Measuring Group Personality with Swarm AI

Gregg Willcox[1], David Askay[2], Louis Rosenberg[1], Lynn Metcalf[2], Bryan Kwong[1], and Richard Liu[3]

[1] Unanimous AI, San Francisco, CA, USA
{gregg,louis,bryan}@unanimous.ai
[2] California Polytechnic State University, San Luis Obispo, CA, USA
{daskay,lmetcalf}@calpoly.edu
[3] University of California, Berkeley, CA, USA
richard.y.liu@berkeley.edu

Abstract. The aggregation of individual personality assessments to predict team performance is widely accepted in management theory but has significant limitations: the isolated nature of individual personality surveys fails to capture much of the team dynamics that drive real-world team performance. Artificial Swarm Intelligence (ASI)—a technology that enables networked teams to think together in real-time and answer questions as a unified system—promises a solution to these limitations by enabling teams to collectively complete a personality assessment, whereby the team uses ASI to converge upon answers that best represent the group's disposition. In the present study, the group personality of 94 small teams was assessed by having teams take a standard Big Five Inventory (BFI) assessment both as individuals, and as a real-time system enabled by an ASI technology known as Swarm AI. The predictive accuracy of each personality assessment method was assessed by correlating the BFI personality traits to a range of real-world performance metrics. The results showed that assessments of personality generated using Swarm AI were far more predictive of team performance than the traditional aggregation methods, showing at least a 91.8% increase in average correlation with the measured outcome variables, and in no case showing a significant decrease in predictive performance. This suggests that Swarm AI technology may be used as a highly effective team personality assessment tool that more accurately predicts future team performance than traditional survey approaches.

Keywords: Group Personality; BFI; Group Performance; Swarm Intelligence; Artificial Swarm Intelligence; Collective Intelligence; Group Consensus.

1   Introduction

Most businesses strive to build high performing teams wherein the combination of skills, personality traits, and work habits of team-members drives effective execution towards organizational goals. One commonly used technique for predicting whether a team will be high performing is to administer a personality assessment to each individual member, aggregate the team's results, and use those aggregated results to forecast whether the combined team is likely to work well together [23,31,11,10]. Prior research has shown a correlation between aggregated results on personality assessments and team performance [24]. The current study reviews these prior methods and explores whether improved forecasts of team performance can be attained using Artificial Swarm Intelligence—a unique AI technology that aims to more accurately assess team personality.

As further background, multilevel approaches to investigating organizational phenomena are critical, yet understudied [29]. Multilevel research often involves aggregating individual-level data (e.g., the personalities of individuals comprising a team) to measure group-level constructs (e.g., team performance). Typically, individual-level data are aggregated to measure group-level phenomena in one of four ways: by calculating a mean score of individual measures, by computing the highest (maximum) individual trait score, by computing the lowest (minimum) individual trait score, or by looking at the variance of individual trait scores within a team [17,4]. Each of these four methods of aggregation have been found to capture unique information about the group [17]. For example, conscientiousness, an individual-level personality trait, is associated with high levels of organization and attention to detail. Averaging the conscientiousness scores of individuals comprising a team assumes that the amount of conscientiousness possessed by each individual team member contributes to the collective pool of conscientiousness available to the team, regardless of how the trait is distributed among team members. Alternatively, the lowest scoring individual on conscientiousness brings the rest of the group down on average, making the group minimum score the most appropriate way to aggregate individual scores. While each of these methods of aggregation provide unique insights, researchers continue to question the efficacy of using individual-level measures to assess group-level traits or phenomena.

An alternative aggregation method, group consensus, offers the potential to improve the accuracy of personality assessments. A consensus approach, whereby all members consider each question on an assessment and jointly agree on a collective score, has been advocated because it better captures the underlying and unique group dynamics present in teams [3,8]. For example, a study of MBA students found that measuring team efficacy through a consensus approach was a better predictor of group performance than when measured through aggregated individual-level constructs [13]. While the consensus method offers a potentially superior way of aggregating individual-level constructs, it suffers from drawbacks. Specifically, the context of a group discussion allows for social influence to silence some members or to encourage conformity. Additionally, achieving consensus is costly in terms of time and logistical organization of participants.

For these reasons, and despite the potential of group-level consensus personality measurement and calls to move away from the aggregation of individual-level data [25], researchers seldom use group-level consensus ratings.

Advances in networking technology and artificial intelligence have led to the development of Artificial Swarm Intelligence (ASI) systems that provides a way for groups of humans to quickly reach a consensus in a way that overcomes these limitations. ASI has been found to significantly amplify decision-making accuracy in human groups across a wide range of tasks [34,35,27], from forecasting [32,33] to subjective judgments [2,39] and prioritizations [36]. Indeed, groups can achieve consensus in less than 60 seconds, while also limiting social influence from group members through anonymous deliberation in a way that captures group dynamics. ASI presents a promising method that answers the call for research using consensus-based aggregation approaches [38]. Specifically, we focus on the potential of using ASI as a method of administering and composing group-level personality assessments, and in predicting team performance based on these personality assessments.

2 Foundations of Swarm Intelligence

In the natural world, Swarm Intelligence (SI) enables social organisms to aggregate their collective insights rapidly and to converge in synchrony on optimal decisions by forming real-time closed-loop systems. Swarm Intelligence has been deeply studied across many social species, from schools of fish and flocks of birds to swarms of honey bees and even slime molds. Unlike birds, bees and fish, humans have not evolved the natural ability to form real-time swarms, as we lack the innate mechanisms used by other species to form closed-loop systems. Schooling fish detect vibrations in the water around them. Flocking birds detect high-speed motions propagating through the group formation. Swarming bees generate complex body vibrations called a "waggle dance" that encode assessment information. To enable networked human groups to form similar closed-loop systems, a cloud-based platform called "swarm.ai" was developed. It enables human groups, connected from remote locations, to make collective predictions, decisions, and assessments by working together as closed-loop swarms.

