task (correct, error, inserted error, and corrected error) to allow typists to distinguish sources of errors and correct responses and, therefore, provide a stronger test of illusions of authorship. We asked 24 skilled typists (WPM = 70.7 ± 16.4) to type 600 words, each of which was followed by a four-alternative explicit report screen. Typists typed 91.8% of the words correctly. Mean interkeystroke intervals, plotted in Fig. 3A, show post-error slowing for incorrect responses ($F_{1,138} = 117.7$, $p < 0.01$) and corrected errors ($F_{1,138} = 120.0$, $p < 0.01$), but not for inserted errors ($F < 1.0$), indicating that inner-loop detection distinguishes between actual errors and correct responses.

Explicit detection probabilities, plotted in Fig. 3B, show good discrimination between correct and error responses. For correct responses, typists said "correct" more than "error" [$t(23) = 97.29$, $p < 0.01$]; for error responses, typists said "error" more than "correct" [$t(23) = 8.22$, $p < 0.01$]. Typists distinguished actual errors from inserted errors well, avoiding an illusion of authorship. They said "error" more than "inserted" for actual errors [$t(23) = 7.06$, $p < 0.01$] and "inserted" more than "error" for inserted errors [$t(23) = 14.75$, $p < 0.01$]. However, typists showed a strong illusion of authorship with corrected errors. They were just as likely to call them correct responses as corrected errors [$t(23) = 1.38$].

The post-error slowing and post-trial report data show a dissociation between inner- and outer-loop error detection. We assessed the dissociation further by comparing post-error slowing on trials in which typists did and did not experience illusions of authorship (*21*). The pattern of post-error slowing was the same for both sets of trials (fig. S6), suggesting that the pattern in Fig. 3A is representative of all trials.

The three experiments found strong dissociations between explicit error reports and post-error slowing. These dissociations are consistent with the hierarchical error-detection mechanism that we proposed, with an outer loop that mediates explicit reports and an inner loop that mediates post-error slowing. This nested-loop description of error detection is consistent with hierarchical models of cognitive control in typewriting (*9*, *10*, *15*–*17*) and with models of hierarchical control in other complex tasks (*2*, *8*, *22*). Speaking, playing music, and navigating through space may all involve inner loops that take care of the details of performance (e.g., uttering phonemes, playing notes, and walking) and outer loops that ensure that intentions are fulfilled (e.g., messages communicated, songs performed, and destinations reached). Hierarchical control may be prevalent in highly skilled performers who have had enough practice to develop an autonomous inner loop. Previous studies of error detection in simple tasks may describe inner-loop processing. The novel contribution of our research is to dissociate the outer loop from the inner loop.

The three experiments demonstrate cognitive illusions of authorship in skilled typewriting (*11*–*14*). Typists readily take credit for correct output on the screen, interpreting corrected errors as their own correct responses. They take the blame for inserted errors, as in the first and second experiments, but they also blame the computer, as in the third experiment. These illusions are consistent with the hierarchical model of error detection, with the outer loop assigning credit and blame and the inner loop doing the work of typing (*10*, *17*). Thus, illusions of authorship may be a hallmark of hierarchical control systems (*2*, *11*, *22*, *23*).

