evaluating the strength of their opinions as they convey their contribution. If they stop adjusting their magnet with respect to the changing position of the puck, the distance grows and their imparted 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, intelligence algorithms monitor the behaviors of all swarm members in real-time, inferring their implied conviction based upon their relative motions over time. This reveals a range of behavioral characteristics within the swarm population and weights their contributions accordingly, from entrenched participants to flexible participants to fickle participants.

IV. Financial Forecasting Study

To assess the ability of human swarms to amplify their accuracy in financial predictions, a study was conducted over a nineteen week period using groups of volunteers who were unaffiliated with the research team. The participants were all self-identified as "active traders" who follow the financial markets daily and make financial trades regularly. Each weekly group consisted of between 7 to 36 participants. To establish a baseline, all participants provided their weekly forecasts as individuals using a standard online survey. The group then congregated online as a real-time swarm using the Swarm platform to make synchronous forecasts.

Across the nineteen week period, predictions were made for the following financial indices: (a) the S&P 500 (**SPX**), (b) the gold shares index fund (**GLD**), (c) the gold miners index fund (**GDX**), and (d) the crude oil index (**CRUDE**). The forecasts were generated every Tuesday at market close. The participants were asked to predict if each index would be higher or lower from the current price at market close on Friday (i.e. 72 hours later). Predictions were recorded first from individuals on private surveys, then from swarms working together as a system. In addition, participants were asked to qualify the expected change in price by indicating if the predicted move would be "by a little" or "by a lot." This was included as a means for evoking participant confidence in their directional forecast rather than as a true predictor of magnitude.

Figure 4 shows an ASI system (i.e. a "human swarm") comprised of 24 participants in the process of forecasting a weekly change in GDX price. It's important to note that this is a snapshot of a single moment time, as it generally takes between 10 and 60 seconds of deliberation for the system to converge upon a solution. As shown in the figure, the group is given four options to choose from, enabling the set of human forecasters to identify which direction the index will move, as well as express a general sense of magnitude. The magnitude indicator is helpful as it causes the swarm to split into multiple different factions and then converge over time on a solution that maximizes their collective confidence and conviction. Figure 5 shows a time-integrated of the deliberation as a heat map, the brightness representing the level of support imparted for each option.

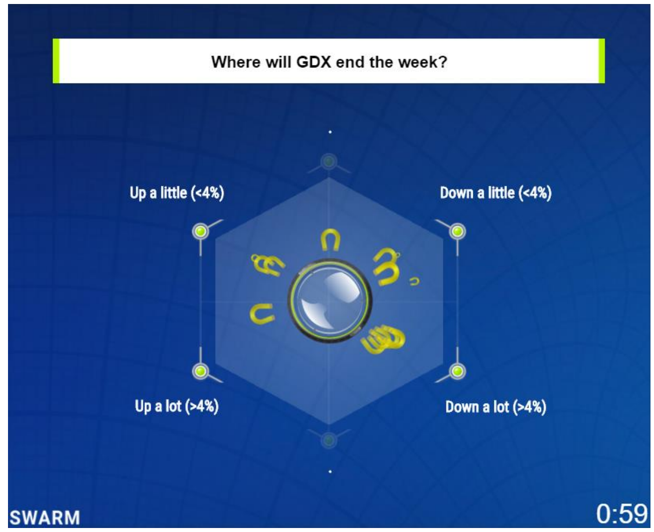

Fig. 4. Snapshot of a human swarm predicting GDX in real-time

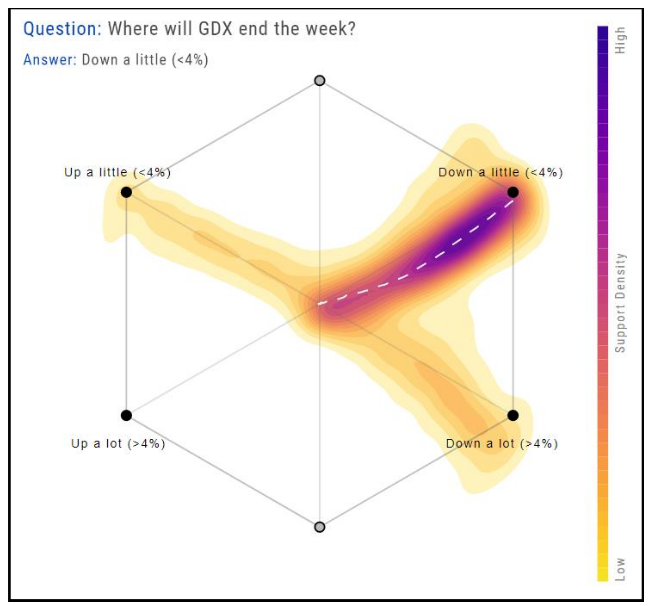

Fig. 5. Support Density heatmap of swarm predicting GDX in real-time

V. Analysis and Results

For each of the nineteen weeks in the testing period, a set of predictions were made for each of the four market indices (SPX, GLD, GDX, CRUDE), providing 76 sets of four predictions. Results were generated indicating: (a) Individual Accuracy – computed as the average performance across the pool of human subjects, (b) Crowd Accuracy – computed by taking the most popular prediction from the participant pool and using that to compute accuracy over time, and (c) Swarm Accuracy – computed by assessing the accuracy of the predictions made by the swarms each week.