When using the swarm.ai platform, networked human teams answer questions by collaboratively moving a graphical pointer to select from a set of answer options. Each participant provides their individual input by manipulating a graphical magnet with a mouse, touchpad, or touchscreen. By adjusting the position and orientation of their magnet with respect to the moving puck, participants express their real-time intent. The input from each user is not a discrete vote, but a stream of vectors that varies freely over time. Because all members of the group can adjust their intent continuously in real-time, the swarm explores the decision-space, not based on the input of any individual member, but based on the emergent dynamics of the full system. This enables a complex behavioral interaction among all members of the population, empowering the group to collectively consider the options and synchronously converge on the most agreeable solution.
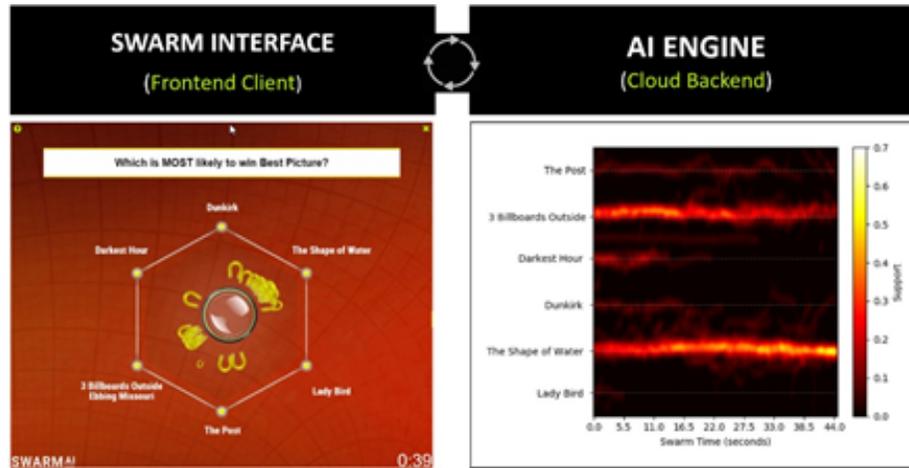
Fig. 1. Architecture of the swarm.ai platform with graphical client and cloud-based AI engine

It is important to note that participants not only vary the direction of their intent but also modulate the magnitude of their intent by adjusting the distance between their magnets and the pointer, which is commonly represented as a graphical puck. Because the graphical puck is in continuous motion across the decision-space, users need to move their magnets continually so that they stay close to the puck's rim. This is significant for it requires that all participants, regardless of group size or composition, be engaged continuously throughout the decision process, evaluating and re-evaluating their intent in real-time. If a participant stops adjusting their magnet with respect to the changing position of the puck, the distance grows and the participant's influence on the group's decision wanes.

Thus, like bees vibrate their bodies to express sentiment in a biological swarm or neurons fire 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. This is generally referred to as a "leaky integrator" structure and common to both swarm-based and neuron-based systems. In addition, intelligence algorithms monitor the behaviors of swarm members in real-time, inferring their relative conviction based on their actions and interactions over time. This reveals a range of behavioral characteristics within the population and weights their contributions accordingly.

Just as ASI provides an effective way for groups to reach a consensus around decision-making, it is a promising method for reaching a consensus around responses to psychometric assessments like a personality test. Through ASI, a question can be answered in less than 60 seconds, participants are anonymous

and less subject to dysfunctional social influence, and consensus is achieved through interactions as participants deliberate visually through the interface.

3 Method

To assess the ability of ASI technology to function as an accurate assessment tool of team personality, a large study was conducted across a set of 94 working groups (i.e. teams), each comprising 3 to 6 members. Each of these teams were engaged in a 10-week group project. In total, 384 human subjects participated in this study. All were college students enrolled in business, communication studies, or engineering courses, for which a team project was a significant component. Participants first completed the personality assessment individually by themselves, then they completed a personality assessment collectively as a group using ASI. The individual results were used to compose group-level team personality through typical aggregation approaches (mean, max, min, and variance). The results from the ASI represent a consensus-based team personality. Finally, at the conclusion of the group project, an outcome survey was administered individually to participants to measure group outcomes (e.g., performance).

The Big Five Inventory (BFI) assessment [20] was used to measure personality for both individual and ASI conditions. Qualitics was used to administer the assessment to individuals and the Swarm® software platform was used for measuring ASI consensus. The BFI assessment is commonly used in literature and industry as a personality assessment tool, and a wide body of research has validated that individual and group scores on this assessment are correlated with performance on real-world tasks [5,7,28,14,22,18,21,6]. The BFI measures five dimensions of personality:

- Extraversion–the degree to which a person enjoys interacting with others, expresses energy, and asserts themselves
- Openness–the degree to which a person is curious, imaginative, and interested in new ideas.
- Agreeableness–the degree to which a person interacts with others with empathy, optimism
- Neuroticism–the degree to which a person is easily upset and emotionally reactive
- Conscientiousness–the degree to which a person is organized and has self-control

The questions that were included in Individual and Swarm versions of the BFI assessment are listed in Appendix A. When answering the BFI individually, participants were asked about their own personalities (e.g., Are you talkative?). When groups were asked questions through ASI, the referent shifted to the group-level (e.g., Is this group talkative?).

The swarms were attended by 297 (77.3%) participants, and any group in which fewer than 2 individuals participated in the swarm was eliminated from the dataset. The swarms had one minute to answer each question, and if they failed to reach a consensus in that time (referred to as a Brain Freeze), the

question was repeated only once. No swarm experienced a brain freeze during the second round.