References and Notes

1. P. M. A. Rabbitt, *J. Exp. Psychol.* **71**, 264 (1966).
2. D. A. Norman, *Psychol. Rev.* **88**, 1 (1981).
3. C. B. Holroyd, M. G. H. Coles, *Psychol. Rev.* **109**, 679 (2002).
4. N. Yeung, M. M. Botvinick, J. D. Cohen, *Psychol. Rev.* **111**, 931 (2004).
5. W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, E. Donchin, *Psychol. Sci.* **4**, 385 (1993).
6. S. Dehaene, M. I. Posner, D. M. Tucker, *Psychol. Sci.* **5**, 303 (1994).
7. C. S. Carter *et al.*, *Science* **280**, 747 (1998).
8. K. S. Lashley, in *Cerebral Mechanisms in Behavior*, L. A. Jeffress, Ed. (Wiley, New York, 1951), pp. 112–136.
9. T. A. Salthouse, *Psychol. Bull.* **99**, 303 (1986).
10. G. D. Logan, M. J. C. Crump, *Psychol. Sci.* **20**, 1296 (2009).
11. T. I. Nielsen, *Scand. J. Psychol.* **4**, 225 (1963).
12. M. M. Botvinick, J. D. Cohen, *Nature* **391**, 756 (1998).
13. D. M. Wegner, *The Illusion of Conscious Will* (MIT Press, Cambridge, MA, 2002).
14. G. Knoblich, T. T. J. Kircher, *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 657 (2004).
15. D. E. Rumelhart, D. A. Norman, *Cogn. Sci.* **6**, 1 (1982).
16. L. H. Shaffer, *Psychol. Rev.* **83**, 375 (1976).
17. X. Liu, M. J. C. Crump, G. D. Logan, *Mem. Cognit.* **38**, 474 (2010).
18. A. M. Gordon, J. F. Soechting, *Exp. Brain Res.* **107**, 281 (1995).
19. J. Long, *Ergonomics* **19**, 93 (1976).
20. P. Rabbitt, *Ergonomics* **21**, 945 (1978).
21. Materials and methods are available as supporting material on *Science* Online.
22. M. M. Botvinick, *Trends Cogn. Sci.* **12**, 201 (2008).
23. R. Cooper, T. Shallice, *Cogn. Neuropsychol.* **17**, 297 (2000).

24. We thank J. D. Schall for comments on the manuscript. This research was supported by grants BCS 0646588 and BCS 0957074 from the NSF.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6004/683/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
References

5 April 2010; accepted 13 September 2010
10.1126/science.1190483

Evidence for a Collective Intelligence Factor in the Performance of Human Groups

Anita Williams Woolley,[1]* Christopher F. Chabris,[2,3] Alex Pentland,[3,4] Nada Hashmi,[3,5] Thomas W. Malone[3,5]

Psychologists have repeatedly shown that a single statistical factor—often called "general intelligence"—emerges from the correlations among people's performance on a wide variety of cognitive tasks. But no one has systematically examined whether a similar kind of "collective intelligence" exists for groups of people. In two studies with 699 people, working in groups of two to five, we find converging evidence of a general collective intelligence factor that explains a group's performance on a wide variety of tasks. This "*c* factor" is not strongly correlated with the average or maximum individual intelligence of group members but is correlated with the average social sensitivity of group members, the equality in distribution of conversational turn-taking, and the proportion of females in the group.

As research, management, and many other kinds of tasks are increasingly accomplished by groups—working both face-to-face and virtually (*1*–*3*)—it is becoming ever more important to understand the determinants of group performance. Over the past century, psychologists made considerable progress in defining and systematically measuring intelligence in individuals (*4*). We have used the statistical approach they developed for individual intelligence to systematically measure the intelligence of groups. Even though social psychologists and others have studied for decades how well groups perform specific tasks (*5*, *6*), they have not attempted to measure group intelligence in the same way individual intelligence is measured—by assessing how well a single group can perform a wide range of different tasks and using that information to predict how that same group will perform other tasks in the future. The goal of the research reported here was to test the hypothesis that groups, like individuals, do have characteristic levels of intelligence, which can be measured and used to predict the groups' performance on a wide variety of tasks.

Although controversy has surrounded it, the concept of measurable human intelligence is based on a fact that is still as remarkable as it was to Spearman when he first documented it in 1904

[1]Carnegie Mellon University, Tepper School of Business, Pittsburgh, PA 15213, USA. [2]Union College, Schenectady, NY 12308, USA. [3]Massachusetts Institute of Technology (MIT) Center for Collective Intelligence, Cambridge, MA 02142, USA. [4]MIT Media Lab, Cambridge, MA 02139, USA. [5]MIT Sloan School of Management, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: awoolley@cmu.edu

Downloaded from www.sciencemag.org on October 31, 2010

(*7*): People who do well on one mental task tend to do well on most others, despite large variations in the tests' contents and methods of administration (*4*). In principle, performance on cognitive tasks could be largely uncorrelated, as one might expect if each relied on a specific set of capacities that was not used by other tasks (*8*). It could even be negatively correlated, if practicing to improve one task caused neglect of others (*9*). The empirical fact of general cognitive ability as first demonstrated by Spearman is now, arguably, the most replicated result in all of psychology (*4*).

