Quantifying Collective Intelligence in Human Groups

Christoph Riedl[1,2,3,4,7,8,*], Young Ji Kim[5], Pranav Gupta[6], Thomas W. Malone[7,8], Anita Williams Woolley[6]

[1]D'Amore-McKim School of Business, Northeastern University, Boston, MA

[2]Khoury College of Computer Sciences, Northeastern University, Boston, MA

[3]Network Science Institute, Northeastern University, Boston, MA

[4]Institute for Quantitative Social Science, Harvard University, Cambridge, MA

[5]Department of Communication, University of California, Santa Barbara, Santa Barbara, CA

[6]Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA

[7]MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA

[8]MIT Center for Collective Intelligence, Massachusetts Institute of Technology, Cambridge, MA

*c.riedl@neu.edu

Collective intelligence is critical to solving many scientific, business, and other problems, but groups often fail to achieve it. Here we analyze data on group performance from 22 studies, including 5,349 individuals in 1,356 groups. Our results support the conclusion that a robust collective intelligence factor characterizes a group's ability to work together across a diverse set of tasks. We further show that collective intelligence is predicted by the average social sensitivity of group members and the proportion of females in the group, and that it predicts performance on various out-of-sample criterion tasks. We also find that group performance on tasks involving selection among alternatives is better predicted by the skill of individual group members while performance on idea generation tasks is better predicted by the group interaction process.

Significance Statement

Collective intelligence is critical to solving many scientific, business, and other problems. We find strong support for a general factor of collective intelligence using meta-analytic methods in a dataset comprising 22 studies, including 5,349 individuals in 1,356 groups. CI can predict performance in a range of out-of-sample criterion tasks. CI, in turn, is predicted by individual skill, group composition, and group interaction processes. Group performance on tasks involving selection among alternatives is better predicted by the skill of individual group members while performance on idea generation tasks is better predicted by the group interaction process. The proportion of women in a group is a significant predictor of group performance, mediated by social perceptiveness.

Introduction

Science relies increasingly on teams, and more and more problems in our society—from addressing climate change to curing diseases and developing complex technologies—can only be solved by the work of groups (*1–3*). Thus the ability of groups to function at a high level is critically important for many aspects of our well-being and our collective capacity to conduct research (*4*).

In one previous approach to understanding the determinants of group performance, Woolley and colleagues (*4*) used an analogy between the individual intelligence of a person and the collective intelligence of a group. The most common way of operationally defining individual intelligence in the research literature is with a statistical factor (often called "g" for general intelligence) that predicts how well a person will perform on a wide range of different tasks (*5*). Woolley and colleagues found that, just as for individuals, there is also a single statistical factor for a group that describes the group's capability to perform many different tasks. More precisely, they found that, in a factor analysis of group performance on a number of tasks, a single factor predicted over 40% of the variance in performance on all of the tasks. They called this factor *collective intelligence* (CI), which they defined as a group's ability to perform a wide variety of tasks. They also found that a group's collective intelligence was correlated with, not only the individual intelligence of the group members, but also the average social sensitivity of the group members and the proportion of females in the group.

A common question about this work is whether "intelligence" could even be a true property of a group in the same way it is of an individual. Of course, one could *define* intelligence as something that only individual humans and not groups (or computers) could have. But most definitions of intelligence focus on the *capabilities* of intelligent entities (*6*, *7*). For instance, one widely cited definition of individual intelligence is "a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (*8*). These are certainly things that groups can do, too, so by this and many similar definitions of intelligence, it is reasonable to consider groups intelligent. In fact, there is a growing literature on how groups can perform various processes such as learning (*9–11*), remembering (*12*), making decisions (*13*), and solving problems (*14*, *15*). And it is sensible to say that the intelligence of a group—like that of an individual—emerges from the interaction of these processes (*7*, *16*, *17*).

A related question is whether there is any underlying causal property of groups that is analogous to the biological mechanisms in a human brain that give rise to "g." Many people do not realize, however, that even for researchers studying the biology of the brain, individual intelligence is still largely a statistical observation. As Haier summarizes, "There is overwhelming empirical evidence that intelligence is best described by a general factor that is common among all tests of mental ability … The *g*-factor … can only be estimated, usually as a latent variable extracted from a battery of tests, and best interpreted for a person as a percentile compared to other individuals" (*18*). In other words, over a hundred years after Spearman discovered the "g" factor, there is still no clear consensus among researchers about what specific biological mechanisms cause some brains to be more intelligent than others (*18*). So the fact that researchers do not yet have detailed causal models of the processes underlying collective intelligence does not mean that the phenomenon does not exist. It just means that further scientific work is needed, and that is one of the goals of the work presented here.