The individual personality assessments were aggregated in post-processing into a group personality assessment using each of four different methods: (1) average group score, (2) minimum score in a group, (3) maximum score in a group, and (4) the variance of individual scores. In this way, the traditional method for assessing group personality (i.e. statistically aggregating individual BFI scores) and a new method for assessing group personality (i.e. enabling teams to take the BFI assessment together as a unified swarm intelligence) could be directly compared.

Several team outcome variables were measured at the conclusion of the group project, which occurred several days or weeks after the swarm assessment. Several performance related self-assessments were administered to each team member:

- Cohesiveness–degree of bonding towards the team, team members, and the task [9]. This is comprised of three sub-dimensions: task attraction, interpersonal cohesiveness, and task commitment.
- Conflict–the degree of relational, task, and process-based conflict experienced in the group [19].
- Psychological Safety–the degree to which group members feel like they can be vulnerable and speak up with other group members [12].
- Potency–general perception of the group's confidence and capability [15].
- Satisfaction–the degree to which group members are pleased with group members and the team [37].
- Viability–the degree to which the group desires to work together again in the future [16].
- Transactive memory–the degree to which group members know about the skills, emotions, and tasks of other group members [26]. This is comprised of three sub-dimensions: specialization, credibility, and coordination.
- Team Effectiveness—a self-rating of how well the group accomplished it's task [1].

Prior studies have established connections between group-level personality and these performance outcome variables. For each group, the aggregated scores (average, min, max, variance) and the swarm scores for the BFI were correlated with the six performance indicators with Pearson's correlation coefficient. The resulting R^2 values were compared and used for statistical tests in analysis.

4   Analysis

The correlation between each personality assessment method and the performance of each team was calculated using a linear regression. The Pearson coefficient of determination (R^2 value) between each BFI dimension and performance metric was calculated for each of the five group personality measurement methods (average, minimum, maximum, variance, ASI). The study originally measured 17 performance metrics, which have been averaged by category down to 9 metrics for ease of viewing.

The R^2 values for each personality measurement method are shown in Appendix A, and the survey average versus swarm correlations with the performance metrics are shown in figure 2 below. Immediately, these plots show that, on average, swarm-based assessments of group personality have a higher correlation with team performance than the survey-based assessments of group personality.
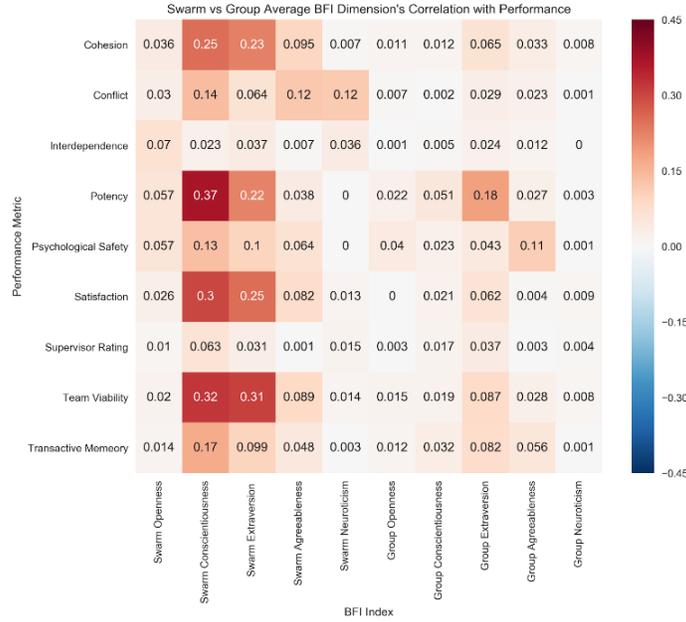
Fig. 2. Architecture of the swarm.ai platform with graphical client and cloud-based AI engine

A bootstrapped significance test was performed to measure whether the swarm could have outperformed the survey methods in this test due to random chance alone. In this process, the observed groups (including the personality assessment by each method and performance metrics) were randomly resampled with replacement 1000 times, and the 90% confidence interval of the difference in R^2 values between the survey and swarm assessments of group personality was calculated. This process was repeated for each group performance metric and each surveying method.

A table of confidence intervals generated using this approach is shown in Appendix B, with the cells in which the swarm's assessment was found to correlate with the performance metric significantly more than the survey's assessment highlighted in yellow, and the cells in which the reverse is true highlighted in green. Table 1 below gives an overview of this statistical significance test: out of the 85 comparisons made between each survey assessment method and the

swarm, the swarm significantly outperformed the survey in at least 25.9% of cases, while the survey never significantly outperformed the swarm.

Table 1. Summary of Bootstrapped Correlation Differences between Swarm and Survey Assessments of Team Personality

Survey Assessment Methods	Swarm Significantly Outperforms the Survey	Survey Significantly Outperforms the Swarm	Average R^2 Increase
Average	30 (35.3%)	0 (0%)	0.0654
Maximum	25 (29.4%)	0 (0%)	0.0687
Minimum	22 (25.9%)	0 (0%)	0.0484
Variance	24 (28.2%)	0 (0%)	0.0684

Next, the correlation of all five facets with each of the outcome variables was calculated for each assessment method using a multivariate linear regression. The correlation, reported as the adjusted R^2, is shown in Table 2. Interestingly, the Swarm method was more correlated with the majority of the outcome variables than the four survey assessment methods: 13 of the 14 measured outcome variables were most correlated with the swarm method. The one exception to this was the Team Specialization outcome variable, to which the Group Minimum method had the highest correlation.

The significance of the difference between the Swarm's correlation with each outcome variable and each survey assessment method's correlation was calculated with a bootstrapping approach: the observed data were resampled with replacement 1,000 times, and the frequency with which the Swarm's correlation was greater than that of each survey assessment method's correlation was recorded. The 90% confidence interval of the difference in correlations and p-value of the correlation difference for each assessment method were calculated using this approach and are recorded in Appendix C. Differences that were significant (p<0.05 and p<0.01) are recorded in Table 2.