Evidence of general intelligence comes from the observation that the average correlation among individuals' performance scores on a relatively diverse set of cognitive tasks is positive, the first factor extracted in a factor analysis of these scores generally accounts for 30 to 50% of the variance, and subsequent factors extracted account for substantially less variance. This first factor extracted in an analysis of individual intelligence tests is referred to as general cognitive ability, or *g*, and it is the main factor that intelligence tests measure. What makes intelligence tests of substantial practical (not just theoretical) importance is that intelligence can be measured in an hour or less, and is a reliable predictor of a very wide range of important life outcomes over a long span of time, including grades in school, success in many occupations, and even life expectancy (*4*).

By analogy with individual intelligence, we define a group's collective intelligence (*c*) as the general ability of the group to perform a wide variety of tasks. Empirically, collective intelligence is the inference one draws when the ability of a group to perform one task is correlated with that group's ability to perform a wide range of other tasks. This kind of collective intelligence is a property of the group itself, not just the individuals in it. Unlike previous work that examined the effect on group performance of the average intelligence of individual group members (*10*), one of our goals is to determine whether the collective intelligence of the group as a whole has predictive power above and beyond what can be explained by knowing the abilities of the individual group members.

The first question we examined was whether collective intelligence—in this sense—even exists. Is there a single factor for groups, a *c* factor, that functions in the same way for groups as general intelligence does for individuals? Or does group performance, instead, have some other correlational structure, such as several equally important but independent factors, as is typically found in research on individual personality (*11*)?

To answer this question, we randomly assigned individuals to groups and asked them to perform a variety of different tasks (*12*). In Study 1, 40 three-person groups worked together for up to 5 hours on a diverse set of simple group tasks plus a more complex criterion task. To guide our task sampling, we drew tasks from all quadrants of the McGrath Task Circumplex (*6*, *12*), a well-established taxonomy of group tasks based on the coordination processes they require. Tasks included solving visual puzzles, brainstorming, making collective moral judgments, and negotiating over limited resources. At the beginning of each session, we measured team members' individual intelligence. And, as a criterion task at the end of each session, each group played checkers against a standardized computer opponent.

The results support the hypothesis that a general collective intelligence factor (*c*) exists in groups. First, the average inter-item correlation for group scores on different tasks is positive (r = 0.28) (Table 1). Next, factor analysis of team scores yielded one factor with an initial eigenvalue accounting for more than 43% of the variance (in the middle of the 30 to 50% range typical in individual intelligence tests), whereas the next factor accounted for only 18%. Confirmatory factor analysis supported the fit of a single latent factor model with the data [χ^2 = 1.66, P = 0.89, df = 5; comparative fit index (CFI) =.99, root mean square error of approximation (RMSEA) = 0.01]. Furthermore, when the factor loadings for different tasks on the first general factor are used to calculate a *c* score for each group, this score strongly predicts performance on the criterion task (r = 0.52, P = 0.01). Finally, the average and maximum intelligence scores of individual group members are not significantly correlated with *c* [r = 0.19, not significant (ns); r = 0.27, ns, respectively] and not predictive of criterion task performance (r = 0.18, ns; r = 0.13, ns, respectively). In a regression using both individual intelligence and *c* to predict performance on the criterion task, *c* has a significant effect (β = 0.51, P = 0.001), but average individual intelligence (β = 0.08, ns) and maximum individual intelligence (β =.01, ns) do not (Fig. 1).

