ACADEMIC PRESS

Organizational Behavior and Human Decision Processes 88 (2002) 605–620

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES

www.academicpress.com

Groups perform better than the best individuals on Letters-to-Numbers problems☆

Patrick R. Laughlin,* Bryan L. Bonner,[1] and Andrew G. Miner[2]

Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA

Abstract

Eighty-two four-person cooperative groups and 328 independent individuals solved a random coding of the letters A–J to the numbers 0–9. On each trial the group or individual proposed an equation in letters (e.g., A + D = ?), received the answer in letters (e.g., A + D = B), proposed one specific mapping (e.g., A = 3), received the answer (e.g., True, A = 3), and proposed the full mapping of the 10 letters to the 10 numbers. As predicted, the groups had fewer trials to solution, proposed more complex equations, and identified more letters per equation than each of the best, second-best, third-best, and fourth-best individuals. We interpret the group superiority as due to the highly intellective nature of Letters-to-Numbers problems, which entail demonstrable recognition of correct answers, demonstrable rejection of erroneous answers, and multiple insights into effective collective information processing strategies.

Keywords: Group problem solving; Intellective tasks; Complementary tasks; Letters-to-Numbers problems

1. Introduction

Group versus individual problem solving is a fundamental and enduring issue in social and organizational psychology. A large amount of research over seven decades has compared cooperative groups and independent individuals in between-subjects designs. There are typically an equal number of groups and individuals in the two conditions, thus comparing groups with the average individual. This research supports the robust generalization that groups perform better than the average individual on a range of problem-solving tasks (for reviews, see Baron, Kerr, & Miller, 1992; Brown, 2000; Davis, 1969; Davis, 1992; Forsyth, 1999; Hackman & Morris, 1975; Hastie, 1986; Hill, 1982; Hinsz, Tindale, & Vollrath, 1997; Kelley & Thibaut, 1969; Kerr, MacCoun, & Kramer, 1996a, 1996b; Levine & Moreland, 1998;

☆ This material is based on work supported by the National Science Foundation under Award No. 9708547.

* Corresponding author.
E-mail address: plaughli@uiuc.edu (P.R. Laughlin).
[1] Present address: Department of Psychology, Williams College, Williamstown, MA, USA.
[2] Present address: Carlson School of Management, University of Minnesota, Minneapolis, MN, USA.

0749-5978/02/$ - see front matter
PII: S0749-5978(02)00003-1

Lorge, Fox, Davitz, & Brenner, 1958; McGrath, 1984; Stasser & Dietz-Uhler, 2001; Steiner, 1972). Traditional theoretical explanations of this superiority of groups over the average individual have emphasized processes by which groups recognize and reject errors (dating from Shaw, 1932) and processes by which groups recognize and accept correct responses proposed by one or more members (dating from Lorge & Solomon, 1955). More recent explanations have emphasized groups as collective information processors who effectively process more information than the average individual, especially on complex problems (Hinsz et al., 1997).

Laughlin (1980) proposed a group task continuum anchored by intellective and judgmental tasks. Intellective tasks have a demonstrably correct solution (e.g., mathematical problems, object transfer problems such as the Tower of Hanoi, vocabulary, or analogies) whereas judgmental tasks are evaluative, behavioral, or aesthetic judgments for which no correct answer exists (e.g., attitudinal judgments, preferences for risk, choice dilemmas, and jury decisions). Laughlin and Ellis (1986) subsequently proposed four conditions of demonstrability: (1) group consensus on a conceptual system; (2) sufficient information; (3) incorrect members are able to recognize the correct response if it is proposed; (4) correct members have sufficient ability, motivation, and time to demonstrate the correct response to the incorrect members.

Hastie (1986) distinguished quantity estimation, problem solving, and vocabulary and world knowledge tasks in his review of group versus individual performance on tasks with a criterion of accuracy (intellective tasks in the terminology of Laughlin, 1980, and Laughlin & Ellis, 1986). The relative superiority of groups over the average individual is greatest on problem solving, intermediate on vocabulary and world knowledge, and least on quantity estimation. Hastie proposed that the basic underlying dimension of this relationship is solution demonstrability: the greater the demonstrability of the correct solution, the greater the relative superiority of group over individual performance. Hill (1982) also organized her review of group versus individual performance by three task types: (1) problem solving, (2) concept learning and concept attainment, (3) vocabulary, analogies, and other world knowledge tasks. She also concluded that the relative superiority of groups over the average individual depends upon task type, with greater superiority for problems with highly demonstrable correct answers than concept learning, concept attainment, and world knowledge. In summary, the essential factor in the relative superiority of group performance over the average individual is solution demonstrability: the greater the demonstrability of the solution, the greater the superiority of group performance over the average individual.