Since the work by Woolley at al (*4*), a number of other studies have replicated or confirmed the initial results about a collective intelligence factor in human groups (*10*, *19–21*). However, others have questioned these results (*22–24*) and reported disparate results regarding the strength of individual skill in predicting group collective intelligence (*25*).

Here, we provide robust evidence of a single collective intelligence factor using accumulated data from 22 different samples, involving 5,349 individuals working together in 1,356 groups. We present an analysis of these data drawn from different populations working together in a variety of settings (online, face-to-face, groups of friends, strangers, etc.; see *SI Appendix* Tables S1-S2). We analyze different combinations of tasks in different samples, all focused on exploring the strength of the inter-item correlations and resulting evidence of a general collective intelligence factor (Fig. 1).

Furthermore, we quantify indicators of known correlates of a group's collective intelligence (individual member skill and group interaction processes) that are drawn from the fine-grained process data available on our research platform. We discuss how our findings can help create situations that reliably foster high collective intelligence. In addition, our novel method for capturing granular, process-level data paves the way for researchers to build testable causal theories of collective intelligence grounded in robust behavioral indicators.

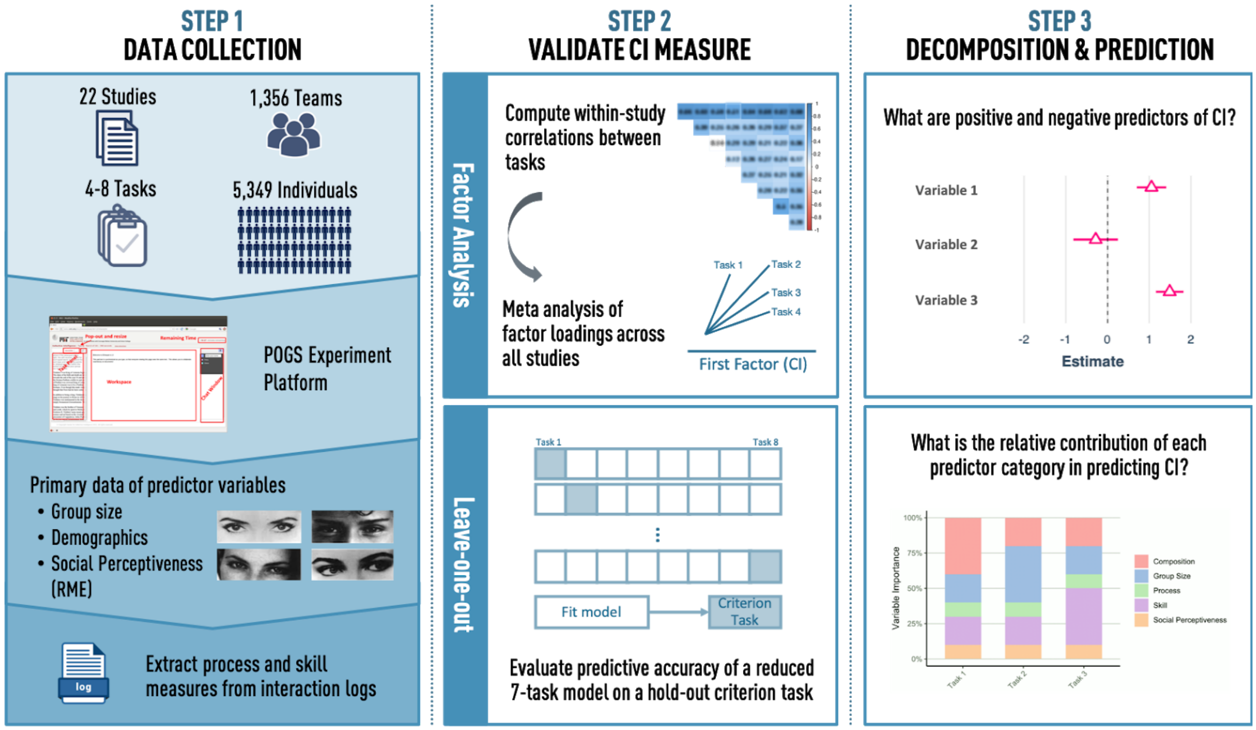

Fig. 1. Study process. Step 1: Using the Platform for Online Group Studies (POGS) we conducted 22 studies involving 1,356 groups, 5,349 individuals, and 4-8 tasks per study. Step 2: We perform meta-analytic factor analysis (across the 22 studies) and leave-one-out analysis to test the robustness of evidence of a general collective intelligence factor that explains a group's performance. Step 3: We use a variety of predictors, including demographics, skill, social perceptiveness, and process measures to predict collective intelligence and to assess the relative predictive power of each set of predictors.

To systematically collect data from both co-located and distributed groups, we have developed an online tool, called the Platform for Online Group Studies (POGS; *26*). This platform enables participants to see the input of other group members in real time and work with them in a collaborative editing application similar to Google Docs. Through POGS, we were able to administer the battery of group tasks in a standardized way with the same instructions, time constraints, and user interface for all groups in all studies.

We have been gradually developing different tasks to include in the measurement, working toward refining a stable collective intelligence test battery (see *SI Appendix*). To date we have tested several types of tasks allowing us to capture 176 individual task measurements clustered into 8 task types in 22 different samples (*SI Appendix* Section S1.1). The samples included various populations from university students to crowd workers to military personnel to online gamers (*SI Appendix* Table S1, S2). In all 22 studies, we used POGS to administer the tasks. In selecting tasks, we sampled from existing taxonomies of group tasks to ensure we had variation in task types. For instance, we included tasks requiring groups to *generate* a range of different ideas, to *choose* one correct answer from among different options, or *execute* specified tasks as quickly and accurately as possible (*15*, *27*, *28*).