In Study 2, we used 152 groups ranging from two to five members. Our goal was to replicate these findings in groups of different sizes, using a broader sample of tasks and an alternative measure of individual intelligence. As expected, this study replicated the findings of Study 1, yielding a first factor explaining 44% of the variance and a second factor explaining only 20%. In addition, a confirmatory factor analysis suggests an excellent fit of the single-factor model with the data (χ^2 = 5.85, P = 0.32, df = 5; CFI = 0.98, NFI = 0.89, RMSEA = 0.03).

In addition, for a subset of the groups in Study 2, we included five additional tasks, for a total of ten. The results from analyses incorporating all ten tasks were also consistent with the hypothesis that a general *c* factor exists (see Fig. 2). The scree test (*13*) clearly suggests that a one-factor model is the best fit for the data from both studies [Akaike Information Criterion (AIC) = 0.00 for single-factor solution]. Furthermore, parallel analysis (*13*) suggests that only factors with an eigenvalue above 1.38 should be retained, and there is only one such factor in each sample. These conclusions are supported by examining the eigenvalues both before and after principal axis extraction, which yields a first factor explaining 31% of

Table 1. Correlations among group tasks and descriptive statistics for Study 1. n = 40 groups; *$P \leq$ 0.05; **$P \leq$ 0.001.

	1	2	3	4	5	6	7	8	9
1 Collective intelligence (*c*)									
2 Brainstorming	0.38*								
3 Group matrix reasoning	0.86**	0.30*							
4 Group moral reasoning	0.42*	0.12	0.27						
5 Plan shopping trip	0.66**	0.21	0.38*	0.18					
6 Group typing	0.80**	0.13	0.50**	0.25*	0.43*				
7 Avg member intelligence	0.19	0.11	0.19	0.12	−0.06	0.22			
8 Max member intelligence	0.27	0.09	0.33*	0.05	−0.04	0.28	0.73**		
9 Video game	0.52*	0.17	0.38*	0.37*	0.39*	0.44*	0.18	0.13	
Minimum	−2.67	9	2	32	−10.80	148	4.00	8.00	26
Maximum	1.56	55	17	81	82.40	1169	12.67	15.67	96
Mean	0	28.33	11.05	57.35	46.92	596.13	8.92	11.67	61.80
SD	1.00	11.36	3.02	10.96	19.64	263.74	1.82	1.69	17.56

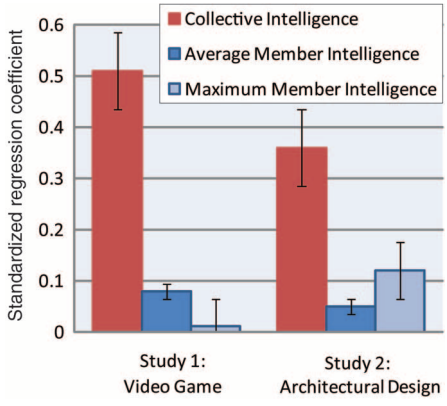
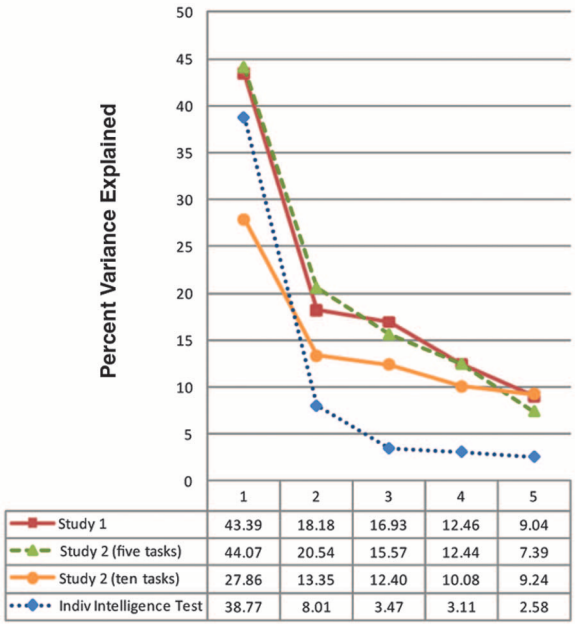


Fig. 1. Standardized regression coefficients for collective intelligence (*c*) and average individual member intelligence when both are regressed together on criterion task performance in Studies 1 and 2 (controlling for group size in Study 2). Coefficient for maximum member intelligence is also shown for comparison, calculated in a separate regression because it is too highly correlated with individual member intelligence to incorporate both in a single analysis (r = 0.73 and 0.62 in Studies 1 and 2, respectively). Error bars, mean ± SE.