To assess whether the human swarms predicted the directional change in market indices (i.e. UP or DOWN) more accurately than individuals, the swarm's performance was compared with the individuals' performance using a bootstrapping procedure. For each of the four investment categories (SPX, GLD, GDX, CRUDE) and each prediction week, we selected the answer provided by an individual sampled at random among the individuals who provided a response for that particular week and investment type. Answers were averaged across the four investment types and the 19 weeks to obtain a percentage accuracy measure. The procedure was repeated 1,000 times in order to obtain a distribution of probabilities for making a correct prediction.

The distribution, shown in Figure 6 as a probability density function, represents the probability of an individual making a correct prediction when responses are randomly sampled from the individual answers provided. With a mean accuracy of **56%,** the individuals were moderately better than random guessing when predicting the directional change in these market indicators. The red line in Figure 6 shows the empirical performance of the swarms, which at **77%** accuracy was significantly higher performing as compared to individuals. The probability that the swarm and the crowd were more accurate than individuals due to random chance was calculated using a bootstrapping procedure, and was found to be extremely low ($p<0.001$) indicating a highly significant result.

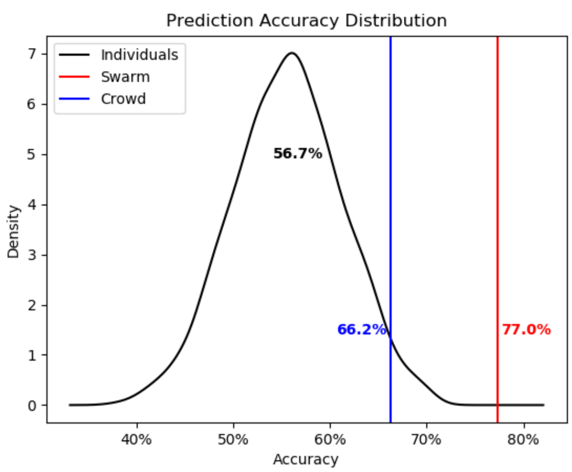

Fig. 6. Individual vs Swarm vs Crowd Accuracy when predicting the directional change in all four indices in the subsequent 72-hour period.

A similar analysis was done using the more traditional "Wisdom of Crowd" method of taking the most popular predictions across the pool of individuals as the forecast. The crowd in this study achieved a 66.2% accuracy, shown as a blue line in the figure above. The probability that the swarm performed better than the crowd due to random chance was low ($p=0.022$), indicating that we can be confident that the swarm significantly outperformed the crowd in aggregate in this study.

Looking at the results as a percentage increase, the swarms, on average, were **36%** more accurate when predicting the directional movement in the financial indices than the individual financial traders who comprised those swarms.

In addition to analyzing the predictive accuracy across all four indices in aggregate (as shown in Figure 5 above), it is also instructive to assess performance with respect to each of the four financial categories in isolation, shown in Figure 7 below. Across 19 weeks, the swarm outperformed or matched the individual traders and the crowd-based forecasts in all four instances.

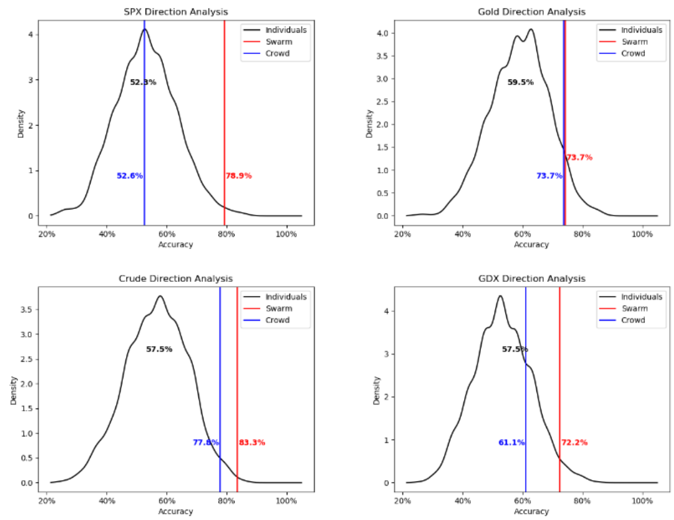

Fig. 7. Individual Accuracy vs Swarm Accuracy when predicting the directional change in each individual index in the subsequent 72-hour period.