Our data contains measures of 1,356 groups (5,349 individuals, with groups ranging between 2-7 individuals; *SI Appendix* Table S1). Three task types (Brainstorm Object, Typing Text, and Unscramble Words) were used in all studies. Seven studies, including the largest individual study with 254 groups (Field Sample 1), administered all eight task types.

To analyze the data from 22 individual studies, we combine meta-analytic techniques (see *Materials and Methods*) with the analysis of primary data. The meta-analytic approach allows us to combine evidence across all our studies, even as the studies administered different subsets of the eight task types. This meta-analytic approach allows us to account for within study error (which depends, for example, on the number of observations in each study, *29*).

Results

In the first stage of the meta-analytic analysis, we compute the within-study correlation coefficients across all tasks. This first step yields raw estimates of pooled correlations between group scores on different tasks. In examining these correlations (Fig. 2A), we observe an average inter-item correlation of 0.27 [0.12-0.50].

In the second stage of the analysis, we fit a single factor structural model (Fig. 2B). We fit a random effects model, which uses weighted least squares to weigh the precision of the pooled correlation based on the number of observations in each study. The fit statistics of the single factor model are excellent (see *SI Appendix*). The standardized factor loadings range between 0.26 and 0.52 (all $p < 0.001$) with average variance extracted of 43.6%. This means that the data across 22 studies support the one-factor structure of collective intelligence reported previously (*4*).

We then use the weights (factor loadings) of the single collective intelligence factor of the meta-analysis to compute corresponding *CI Scores* for all groups in our dataset. We compute a group's *CI Score* as the average weighted z-score of all tasks used in the study (*30*). Consistent with convention in individual intelligence tests, we then multiply the scores by 15 and add 100.

As a test of the power of our *CI Scores* in predicting a group's performance on a group task, we performed a variation of leave-one-out cross-validation—a method commonly used in machine learning to evaluate the predictive performance of models and avoid over-fitting these models to the available data (*31*). To do this, we compute the *CI Scores* for a group's performance eight times, each time leaving out a different one of the eight tasks, so that the resulting model is estimated based on the data of only seven remaining tasks. Then we use these *CI Scores* to predict performance on the tasks that were left out (Fig. 2C). In other words, we repeat the full two-stage meta analysis eight times, each time excluding a different task.