Downloaded from www.sciencemag.org on October 31, 2010



	1	2	3	4	5
Study 1	43.39	18.18	16.93	12.46	9.04
Study 2 (five tasks)	44.07	20.54	15.57	12.44	7.39
Study 2 (ten tasks)	27.86	13.35	12.40	10.08	9.24
Indiv Intelligence Test	38.77	8.01	3.47	3.11	2.58

Fig. 2. Scree plot demonstrating the first factor from each study accounting for more than twice as much variance as subsequent factors. Factor analysis of items from the Wonderlic Personnel Test of Individual intelligence administered to 642 individuals is included as a comparison.

the variance in Study 1 and 35% of the variance in Study 2. Multiple-group confirmatory factor analysis suggests that the factor structures of the two studies are invariant ($\chi^2 = 11.34$, $P = 0.66$, df = 14; CFI = 0.99, RMSEA = 0.01). Taken together, these results provide strong support for the existence of a single dominant *c* factor underlying group performance.

The criterion task used in Study 2 was an architectural design task modeled after a complex research and development problem (*14*). We had a sample of 63 individuals complete this task working alone, and under these circumstances, individual intelligence was a significant predictor of performance on the task ($r = 0.33$, $P = 0.009$).

When the same task was done by groups, however, the average individual intelligence of the group members was not a significant predictor of group performance ($r = 0.18$, ns). When both individual intelligence and *c* are used to predict group performance, *c* is a significant predictor ($\beta = 0.36$, $P = 0.0001$), but average group member intelligence ($\beta = 0.05$, ns) and maximum member intelligence ($\beta = 0.12$, ns) are not (Fig. 1).

If *c* exists, what causes it? Combining the findings of the two studies, the average intelligence of individual group members was moderately correlated with *c* ($r = 0.15$, $P = 0.04$), and so was the intelligence of the highest-scoring team member ($r = 0.19$, $P = 0.008$). However, for both studies, *c* was still a much better predictor of group performance on the criterion tasks than the average or maximum individual intelligence (Fig. 1).

We also examined a number of group and individual factors that might be good predictors of *c*. We found that many of the factors one might have expected to predict group performance—such as group cohesion, motivation, and satisfaction—did not.

However, three factors were significantly correlated with *c*. First, there was a significant correlation between *c* and the average social sensitivity of group members, as measured by the "Reading the Mind in the Eyes" test (*15*) ($r = 0.26$, $P = 0.002$). Second, *c* was negatively correlated with the variance in the number of speaking turns by group members, as measured by the sociometric badges worn by a subset of the groups (*16*) ($r = -0.41$, $P = 0.01$). In other words, groups where a few people dominated the conversation were less collectively intelligent than those with a more equal distribution of conversational turn-taking.

Finally, *c* was positively and significantly correlated with the proportion of females in the group ($r = 0.23$, $P = 0.007$). However, this result appears to be largely mediated by social sensitivity (Sobel $z = 1.93$, $P = 0.03$), because (consistent with previous research) women in our sample scored better on the social sensitivity measure than men [$t(441) = 3.42$, $P = 0.001$]. In a regression analysis with the groups for which all three variables (social sensitivity, speaking turn variance, and percent female) were available, all had similar predictive power for *c*, although only social sensitivity reached statistical significance ($\beta = 0.33$, $P = 0.05$) (*12*).

These results provide substantial evidence for the existence of *c* in groups, analogous to a well-known similar ability in individuals. Notably, this collective intelligence factor appears to depend both on the composition of the group (e.g., average member intelligence) and on factors that emerge from the way group members interact when they are assembled (e.g., their conversational turn-taking behavior) (*17*, *18*).