A large number of significant differences were recorded using this approach: The Swarm method significantly outperformed each survey assessment method on at least 7 (as compared to the Group Minimum method) and at maximum 10 (as compared to the Group Average method) of the 14 outcome variables.

The average correlation of each method with the major outcome variables (i.e. not the sub-dimensions of outcome variables, such as Team Coordination) was calculated. On average, the swarm method logged a 0.351 R^2 correlation with these major outcome variables, indicating that the swarm-measurements of group personality explained 35.1% of the variance in the measured team behaviors. The survey assessment method that performed the best on average was the Group Minimum method, which achieved a 0.183 R^2 correlation with the outcome variables on average. As a result, the swarming method explained 91.8% more of the variance in measured team behaviors than the best-performing sur-

vey assessment method did. This indicates that swarming to produce an estimate of group personality may produce significantly more insightful estimates of the group dynamics and performance as compared to traditional approaches to group personality analysis.

Table 2. Correlation of Multivariate Regression with Outcome Variables by Personality Measurement Method (Correlation Displayed as Adjusted R^2)

Outcome Variables	Assessment Method				
	Swarm Group	Avg	Group Max	Group Min	Group Var
Task Attraction	**0.352**	0.116**	0.134**	0.134*	0.094**
Interpersonal Cohesiveness	**0.376**	0.156*	0.194*	0.134*	0.129*
Task Commitment	**0.357**	0.073**	0.108**	0.117**	0.091**
Cohesiveness Total	**0.395**	0.126**	0.160**	0.144**	0.113**
Relationship Conflict	**0.307**	0.089*	0.121	0.162	0.124
Task Conflict	**0.208**	0.132	0.083	0.178	0.084
Process Conflict	**0.251**	0.057*	0.082	0.162	0.043*
Conflict Total	**0.292**	0.095*	0.102	0.187	0.087*
Team Specialization	0.104	0.119	0.105	**0.160**	0.087
Team Credibility	**0.217**	0.138	0.198	0.124	0.080
Team Coordination	**0.338**	0.152*	0.143*	0.182*	0.118*
Team Total	**0.244**	0.152	0.159	0.175	0.090
Team Satisfaction	**0.407**	0.090**	0.114**	0.124**	0.121**
Team Potency	**0.442**	0.196**	0.160*	0.305*	0.173*
Average	**0.351**	0.129**	0.137**	0.183**	0.118**

Bolded entries indicate the highest-performing assessment methods for each outcome variable.
* = Significant at p=0.05 level relative to Swarm
** = Significant at p=0.01 level relative to Swarm

5   Conclusion

The group personality of 94 small teams was assessed by asking the teams to respond to a standard set of 45 Big Five Inventory questions using both traditional surveys of individual personality and a real-time collaboration interface (Swarm AI) to establish a group consensus of the team's own personality. Four different multilevel approaches to aggregating the team member's answers to the survey BFI questions were studied: the average, variance, minimum, and maximum of the team's answers.

The performance of the surveying methods was compared to the swarming methods by correlating the BFI dimensions, as calculated by each method, to various metrics of the team's self-reported performance. The swarming method significantly outperformed each of the survey aggregation methods at predicting

a wide range of performance metrics (at least 25.9%, n=85), and were never significantly outperformed by the survey aggregation methods.

The performance of the surveying methods was also compared to the swarming methods using a multivariate regression to correlate the BFI dimensions with various metrics of the team's self-reported performance. The swarming method significantly outperformed each of the survey aggregation methods in over half of the outcome variables studied and logged an 91.8% higher average correlation with the outcome variables than the Group Minimum method (the best-performing survey assessment method).

This result suggests that ASI can be used to evaluate team personality, and predict team performance, more accurately than traditional individual surveying methods. There are several advantages to this approach. First, it overcomes concerns about both time and social influence of the consensus-based approach to aggregation. The average time to reach a consensus was 18.8 seconds. The anonymity provided by the platform enables participants to interact and deliberate visually, while protecting the identities of team members. Second, the analysis reveals that the BFI results of the ASI-based group consensus was a stronger predictor of important group outcomes, such as performance, viability, and cohesion. In doing so, it provides a response to calls for consensus-based aggregation and support for consensus being a superior method of aggregating group-level variables [9]. Future research is needed to replicate and extend these findings to new contexts and different group-level variables.

This research was limited by the availability and participation rate of participants, as 72.9% of participants did not take the pre-swarm survey, and 77.3% did not participate in the swarm. This research also did not investigate whether the presentation of the question itself contributed to the higher success rate of the swarm in predicting team performance, since participants were asked directly about the team's personality in the swarm, but were asked about their own personality in the surveys.

Future research could also investigate whether the behavioral data collected from swarms, such as the average pull direction or time taken to reach a consensus, could be used to more accurately forecast group personality and team performance. This real-time behavioral data has been shown in the past to be predictive of other real-world outcomes [40,30].