We find that these restricted *CI Scores* are strong predictors of a group's performance on the left-out tasks (average Pearson correlation of 0.39 [0.26-0.53]; all $p < 0.001$). All effect sizes are between medium and large (*32*). This provides strong evidence that a single factor *CI score* computed for any seven of the tasks is a reliable predictor of group performance on the remaining task.

Having established CI as a reliable measure of a group's ability to perform a wide range of tasks, we turn to an exploratory analysis of what predicts CI using primary data of predictors, and the meta-analytic dependent variable (*CI Score*; Fig. 2D). In previous studies, all the following aspects of a group's composition have been shown to predict CI: group size, average or maximum individual ability, proportion of female group members, and a group's average social perceptiveness measured using the *Reading the Mind in the Eyes* test (*33*). We find a positive effect of the proportion of women and the level of social perceptiveness, where social perceptiveness mediates the effect of proportion of women on CI, consistent with prior studies (*4*, *19*; *SI Appendix* Table S8). We find negative effects associated with high levels of age diversity, suggesting that this form of diversity impedes collaboration.

Since all tasks are computer-mediated through the POGS system, we are also able to quantify a rich set of process measures. Specifically, we measure estimates of both individual group members' skill on the different tasks performed, and we capture three specific aspects of the groups' process: skill congruence, strategy, and effort (see *SI Appendix* for description of process measure calculations). These critical group process attributes were first identified as essential to group performance in seminal work by Hackman (*35*). *Skill congruence* gauges a group's proficiency at achieving agreement between relative member skills and their contributions to work on a task; *strategy* captures a group's ability to coordinate their work to accomplish coverage of the entire task; and *effort* captures the total amount of activity members contribute to task completion. In our data, skill congruence and strategy are strong positive predictors of group performance, while effort is not a significant predictor of group performance (Fig. 2D Model 4). Once all of our composition and process variables are in the model we see that the individual skill is more predictive of *CI* than either effort, strategy, or skill congruence are on their own (for an analysis of individual tasks see *SI Appendix* Table S6).