These findings raise many additional questions. For example, could a short collective inteligence test predict a sales team's or a top management team's long-term effectiveness? More importantly, it would seem to be much easier to raise the intelligence of a group than an individual. Could a group's collective intelligence be increased by, for example, better electronic collaboration tools?

Many previous studies have addressed questions like these for specific tasks, but by measuring the effects of specific interventions on a group's *c*, one can predict the effects of those interventions on a wide range of tasks. Thus, the ability to measure collective intelligence as a stable property of groups provides both a substantial economy of effort and a range of new questions to explore in building a science of collective performance.

References and Notes

1. S. Wuchty, B. F. Jones, B. Uzzi, *Science* **316**, 1036 (2007).
2. T. Gowers, M. Nielsen, *Nature* **461**, 879 (2009).
3. J. R. Hackman, *Leading Teams: Setting the Stage for Great Performances* (Harvard Business School Press, Boston, 2002).
4. I. J. Deary, *Looking Down on Human Intelligence: From Psychometrics to the Brain* (Oxford Univ. Press, New York, 2000).
5. J. R. Hackman, C. G. Morris, in *Small Groups and Social Interaction*, Volume 1, H. H. Blumberg, A. P. Hare, V. Kent, M. Davies, Eds. (Wiley, Chichester, UK, 1983), pp. 331–345.
6. J. E. McGrath, *Groups: Interaction and Performance* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
7. C. Spearman, *Am. J. Psychol.* **15**, 201 (1904).
8. C. F. Chabris, in *Integrating the Mind: Domain General Versus Domain Specific Processes in Higher Cognition*, M. J. Roberts, Ed. (Psychology Press, Hove, UK, 2007), pp. 449–491.
9. C. Brand, *The g Factor* (Wiley, Chichester, UK, 1996).
10. D. J. Devine, J. L. Philips, *Small Group Res.* **32**, 507 (2001).
11. R. R. McCrae, P. T. Costa Jr., *J. Pers. Soc. Psychol.* **52**, 81 (1987).
12. Materials and methods are available as supporting material on *Science* Online.
13. R. B. Cattell, *Multivariate Behav. Res.* **1**, 245 (1966).
14. A. W. Woolley, *Organ. Sci.* **20**, 500 (2009).
15. S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, *J. Child Psychol. Psychiatry* **42**, 241 (2001).
16. A. Pentland, *Honest Signals: How They Shape Our World* (Bradford Books, Cambridge, MA, 2008).
17. L. K. Michaelsen, W. E. Watson, R. H. Black, *J. Appl. Psychol.* **74**, 834 (1989).
18. R. S. Tindale, J. R. Larson, *J. Appl. Psychol.* **77**, 102 (1992).

19. This work was made possible by financial support from the National Science Foundation (grant IIS-0963451), the Army Research Office (grant 56692-MA), the Berkman Faculty Development Fund at Carnegie Mellon University, and Cisco Systems, Inc., through their sponsorship of the MIT Center for Collective Intelligence. We would especially like to thank S. Kosslyn for his invaluable help in the initial conceptualization and early stages of this work and I. Aggarwal and W. Dong for substantial help with data collection and analysis. We are also grateful for comments and research assistance from L. Argote, E. Anderson, J. Chapman, M. Ding, S. Gaikwad, C. Huang, J. Introne, C. Lee, N. Nath, S. Pandey, N. Peterson, H. Ra, C. Ritter, F. Sun, E. Sievers, K. Tenabe, and R. Wong. The hardware and software used in collecting sociometric data are the subject of an MIT patent application and will be provided for academic research via a not-for-profit arrangement through A.P. In addition to the affiliations listed above, T.W.M. is also a member of the Strategic Advisory Board at InnoCentive, Inc.; a director of Seriosity, Inc.; and chairman of Phios Corporation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1193147/DC1
Materials and Methods
Tables S1 to S4
References

2 June 2010; accepted 10 September 2010
Published online 30 September 2010;
10.1126/science.1193147
Include this information when citing this paper.

Downloaded from www.sciencemag.org on October 31, 2010