Laughlin, VanderStoep, and Hollingshead (1991) extended the traditional comparison of groups and the average individual to a comparison of groups and an equivalent number of individuals. Experiment 1 compared 96 four-person groups (384 group members) and 384 individuals in rule induction. On each trial the group or individual proposed a hypothesis and selected evidence on either one, two, three, or four arrays of information. This design allowed comparison of the four-person groups with the best, second-best, third-best, and fourth-best individuals. The groups performed below the level of the best individuals, at the level of the second-best individuals, and better than the third-best and fourth-best individuals. Laughlin et al. (1991) then proposed that the groups would perform at the level of the best individual with more information and time. Accordingly, Experiment 2 compared 18 four-person groups (72 participants) and 72 individuals with five arrays of information and 10 more minutes for solution. As predicted, the groups performed at the level of the best individuals with this large amount of information and additional time. In a subsequent experiment with a large amount of information Laughlin, Bonner, and Altermatt (1998) compared 48 four-person groups (192 participants) and 192 individuals. The group or individual proposed hypotheses and selected evidence on four arrays on each trial. Again, the groups performed at the level of the best individuals. Thus, the two experiments that have compared groups to more than the average individual have found that groups perform at the level of the best of an equivalent number of individuals on information-rich rule induction problems.

Until now we have considered comparisons of groups and independent individuals in between-subjects designs. In a second approach to group versus individual performance participants first take a cognitive task alone, and then retake the same task as a cooperative group. This allows comparison of the performance of groups of N members with both the average member of the group and the previously identified performance of the best, second-best, ..., Nth best member of the group. Groups perform at the level of their best member on highly demonstrable mathematical and Eureka problems, as indicated by a best-fitting truth-wins social combination model (Laughlin & Ellis, 1986; Laughlin, Kerr, Munch, & Haggarty, 1976; Stasson, Kameda, Parks, Zimmerman, & Davis, 1991). Groups perform at the level of their second-best member on less demonstrable vocabulary and analogy problems, as indicated by a best-fitting truth-supported wins social combination model (Laughlin & Adamopoulos, 1980; Laughlin, Kerr, Davis, Halff, & Marciniak, 1975). Groups perform at the level of their second-best member on survival problems, such as Lost on the Moon, where a list of objects (tank of oxygen, rope, etc.) is rank ordered for usefulness for a group of downed astronauts on a walk to safety (e.g., Bottger & Yetton, 1988; Littlepage, Schmidt, Whisler, & Frost, 1995). On these problems the criterion of performance is correspondence to the rankings of experts (e.g., NASA survival specialists), and hence more judgmental than intellective problems with completely demonstrable solutions. In the research on quantity estimation reviewed by Hastie (1986) groups also perform below the level of their best member and slightly better than their average member. Gigone and Hastie (1997, p. 153) summarize their review of subsequent research on member and group quantity estimation: "Only one study (Reagan-Cirincione, 1994) reported group judgments as being more accurate than the mean of the member's judgments."

In summary, research with both between-subjects and within-subjects designs supports the generalization that group performance increases over independent individual or group member performance with the demonstrability of the problem solution or answer. Groups perform at the level of the best independent individual or the best group member on highly demonstrable mathematical, insight, and information-rich rule induction problems. Groups perform at the level of the second-best individual or group member on world knowledge problems such as vocabulary, analogies, and rankings of objects for usefulness. Groups perform at the level of the average individual or group member on weakly demonstrable estimations of quantities.

The current experiment compared 82 four-person groups (328 members) and an equivalent number of 328 independent individuals on highly intellective Letters-to-Numbers problems. Because these problems entail demonstrable recognition of correct answers, demonstrable rejection of erroneous answers, and multiple insights into effective collective information processing strategies, we predicted that the groups would perform better than the best of an equivalent number of individuals.