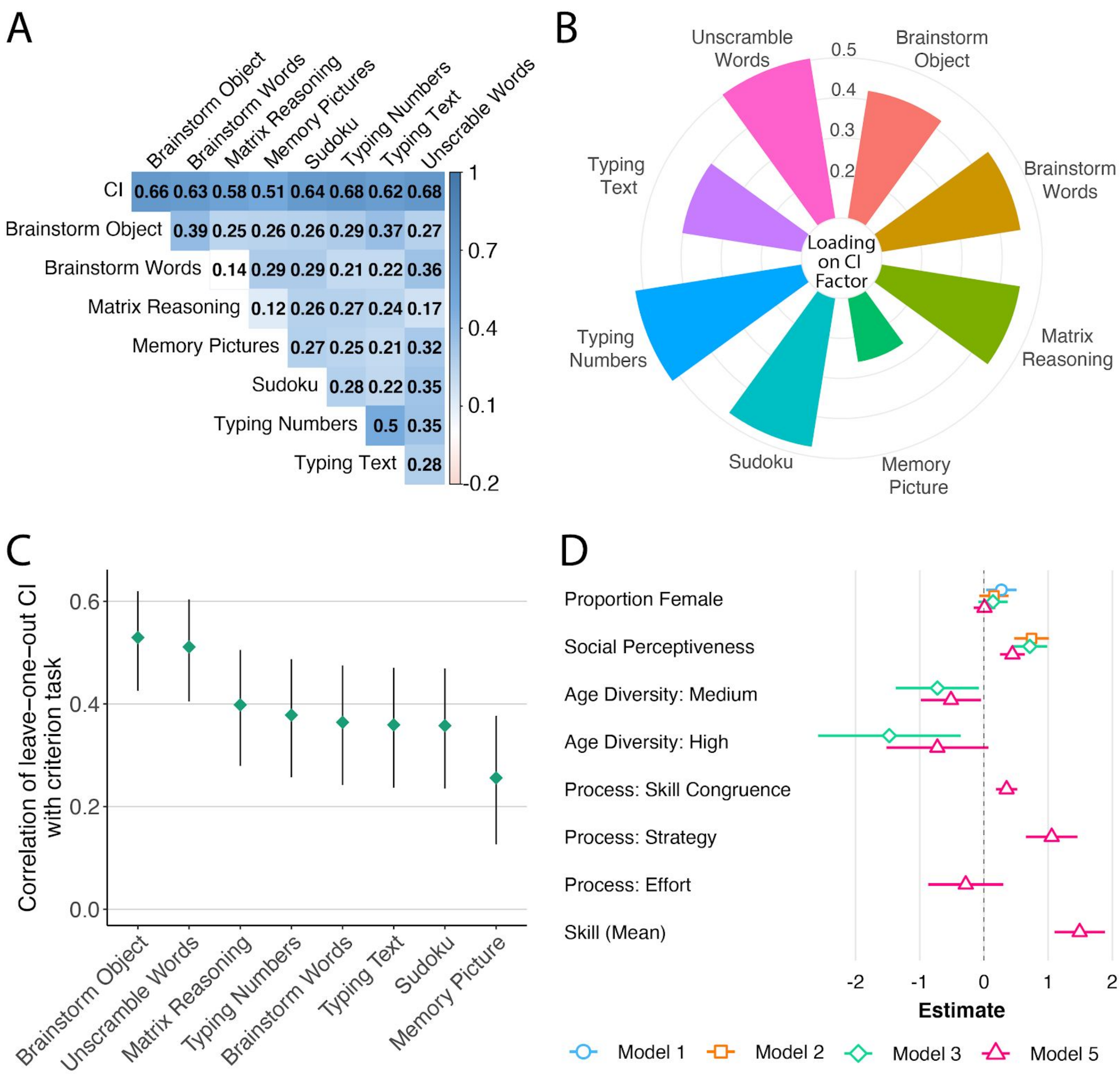

Fig. 2. Collective Intelligence factor analysis and prediction. (A) Raw Pearson correlations between tasks and CI of the pooled data (all correlations are significant with at least $p < 0.026$). **(B)** Standardized factor loadings of the meta-analysis of each task on the first factor (Collective Intelligence). **(C)** Treating each of the eight tasks as a criterion task, we repeat the meta factor analysis (using the remaining seven tasks) to compute a restricted CI factor and predict the excluded criterion task (Pearson correlation with 95% confidence interval). **(D)** Regression coefficients for four different linear models predicting CI. Proportion of female group members is a significant predictor in models that do not control for Social Perceptiveness (showing coefficients from *SI Appendix* Table S5).

The next natural question is: What is the relative contribution of each group of variables—individual skill, group process, and group composition—to CI? We use random forests (*36*) to investigate the relative importance of these variables (Fig. 3). A key advantage of this data-driven machine learning method over the regression-based approach is that it does not depend on the order in which variables are entered into a stepwise model and that it accounts for non-linear and complex relationships between the variables. The largest proportion of variation in CI is explained by our group of process measures (congruence, strategy, and effort), followed by individual member skill (measured as the group mean and maximum). Substantially less variation is explained by group size, followed by social perceptiveness, and group composition (proportion female and age diversity). The importance of group process versus skill varies by task (Fig. S1). For example, more than 51% of the explained variation in performance on Sudoku is due to individual member skill, with group processes playing a smaller role, while for the Unscrambling Words tasks, 55% of the variation in performance is due to group processes, with individual skill playing a smaller role.

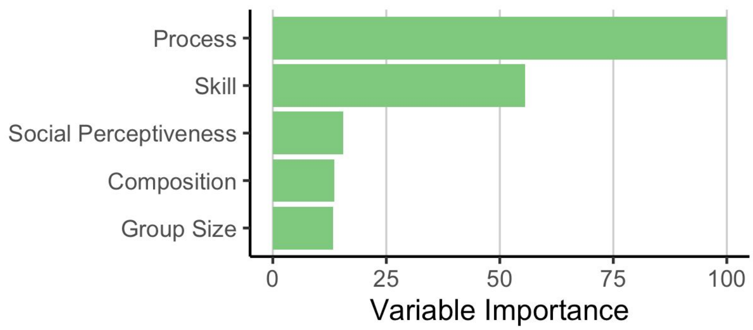

Fig. 3. Variable importance in predicting collective intelligence. Variable importance based on a random forest prediction model computed as the decrease of accuracy in prediction when a given variable is excluded from the model and expressed relative to the maximum.