Focusing on the ability of swarming to amplify the accuracy of financial predictions, the improvements for each of the four assets above are summarized in Table 1 below. As shown, the largest accuracy increase achieved by swarm-based forecasting was recorded in SPX predictions, which showed an impressive **26.6** percentage point gain over the individuals, corresponding to a **43%** amplification in accuracy. The swarm-based forecasts also outperformed the crowd-based forecasts, achieving an average increase of **10.8** percentage points across the four assets tested. This corresponds to a net **16%** amplification in total accuracy for swarm-based forecasts vs crowd-based forecasts.

Index	Swarm Accuracy	Crowd Accuracy	Individual Accuracy
SPX	**78.9%**	52.6%	52.3%
Gold	**73.7%**	**73.7%**	59.5%
Crude	**83.3%**	77.8%	57.5%
GDX	**72.2%**	61.1%	57.5%
Average	***77.0%***	*66.2%*	*56.7%*

Table 1. Individual Accuracy vs Swarm Accuracy across each index

A paired t-test was used to calculate the likelihood that the swarm was more accurate than the crowd at predicting the direction of stock movement due to random chance alone. The results of this test, as shown in Table 2 below, reveal that we can be confident that the swarm outperforms individuals in each index ($p<0.05$ for each individual index), and we can also be confident that the Swarm outperformed the crowd on average (p=0.022) and the crowd when predicting SPX only (p=0.010).

Index	Swarm vs Crowd p-value	Swarm vs Individual p-value
SPX	0.010	0.002
Gold	0.500	0.039
Crude	0.289	0.002
GDX	0.166	0.027
Overall	0.022	< 0.001

Table 2. Significance between Swarm and Crowd or Individual Directional Forecast Accuracy

To make the difference in accuracy between these predictive methods more concrete, a financial simulation was conducted to calculate the financial impact of investing using the guidance of swarms versus individual forecasts and the crowd's average forecast. In this simulation, each forecasting method started with a $1000 bankroll, and invested 100% of its bankroll each week evenly across the four predicted indexes. If the forecasting method predicted the index would increase in price, a "long" position was taken, while if the method predicted a decrease in price, the index was "shorted". The net bankroll was tallied at the end of each week, accounting for the position that was taken and the decrease or increase in the price of each of the assets that week, and the new bankroll was then re-invested according to the next week's predictions. The final return on investment of the forecasting method was calculated as the final bankroll divided by the initial bankroll ($1000).

The results of this simulation are shown in figure 8 below and summarized in table 3. The swarm again outperforms the crowd, ending the 19-week simulation with a 13.28% ROI, while the crowd ends with an 8.87% ROI. The individuals were the lowest performers, ending with a positive, but lower 3.60% ROI. To put these results into perspective, the performance that would have resulted from simply investing "long" (i.e. buy and hold without trading) in the four assets is plotted in red and ends up with a 1.96% ROI. Clearly, both the crowd and the swarm were able to predict weekly price swings to some degree, and as a result outperform the market in the long term in this study.

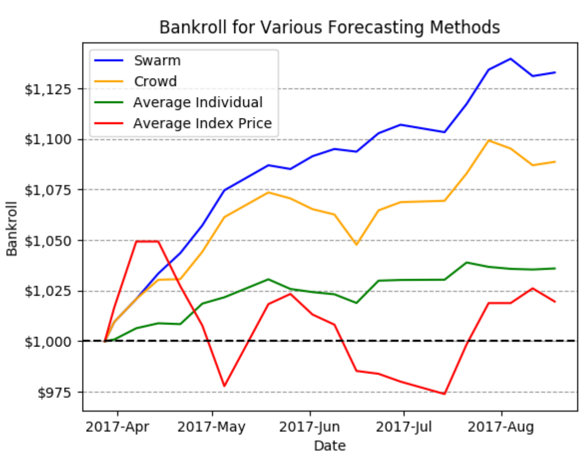

Fig. 8. Simulated Bankroll by Week for each Forecasting Method

Forecast	Final Bankroll	Percent Increase
Swarm	$1,132.81	13.28%
Crowd	$1,088.68	8.87%
Individual	$1,035.95	3.60%
Short	$980.41	-1.96%
Long	$1,019.60	1.96%

Table 3. Simulated Bankroll by Week for each Forecasting Method

To color these results further, the probability that the swarm-based ROI outperformed the crowd-based ROI and the average Individual's ROI by random chance is calculated using a bootstrapping test. In this test, the forecasts that each method makes are resampled 1,000 times, and the average ROI per dollar investment is calculated. The average ROI per dollar investment is used instead of the compounded ROI at the end of the study to mitigate the effect of compounding on the final results (i.e. to ensure that early-week correct predictions don't artificially inflate the outcome). This histogram of bootstrapped average ROI per dollar investments is shown in figure 9.