Acknowledgment

References

1. Alper, S., Tjosvold, D., Law, S.A.: Conflict management, efficacy, and performance in self-managing work teams. Personnel Psychology **53**, 625–642 (2000)

2. Baltaxe, D., Rosenberg, L., Pescetelli, N.: Amplifying Prediction Accuracy using Human Swarms, Collective Intelligence 2017. New York, NY ; (2017)
3. Bar-Tal, D.: Group beliefs. Springer-Verlag, New York (1990)
4. Barrick, M.R., Stewart, G.L., Neubert, M.J., Mount, M.K.: Relating member ability and personality to work-team processes and team effectiveness. Journal of applied psychology **83**(3), 377–391 (1998)
5. Bell, S.T.: Deep-level composition variables as predictors of team performance: a meta-analysis. Journal of applied psychology **92**(3), 595 (2007)
6. Boudreau, J.W., Boswell, W.R., Judge, T.A.: Effects of personality on executive career success in the united states and europe. Journal of vocational behavior **58**(1), 53–81 (2001)
7. Bradley, B.H., Klotz, A.C., Postlethwaite, B.E., Brown, K.G.: Ready to rumble: How team personality composition and task conflict interact to improve performance. Journal of Applied Psychology **98**(2), 385 (2013)
8. Campion, M.A., Medsker, G.J., AC., H.: Relations between work group characteristics and effectiveness: Implications for designing effective work groups. PERSONNEL PSYCHOLOGY **46**, 823–850 (1993)
9. Carless, S.A., De Paola, C.: The measurement of cohesion in work teams. Small group research **31**(1), 71–88 (2000)
10. De Raad, B.: The Big Five personality factors. The psycholexical approach to personality. Hogrefe and Huber Publishers, Gottingen, Germany (2000)
11. Digman, J.M.: Personality structure: Emergence of the five-factor model. Annual Review of Psychology **41**, 417–440 (1990)
12. Edmondson, A.: Psychological safety and learning behavior in work teams. Administrative Science Quarterly **44**(2), 350–383 (1999)
13. Gibson, C.B., Randel, A.E., AE., E.: Understanding group efficacy: An empirical test of multiple assessment methods. Group and Organization Management **25**, 67–97 (2000)
14. Gregory, R.J.: Psychological testing: History, principles, and applications. Allyn and Bacon (2004)
15. Guzzo, R.A., Yost, P.R., Campbell, R.J., Shea, G.P.: Potency in groups: Articulating a construct. British journal of social psychology **32**(1), 87–106 (1993)
16. Hackman, J.R.: The design of work teams. In: Lorsch, J.W. (ed.) Handbook of organizational behavior:, pp. 315–342. Prentice-Hall, Englewood Cliffs, NJ (1987)
17. Hollenbeck, J.R., DeRue, D.S., Guzzo, R.: Bridging the gap between i/o research and hr practice: Improving team composition, team training, and team task design. Human Resource Management **43**(4), 353–366 (2004)
18. Howard, P.J., Howard, J.M.: Owner's Manual for Personality at Work. Bard Press (Austin, TX (2001)
19. Jehn, K.A., Mannix, E.A.: The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. Academy of management journal **44**(2), 238–251 (2001)
20. John, O.P., Naumann, L.P., Soto, C.J.: Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual issues. In: John, O.P., Robins, R.W., Pervin, L.A. (eds.) Handbook of personality: Theory and research, pp. 114–158. Guilford Press, New York, NY (2008)
21. Judge, T.A., Higgins, C.A., Thoresen, C.J., Barrick, M.R.: The big five personality traits, general mental ability, and career success across the life span. Personnel psychology **52**(3), 621–652 (1999)

22. Judge, T.A., Zapata, C.P.: The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the big five personality traits in predicting job performance. Academy of Management Journal **58**(4), 1149–1179 (2015)
23. Kichuk, S.L., Wiesner, W.H.: The big five personality factors and team performance: implications for selecting successful product design teams. Journal of Engineering and Technology Management,Volume **14**(3-4), 195–221 (1997)
24. Kichuk, S.L., Wiesner, W.H.P., Performance, T.: Implications for selecting successful product design teams. Tech. rep., Innovation Research Centre Working Paper 51 (1996)
25. Kirkman, B.L., Tesluk, P.E., Rosen, B.: Assessing the incremental validity of team consensus ratings over aggregation of individual?level data in predicting team effectiveness. Personnel Psychology **54**(3), 645–667 (2001)
26. Lewis, K.: Measuring transactive memory systems in the field: Scale development and validation. Journal of applied psychology **88**(4), 587–603 (2003)
27. Metcalf, L., Askay, D.A., Rosenberg, L.B.: Keeping humans in the loop: Pooling knowledge through artificial swarm intelligence to improve business decision making. California Management Review **4**(61), 84–109 (2019). https://doi.org/https://doi.org/10.1177/0008125619862256
28. O'Neill, T.A., Allen, N.J.: Team task conflict resolution: An examination of its linkages to team personality composition and team effectiveness outcomes. Group Dynamics: Theory, Research, and Practice **18**(2), 159 (2014)
29. Paruchuri, S., Perry-Smith, J.E., Chattopadhyay, P., Shaw, J.D.: New ways of seeing: Pitfalls and opportunities in multilevel research (2018)
30. Patel, B.N., Rosenberg, L., Willcox, G., et al.: Human-machine partnership with artificial intelligence for chest radiograph diagnosis. In: npj Digital Medicine. 2, 111 (2019). https://doi.org/10.1038/s41746-019-0189-7
31. Peeters, M., van Tuijl, H., Rutte, C., Reymen, I.: Personality and team performance: A meta-analysis. European Journal of Personality **20**(5), 377 (2006)
32. Rosenberg, L., Pescetelli, N., Willcox, G.: Artificial swarm intelligence amplifies accuracy when predicting financial markets **2017**, 58–62 (2017)
33. Rosenberg, L., Willcox, G.: Artificial swarm intelligence vs vegas betting markets. In: Cambridge, U.K. (ed.) 2018 11th International Conference on Developments in eSystems Engineering (DeSE). pp. 36–39 (2018)
34. Rosenberg, L.B.: Human swarms. In: a real-time method for collective intelligence. pp. 658–659. Proceedings of the European Conference on Artificial Life (2015)
35. Rosenberg, L.: Artificial swarm intelligence vs human experts. IJCNN **3** (2016)
36. Rosenberg, L.: Baltaxe, david and pescetelli, nicollo. "crowds vs swarms, a comparison of intelligence," ieee 2016 swarm/human blended intelligence (shbi), cleveland. OH pp pp. 1–4 (2016)
37. Vegt, G.S., Emans, B.J., Vliert, E.: Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction. Personnel Psychology **54**(1), 51–69 (2001)
38. Willcox, G., Askay, D., Rosenberg, L., Metcalf, L., Kwong, B., Liu, R.: Measuring group personality with swarm ai. In: IEEE Conference on Transdisciplinary AI (Trans AI) (2019)
39. Willcox, G., Rosenberg, L., Askay, D., Metcalf, L., Harris, E., Domnauer, C.: Artificial swarming shown to amplify accuracy of group decisions in subjective judgment tasks. In: Bhatia, R. (ed.) Arai K. Advances in Information and Communication. FICC 2019. Lecture Notes in Networks and Systems, vol 70. Springer, Cham (2020)