Discussion

In conclusion, we interpret our data from 22 different samples encompassing over 5,000 participants in over 1,300 groups across a diverse set of contexts, populations, and tasks to strongly support a general factor of collective intelligence, with important theoretical and empirical implications for future research. In addition, with the wealth of fine-grained behavioral data we collected, we were able to quantify the relative contributions of individual skill and group process in predicting CI.

This work advances the science of collective performance in two principal ways. The first is methodological. Typically, in research on groups, performance is operationalized with a single variable or task type, and we know that single item measures are inherently less stable or generalizable (*25*). A metric of collective intelligence, based on a variety of tasks, provides a more generalizable measure than typical measures of group performance.

The second advance is conceptual. Existing research does not typically distinguish between capability (i.e., a group's potential) and performance (i.e., potential actualized). However, we argue that we are capturing a group's capability to work together, which is enabled by humans' ability to form mental representations of the intentions and goals of others which they can use to optimize mutual interaction (*9*). Under this view, the notion of group-level intelligence fundamentally captures both the intellect of the individuals as well as the effective alignment of the individuals' activities and beliefs (*41*). This capability, in turn, predicts performance. Separating the concept and measurement of capability from performance is important, as performance in a particular setting at a particular point in time is an imprecise gauge of capability, since (particularly in field settings) it is influenced by a number of factors outside of a group's capability to work together, such as competition, opportunities, or resources.

Taken together, these two advances mean that using CI as a metric of group capability can provide more reliable ways of measuring the effects of interventions than typical laboratory studies (due to the incorporation of multiple measures) and without the confounds that can come with performance metrics gathered in field settings. Therefore this work provides both a method and a concept for guiding research to advance the science of collective performance.

To further this point, an important additional insight from the analysis presented is that we see wide variability in the degree to which group process versus composition are strong predictors of performance on any particular task. This builds on classic work on task types and process loss in groups (*28*). For instance, we see that group composition is a bigger contributor to closed-ended *choose*-type tasks such as Sudoku, while *generate* tasks such as brainstorming and *execute* tasks such as typing are more influenced by the quality of group interaction process. Choose-type tasks tend to be solved with a single, demonstrably correct answer requiring just one person in the group to arrive at it. By contrast, generate-type tasks tend to be additive, meaning that better coordinated contributions from more people will yield better performance.

Our work builds on those distinctions by taking them one step further and quantifying more precisely the relative contribution of different group processes and individual skills to performance on each task. Using this approach, researchers could then characterize tasks on the basis of the quantified contribution of group process versus individual skill. This new understanding could help consolidate the existing literature by allowing researchers to more precisely specify the types of tasks to which their findings generalize.

The variation in relative contribution of group process and skill to task performance also suggests that disparate findings in the literature, such as those regarding the strength of individual skill in predicting collective intelligence (*25*), are likely explained by the selection of tasks used to measure group performance or CI. Furthermore, methodological choices that constrain the interaction processes of groups, such as restricting the number of members who can record group answers (*24*, *37*), would obscure group capability on tasks that are highly reliant on group process, and depress the correlation among performance scores of those tasks with other types of tasks (since they do not fully reflect the capability of the group or its members). Taken together, these analyses suggest that a number of small but important methodological choices researchers make likely influence the extent to which findings of different studies in this literature replicate or diverge.

It has taken over a century to develop detailed causal theories about the biological basis of individual intelligence, and the task is still far from complete. We believe that similar work remains to be done in linking the phenomenon of collective intelligence to existing and new theories of group performance. We believe that some such links are already clear. For instance, our results help clarify the question of the effect of gender composition on group performance by showing that the correlation between collective intelligence and the proportion of females in the group is mediated by the social perceptiveness of the group members. This result can be explained by previous research showing that women, on average, have scores higher than men on the tests of social perceptiveness (*33*). In addition, existing studies have sought to quantify the contribution of individual cognitive ability to group performance (e.g., *38*, *39*) and we complement this work with more focused skill-based measures and the ability to examine variation in contribution across tasks of different types together with detailed process measures.