The probability that the Swarm outperformed the Market due to random chance was low (p<0.001), so we can be confident that this swarm of financial traders over these 19 weeks would on average outperform the market. The probability that the swarm outperformed the crowd due to random chance was also low (p=0.077).

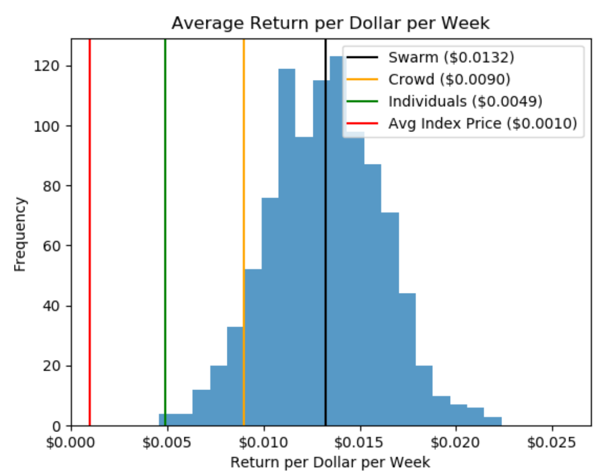

Fig. 9. Histogram of Swarm Average Return per Dollar per Week to Crowd, Individual, and Market.

VI. CONCLUSIONS

This study explored if real-time swarms of financial traders could outperform the predictive accuracy of either (i) individual traders and/or (ii) groups of traders aggregated using traditional Wisdom of Crowd (WoC) techniques. The results showed that groups of forecasters, working together in real-time swarms, can significantly outperform the accuracy of individual traders when predicting the directional movement of four common financial assets (SPX, GLD, GDX, and CRUDE).

The results also show that the swarm-based forecasts could outperform crowd-based forecasts, with the most significant results being achieved in the prediction of the S&P index fund (SPX). In addition, the results of this study show that when investments are made using these swarm-based forecasts, a significantly higher return on investment (ROI) is achieved compared to investments made using either (i) individual forecasts or (ii) crowd-based forecasts.

Additional research is warranted to further validate the benefits of swarm-based forecasting for financial applications. Of particular interest is the ability of ASI technology to amplify prediction accuracy in longer term predictions, as the current study used a relatively short 72-hour forecasting window. Other topics recommended for ongoing research include exploring swarm-based forecasting using participant groups of larger sizes, comparing participants of varying expertise levels, and testing improved swarming algorithms.

ACKNOWLEDGMENT

Thanks to Chris Hornbostel and David Baltaxe for their efforts in coordinating the weekly swarms of financial traders that generated predictions. Also, thanks to Unanimous AI for the use of the Swarm platform for this ongoing work. This work was partially funded by NSF Grant #1840937.

REFERENCES

[1] Galton, F. (1907). Vox Populi. Nature, 75, 450-451.

[2] Steyvers, M., Lee, M.D., Miller, B., & Hemmer, P. (2009). The Wisdom of Crowds in the Recollection of Order Information. In Y. Bengio and D. Schuurmans and J. Lafferty and C. K. I. Williams

[3] Philip E. Tetlock and Dan Gardner. 2015. Superforecasting: The Art and Science of Prediction. Crown Publishing Group, New York, NY, USA.

[4] J Dana, P Atanasov, P Tetlock, B Mellers (2019), Are markets more accurate than polls. The surprising informational value of "just asking." Judgment and Decision Making 14 (2), 135-147

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

[6] Rosenberg, Louis. "Artificial Swarm Intelligence vs Human Experts," Neural Networks (IJCNN), 2016 International Joint Conference on. IEEE.

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

[8] Baltaxe, David, Rosenberg, Louis and N. Pescetelli, "Amplifying Prediction Accuracy using Human Swarms", Collective Intelligence 2017. New York, NY ; 2017.

[9] Willcox G., Rosenberg L., Askay D., Metcalf L., Harris E., Domnauer C. (2020) Artificial Swarming Shown to Amplify Accuracy of Group Decisions in Subjective Judgment Tasks. In: Arai K., Bhatia R. (eds) Advances in Information and Communication. FICC 2019. Lecture Notes in Networks and Systems, vol 70. Springer, Cham

[10] L. Rosenberg, N. Pescetelli and G. Willcox, "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.

[11] L. Rosenberg and G. Willcox, "Artificial Swarm Intelligence vs Vegas Betting Markets," 2018 11th International Conference on Developments in eSystems Engineering (DeSE), Cambridge, United Kingdom, 2018, pp. 36-39

[12] L. Rosenberg, M. Lungren, S. Halabi, G. Willcox, D. Baltaxe and M. Lyons, "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, pp. 1186-1191.

[13] 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

[14] 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.

[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