40. Willcox, G., Rosenberg, L., Donovan, R., Schumann, H.: Dense neural network used to amplify the forecasting accuracy of real-time human swarms. Computational Intelligence and Communication Networks **2019** (2019)

# Appendix A: Correlation Heat Maps
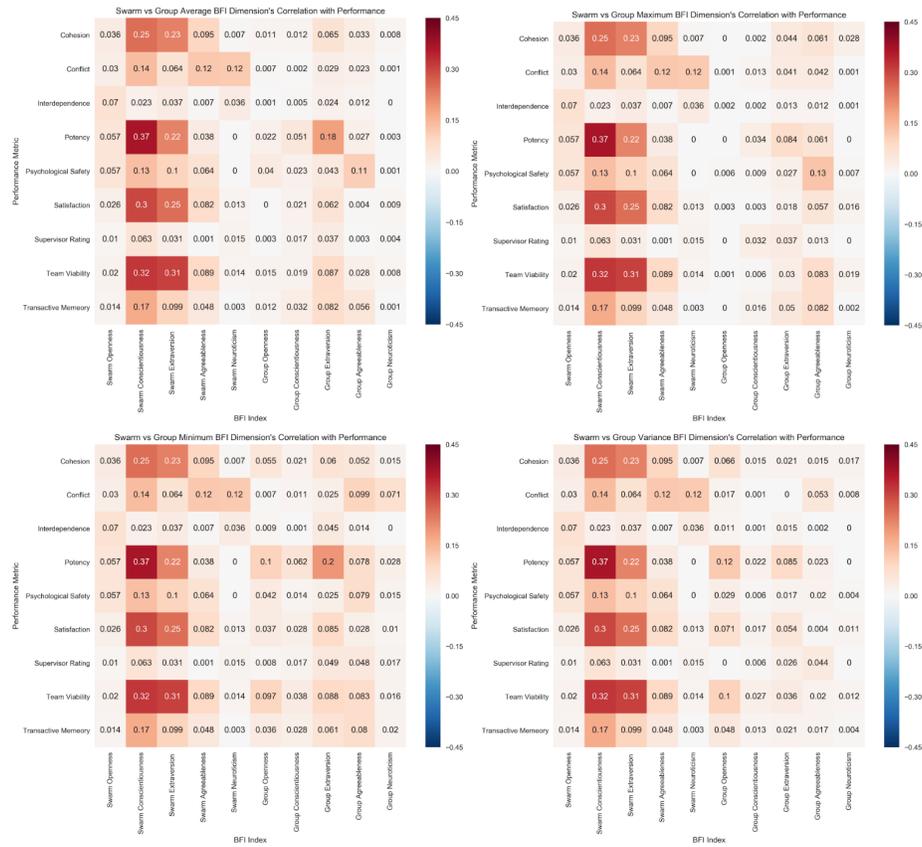
Fig. 3. Swarm and Survey Correlation with Performance Metrics (Reported as R^2). From top left clockwise: Swarm vs Survey Average, Swarm vs Survey Maximum, Swarm vs Survey Minimum, Swarm vs Survey Variance

Appendix B: BFI Trait Correlation Analysis: Bootstrapping Results

	Swarm R² - Survey Average R²				
	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Cohesiveness Task Attraction	-0.041:0.27	0.125:0.374	0.039:0.314	-0.062:0.171	-0.042:0.105
Interpersonal Cohesiveness	-0.06:0.091	0.074:0.341	0.065:0.381	-0.081:0.238	-0.067:0.108
Task Commitment	-0.037:0.137	0.139:0.432	0.021:0.263	-0.003:0.288	-0.067:0.081
Cohesiveness Total	-0.047:0.195	0.146:0.439	0.063:0.386	-0.043:0.258	-0.064:0.124
Relationship Conflict	-0.049:0.118	0.042:0.334	-0.051:0.09	-0.073:0.317	-0.017:0.293
Task Conflict	-0.09:0.068	0.006:0.233	-0.089:0.088	-0.01:0.177	0.0:0.21
Process Conflict	-0.041:0.125	0.007:0.233	-0.013:0.147	-0.064:0.228	-0.028:0.242
Conflict Total	-0.065:0.107	0.029:0.307	-0.051:0.123	-0.03:0.279	-0.002:0.286
Team Specialization	-0.104:0.111	-0.012:0.107	-0.128:0.086	-0.129:0.033	-0.081:0.043
Team Credibility	-0.057:0.085	0.037:0.263	-0.053:0.118	-0.157:0.083	-0.055:0.028
Team Coordination	-0.032:0.093	0.116:0.36	-0.094:0.114	-0.018:0.212	-0.031:0.183
Team Transactive Memory Total	-0.05:0.137	0.07:0.296	-0.105:0.136	-0.103:0.11	-0.049:0.053
Psychological Safety	-0.067:0.18	0.032:0.223	-0.031:0.161	-0.178:0.051	-0.027:0.04
Team Viability	-0.058:0.094	0.158:0.434	0.098:0.373	-0.048:0.199	-0.072:0.097
Team Satisfaction	-0.038:0.153	0.139:0.434	0.057:0.355	-0.01:0.257	-0.075:0.114
Team Potency	-0.055:0.18	0.206:0.477	-0.108:0.176	-0.049:0.113	-0.061:0.054
Team Effectiveness by Member	-0.015:0.225	0.174:0.455	-0.009:0.272	0.004:0.255	-0.05:0.069