Our results also have implications for managerial practice. Human resource managers today often focus on evaluating *individuals*, but the work presented here suggests ways to systematically evaluate *groups* as well. General results, for instance, like those above about the positive effects of social perceptiveness and the negative effects of age diversity, have implications for how to select promising *combinations* of people for teams. More interestingly, companies might also give collective intelligence tests to their internal teams and use the results as early indicators to intervene in various ways. If a team performed poorly, for instance, managers might change some of the people on the team or provide external coaching. And teams that performed well might be given more important assignments.

Furthermore, given the relative importance of group processes we see in our data, our study suggests detailed ways of scaffolding the interaction processes of the group via facilitation or technological aids. For example, giving members feedback regarding relative member effort (*40*) or nudging them toward more effective group coordination strategy (*31*) might enable groups to gain better leverage from the knowledge and skills of their members. Importantly, our analyses suggest that the impact on performance from changes in CI can be substantial. Other things being equal, a group with one standard deviation higher CI would increase task performance by 18%, plus or minus about 11% (see section S4.7). In summary, our research suggests that groups can be characterized by a quantifiable form of collective intelligence, that can yield substantial benefits in many important contexts. And building a better science of collective intelligence will enable us to more effectively advance the performance of groups working on the complex and critical issues that threaten our society the most.

Materials and Methods

We performed meta-analytic confirmatory factor analysis (CFA) using two-stage structural equation modeling (TSSEM) following the approach developed by Cheung and colleagues (13, 14). The first stage applies multi-group structural equation model (SEM) to pool correlation matrices. Two diagnostic test statistics suggest that a random-effects model is most appropriate to aggregate the correlation matrices in the first stage. In the second stage of the meta-analysis, we fit a single factor structural model (*SI Appendix* Table S4). The model is fit using weighted least squares to weigh the precision of the pooled correlation (based on the number of observations in each study) in the second stage of analysis. Each element of the pooled correlation table can thus be weighted based on the exact sample sizes available for each element. We explore different models, all of which support the one-factor structure reported here. We use the factor loadings from this one-factor model to compute CI Scores for each group (see *SI Appendix* for detailed equations). For the remainder of the analyses, we then rely on ordinary least squares regression and Pearson correlation coefficients to predict performance and quantify strengths of correlations.

References

1. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science*. **316**, 1036–1039 (2007).
2. L. Wu, D. Wang, J. A. Evans, Large teams develop and small teams disrupt science and technology. *Nature*. **566**, 378–382 (2019).
3. E. Bernstein, J. Shore, D. Lazer, How intermittent breaks in interaction improve collective intelligence. *Proc. Natl. Acad. Sci.*, 201802407 (2018).
4. A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, T. W. Malone, Evidence for a collective intelligence factor in the performance of human groups. *Science*. **330**, 686–688 (2010).
5. C. Spearman, General intelligence, objectively determined and measured. *Am. J. Psychol.* **15**, 201 (1904).
6. S. Legg, M. Hutter, A collection of definitions of intelligence. *Front. Artif. Intell. Appl.* **157**, 17–24 (2007).
7. T. W. Malone, *Superminds: The suprising power of people and computers thinking together* (Little, Brown and Company, New York, 2018).
8. L. S. Gottfredson, Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*. **24**, 13–23 (1997).
9. S. P. Veissière, A. Constant, M. J. Ramstead, K. J. Friston, L. J. Kirmayer, Thinking through other minds: A variational approach to cognition and culture. *Behav. Brain Sci.* **43** (2020).
10. I. Aggarwal, A. W. Woolley, C. F. Chabris, T. W. Malone, The impact of cognitive style

diversity on implicit learning in teams. *Front. Psychol.*, 112 (2019).
11. A. W. Woolley, I. Aggarwal, in *Handbook of Group and Organizational Learning*, L. Argote, J. M. Levine, Eds. (Oxford University Press, London, UK, 2020), pp. 491–506.
12. Y. Ren, L. Argote, Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *Acad. Manag. Ann.* **5**, 189–229 (2011).
13. J. H. Davis, Group decision and social interaction: A theory of social decision schemes. *Psychol. Rev.* **80**, 97–125 (1973).
14. P. R. Laughlin, Social combination processes of cooperative problem-solving groups on verbal intellective tasks. *Prog. Soc. Psychol.* **1**, 127–155 (1980).
15. J. R. Larson, *In Search of Synergy in Small Group Performance* (Psychology Press, New York, NY, 2010).
16. A. R. Luria, *The working brain: An introduction to neuropsychology* (Basic Books, New York, 1973).
17. A. T. Mayo, A. W. Woolley, Variance in group ability to transform resources into performance, and the role of coordinated attention. *Acad. Manag. Discov.* (in press).
18. R. J. Haier, The biological basis of intelligence. *Camb. Handb. Intell.* (Cambridge University Press, Cambridge, UK), 942 (2020).
19. D. Engel, A. W. Woolley, L. X. Jing, C. F. Chabris, T. W. Malone, Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PLoS ONE*. **9**, e115212 (2014).
20. D. Engel, A. W. Woolley, I. Aggarwal, C. F. Chabris, M. Takahashi, K. Nemoto, C. Kaiser, Y. J. Kim, T. W. Malone, Collective intelligence in online collaboration emerges in different contexts and cultures. *CHI 15 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.* (2015).
21. N. Meslec, I. Aggarwal, P. L. Curşeu, The insensitive ruins it all: Compositional and compilational influences of social sensitivity on collective intelligence in groups. *Front. Psychol.* (2016).
22. M. Credé, G. Howardson, The structure of group task performance—A second look at "collective intelligence": Comment on Woolley et al. (2010). **102**, 1483–1492 (2017).
23. A. W. Woolley, Y. Kim, T. W. Malone, "Measuring Collective Intelligence in Groups: A Reply to Credé and Howardson" (SSRN Scholarly Paper ID 3187373, Social Science Research Network, Rochester, NY, 2018), (available at https://papers.ssrn.com/abstract=3187373).
24. J. B. Barlow, A. R. Dennis, Not As Smart As We Think: A Study of Collective Intelligence in Virtual Groups. *J. Manag. Inf. Syst.* **33**, 684–712 (2016).
25. T. C. Bates, S. Gupta, Smart groups of smart people: Evidence for IQ as the origin of collective intelligence in the performance of human groups. *Intelligence*. **60**, 46–56 (2017).
26. Y. J. Kim, D. Engel, A. W. Woolley, J. Lin, N. McArthur, T. W. Malone, What makes a strong team? Using collective intelligence to predict performance of teams in League of Legends. *Proc. 20th ACM Conf. Comput.-Support. Coop. Work Soc. Comput. CSCW 2017* (2017).
27. J. E. McGrath, *Groups: Interaction and performance* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
28. I. D. Steiner, *Group process and productivity* (Academic Press, New York, 1972).
29. M. Borenstein, L. V. Hedges, J. P. T. Higgins, H. R. Rothstein, *Introduction to*

meta-analysis (John Wiley & Sons, Chichester, England, 2009).
30. P. Gupta, Y. J. Kim, E. Glikson, A. W. Woolley, Digitally nudging team processes to enhance collective intelligence. *Proc. Collect. Intell. 2019* (2019).
31. I. H. Witten, E. Frank, *Data Mining: Practical machine learning tools and techniques* (Morgan Kaufmann, New York, NY, 2005).
32. J. Cohen, A power primer. *Psychol. Bull.* **112**, 155 (1992).
33. S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry*. **42**, 241–251 (2001).
34. C. Riedl, P. Gupta, Y. J. Kim, T. W. Malone, A. W. Woolley, Supplementary material for quantifying collective intelligence in human groups. *Proc. Natl. Acad. Sci.* (under review).
35. J. R. Hackman, in *Handbook of organizational behavior*, J. W. Lorsch, Ed. (Prentice Hall, Englewood Cliffs, NJ, 1987), pp. 315–342.
36. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
37. N. Hashmi, thesis, Massachusetts Institute of Technology (2017).
38. D. J. Devine, J. L. Philips, Do smarter teams do better? A meta-analysis of cognitive ability and team performance. *Small Group Res.* **32**, 507–532 (2001).
39. J. A. LePine, J. A. Colquitt, A. Erez, Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Pers. Psychol.* **53**, 563–593 (2000).
40. E. Glikson, A. W. Woolley, P. Gupta, Y. J. Kim, Visualized automatic feedback in virtual teams. *Front. Psychol.* (2019).
41. J. Vasil, P. B. Badcock, A. Constant, K. Friston, M. J. Ramstead, A World Unto Itself: Human Communication as Active Inference. *Front. Psychol.* **11**, 417 (2020).