Fig. 4. Bootstrapped difference in Pearson R values between the Swarm and the Survey Average methods of team personality assessments

	Swarm R2 - Survey Maximum R2				
	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
Cohesiveness Task Attraction	-0.001:0.317	0.132:0.421	0.062:0.344	-0.098:0.148	-0.077:0.088
Interpersonal Cohesiveness	-0.023:0.109	0.073:0.36	0.081:0.396	-0.121:0.179	-0.123:0.079
Task Commitment	-0.017:0.154	0.146:0.446	0.018:0.29	-0.029:0.261	-0.075:0.076
Cohesiveness Total	-0.008:0.217	0.142:0.467	0.074:0.401	-0.098:0.24	-0.089:0.094
Relationship Conflict	-0.038:0.128	0.046:0.312	-0.149:0.084	-0.061:0.308	-0.011:0.284
Task Conflict	-0.049:0.074	0.001:0.208	-0.072:0.126	-0.031:0.169	0.005:0.214
Process Conflict	-0.011:0.141	0.004:0.217	-0.076:0.164	-0.097:0.22	-0.004:0.237
Conflict Total	-0.027:0.104	0.027:0.28	-0.095:0.138	-0.071:0.265	0.006:0.289
Team Specialization	-0.041:0.159	-0.024:0.118	-0.126:0.117	-0.156:0.03	-0.121:0.043
Team Credibility	-0.025:0.081	0.043:0.285	-0.097:0.153	-0.195:0.064	-0.082:0.029
Team Coordination	-0.028:0.092	0.144:0.39	-0.067:0.191	-0.078:0.177	-0.032:0.18
Team Transactive Memory Total	-0.013:0.14	0.086:0.33	-0.105:0.187	-0.177:0.091	-0.044:0.065
Psychological Safety	-0.008:0.221	0.039:0.249	-0.036:0.173	-0.211:0.038	-0.046:0.038
Team Viability	-0.019:0.123	0.172:0.457	0.126:0.429	-0.098:0.159	-0.087:0.083
Team Satisfaction	-0.027:0.173	0.136:0.465	0.085:0.41	-0.078:0.171	-0.065:0.117
Team Potency	-0.009:0.207	0.202:0.505	-0.018:0.292	-0.103:0.09	-0.047:0.054
Team Effectiveness by Member	-0.036:0.213	0.168:0.476	0.02:0.324	-0.067:0.201	-0.052:0.076

Fig. 5. Bootstrapped difference in Pearson R values between the Swarm and the Survey Maximum methods of team personality assessments

	Swarm R² - Survey Minimum R²				
	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
Cohesiveness Task Attraction	-0.072:0.24	0.112:0.362	0.035:0.33	-0.059:0.176	-0.079:0.071
Interpersonal Cohesiveness	-0.169:0.054	0.051:0.32	0.094:0.415	-0.118:0.192	-0.087:0.104
Task Commitment	-0.092:0.111	0.13:0.402	0.016:0.279	-0.056:0.245	-0.081:0.057
Cohesiveness Total	-0.109:0.125	0.121:0.414	0.059:0.387	-0.08:0.23	-0.089:0.094
Relationship Conflict	-0.083:0.094	0.033:0.334	-0.023:0.103	-0.102:0.28	-0.122:0.252
Task Conflict	-0.049:0.065	-0.006:0.227	-0.091:0.104	-0.171:0.099	-0.067:0.172
Process Conflict	-0.03:0.116	-0.012:0.222	-0.006:0.158	-0.134:0.191	-0.141:0.194
Conflict Total	-0.055:0.096	0.016:0.303	-0.04:0.136	-0.14:0.203	-0.122:0.234
Team Specialization	-0.109:0.119	-0.037:0.098	-0.129:0.052	-0.132:0.02	-0.068:0.046
Team Credibility	-0.108:0.049	0.046:0.285	0.004:0.185	-0.123:0.11	-0.157:0.025
Team Coordination	-0.069:0.078	0.095:0.343	-0.028:0.188	-0.127:0.128	-0.066:0.141
Team Transactive Memory Total	-0.104:0.112	0.066:0.289	-0.022:0.175	-0.156:0.086	-0.125:0.026
Psychological Safety	-0.067:0.198	0.024:0.231	0.015:0.195	-0.114:0.081	-0.099:0.024
Team Viability	-0.191:0.036	0.132:0.418	0.095:0.396	-0.16:0.162	-0.065:0.102
Team Satisfaction	-0.098:0.104	0.125:0.414	0.038:0.335	-0.077:0.214	-0.056:0.134
Team Potency	-0.147:0.101	0.192:0.444	-0.077:0.177	-0.163:0.071	-0.097:0.04
Team Effectiveness by Member	-0.092:0.15	0.112:0.391	-0.017:0.238	-0.133:0.177	-0.063:0.059

Fig. 6. Bootstrapped difference in Pearson R values between the Swarm and the Survey Minimum methods of team personality assessments

	Swarm R^2 - Survey Variance R^2				
	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
Cohesiveness Task Attraction	-0.079:0.254	0.111:0.395	0.052:0.395	-0.022:0.21	-0.06:0.092
Interpersonal Cohesiveness	-0.195:0.04	0.03:0.312	0.098:0.488	-0.062:0.229	-0.084:0.12
Task Commitment	-0.076:0.135	0.137:0.431	0.013:0.333	-0.002:0.305	-0.071:0.076
Cohesiveness Total	-0.111:0.152	0.124:0.431	0.067:0.465	-0.03:0.308	-0.073:0.101
Relationship Conflict	-0.114:0.091	0.04:0.352	-0.016:0.111	-0.083:0.301	-0.039:0.268
Task Conflict	-0.09:0.059	-0.005:0.236	-0.026:0.152	-0.115:0.133	0.009:0.21
Process Conflict	-0.024:0.129	0.003:0.263	0.0:0.185	-0.047:0.202	-0.008:0.248
Conflict Total	-0.07:0.096	0.022:0.319	-0.004:0.177	-0.075:0.248	-0.0:0.276
Team Specialization	-0.082:0.134	-0.056:0.114	-0.072:0.157	-0.051:0.064	-0.136:0.05
Team Credibility	-0.122:0.035	0.048:0.321	-0.007:0.203	-0.011:0.194	-0.099:0.021
Team Coordination	-0.114:0.064	0.113:0.394	-0.016:0.246	-0.054:0.186	-0.046:0.157
Team Transactive Memory Total	-0.11:0.09	0.07:0.343	-0.008:0.263	-0.044:0.17	-0.05:0.068
Psychological Safety	-0.052:0.216	0.022:0.251	-0.011:0.219	-0.028:0.165	-0.046:0.045
Team Viability	-0.202:0.058	0.127:0.448	0.093:0.46	-0.05:0.199	-0.063:0.101
Team Satisfaction	-0.127:0.087	0.117:0.446	0.022:0.396	-0.027:0.237	-0.045:0.14
Team Potency	-0.167:0.126	0.209:0.49	-0.015:0.342	-0.073:0.12	-0.053:0.047
Team Effectiveness by Member	-0.13:0.143	0.141:0.452	0.021:0.345	-0.036:0.254	-0.028:0.077

Fig. 7. Bootstrapped difference in Pearson R values between the Swarm and the Survey Variance methods of team personality assessments

Appendix C: Multivariate Correlation Analysis: Bootstrapping Results

Dimensions: Average	90% CI Lower Bound	90% CI Upper Bound	P-Value	More Significant
Task Attraction	0.13	0.42	0.00	Swarm
Interpersonal Cohesiveness	0.10	0.38	0.01	Swarm
Task Commitment	0.15	0.41	0.00	Swarm
Cohesiveness Total	0.16	0.46	0.00	Swarm
Relationship Conflict	0.06	0.32	0.03	Swarm
Task Conflict	-0.05	0.18	0.22	Inconclusive
Process Conflict	0.05	0.30	0.03	Swarm
Conflict Total	0.05	0.29	0.04	Swarm
Team Special	-0.10	0.08	0.51	Inconclusive
Team Credit	-0.03	0.18	0.16	Inconclusive
Team Coordination	0.09	0.29	0.01	Swarm
Team Total	0.01	0.19	0.08	Swarm
Team Satisfaction	0.19	0.42	0.00	Swarm
Team Potency	0.16	0.35	0.00	Swarm

Fig. 8. Bootstrapped difference in Multivariate Regression Adjusted-R values between the Swarm and Group Average Assessment Methods

Dimensions: Minimum	90% CI Lower Bound	90% CI Upper Bound	P-Value	More Significant
Cohesiveness Task Attraction	0.11	0.35	0.01	Swarm
Interpersonal Cohesiveness	0.09	0.41	0.02	Swarm
Task Commitment	0.12	0.35	0.00	Swarm
Cohesiveness Total	0.13	0.39	0.00	Swarm
Relationship Conflict	-0.06	0.33	0.20	Inconclusive
Task Conflict	-0.09	0.15	0.38	Inconclusive
Process Conflict	-0.06	0.23	0.28	Inconclusive
Conflict Total	-0.06	0.26	0.23	Inconclusive
Team Special	-0.13	0.05	0.69	Inconclusive
Team Credit	-0.04	0.20	0.19	Inconclusive
Team Coordination	0.07	0.26	0.02	Swarm
Team Total	-0.03	0.16	0.17	Inconclusive
Team Satisfaction	0.13	0.43	0.00	Swarm
Team Potency	0.05	0.24	0.04	Swarm

Fig. 9. Bootstrapped difference in Multivariate Regression Adjusted-R values between the Swarm and Group Minimum Assessment Methods

Dimensions: Maximum	90% CI Lower Bound	90% CI Upper Bound	P-Value	More Significant
Cohesiveness Task Attraction	0.10	0.47	0.00	Swarm
Interpersonal Cohesiveness	0.05	0.42	0.01	Swarm
Task Commitment	0.10	0.45	0.00	Swarm
Cohesiveness Total	0.12	0.50	0.00	Swarm
Relationship Conflict	-0.01	0.34	0.06	Inconclusive
Task Conflict	0.01	0.25	0.04	Swarm
Process Conflict	0.00	0.33	0.06	Inconclusive
Conflict Total	0.02	0.34	0.03	Swarm
Team Special	-0.16	0.13	0.45	Inconclusive
Team Credit	-0.13	0.19	0.36	Inconclusive
Team Coordination	0.06	0.33	0.01	Swarm
Team Total	-0.03	0.24	0.11	Inconclusive
Team Satisfaction	0.13	0.49	0.00	Swarm
Team Potency	0.16	0.43	0.00	Swarm

Fig. 10. Bootstrapped difference in Multivariate Regression Adjusted-R values between the Swarm and Group Maximum Assessment Methods