2. Letters-to-Numbers problems

The 10 letters, A, B, C, D, E, F, G, H, I, J, were each randomly assigned without replacement to one of the 10 numbers, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The objective was to identify this mapping of the 10 letters to the 10 numbers in as few trials as possible. A trial consisted of proposing an equation that added or subtracted any of the 10 letters (e.g., A + D = ?), receiving the solution to the equation in letters from the experimenter (e.g., A + D = B), proposing one mapping of a letter to a number (e.g., A = 3), feedback whether the mapping is true or false (e.g., True, A = 3), and proposing the full mapping of the 10 letters to the 10 numbers (e.g., A = 3, B = 7, ..., J = 4). On each trial the experimenter checked the proposed mapping of the 10 letters to the 10 numbers. The full correct mapping solved the problem, whereas less than the full correct mapping required another trial. The experimenter did not indicate how many or which of the 10 letters were correctly mapped to the 10 numbers.[3]

An illustration of four trials follows. Assume the correct coding:

A = 3, B = 5, C = 8, D = 2, E = 1, F = 6, G = 4, H = 7,

I = 0, J = 9

Trial	Equation	Hypothesis	Feedback
1	A + B = C	A = 1	False
2	B + C = EA	A = 8	False
3	F + A + D = EE	E = 1	True
4	H − J = −D	I = 0	True

3. Predictions

3.1. Trials to solution

The correct coding of a given letter to a number may be conclusively demonstrated in at least three ways. First, the experimenter gave feedback on a proposed hypothesis on each trial (e.g., True, A is 3). Second, correct codings may be demonstrated by arithmetic and algebra (e.g., if A is known to be 3 and B is known to be 5, the answer to the proposed equation A + B = C identifies C as 8). Third, correct codings may be demonstrated by logic (e.g., if F + G = EI, E must be 1). Moreover, codings may be demonstrated to be plausible or nonplausible, although not definitively correct or incorrect, by transitivity and the other properties of the ordinary

[3] Although Letters-to-Numbers has some similarities to traditional cryptarithmetic problems (e.g., Bartlett, 1958; Newell & Simon, 1972), there are essential differences. In cryptarithmetic problems the objective is also to decode the unique mapping of letters to numbers. Typically one letter is given, as D = 5 in the following problem:

```
  DONALD
+ GERALD
  ROBERT
```

In cryptarithmetic problems all of the necessary information for problem solution is contained in the problem. The problem solver does not propose further equations in letters and learn the answer in letters, or propose specific codings (e.g., T = 0) and learn whether they are correct or incorrect, as in the current Letters-to-Numbers problems. Although the objective in both cryptarithmetic and Letters-to-Numbers is to determine the correspondence between letters and numbers, Letters-to-Numbers goes beyond decoding crytarithmetic problems that contain all necessary information for problem solution to generating further evidence (equations), hypothesis testing (one proposed coding per trial), and reasoning from the resulting feedback on equations and hypotheses to the full correct coding.

number system (e.g., if A + B = C both A and B must be less than C). Conversely, proposed codings may be demonstrated to be erroneous by inconsistency with experimenter feedback, arithmetic, algebra, logic, or the properties of the ordinary number system.

These considerations show that Letters-to-Numbers problems strongly fulfill the four conditions of demonstrability of Laughlin and Ellis (1986) and are therefore highly intellective. The instructions and previous knowledge of arithmetic, algebra, logic, and properties of the ordinary number system comprise group member agreement on the underlying conceptual system (Condition 1). The group members should be able to generate sufficient information for a correct solution from the answers to their proposed equations and feedback on their proposed hypotheses (Condition 2). The incorrect members should be able to recognize a correct inference or coding when it is proposed (Condition 3). The correct members should have sufficient ability, motivation, and time to demonstrate correct inferences and codings to the incorrect members (Condition 4). Consistent with previous research on highly demonstrable problems, the groups should therefore perform as well as the best of an equivalent number of individuals. Beyond this fulfillment of the conditions of demonstrability, a consideration of possible strategies leads to the prediction that the groups will perform significantly *better than* the best individuals.

3.2. Two-letter substitution strategy

An obvious approach is to solve the problems by a two-letter substitution strategy. After identifying one or more letters by feedback on proposed hypotheses, arithmetic, or logic, the problems may be solved by a series of proposed equations that add two of these letters, or subtract one letter from one letter. An illustration follows, omitting the proposed coding of the 10 letters to the 10 numbers on each trial. Again assume the correct coding:

A = 3, B = 5, C = 8, D = 2, E = 1, F = 6, G = 4, H = 7, I = 0, J = 9

Trial	Equation	Hypothesis	Feedback
1	A + J = ED	E = 1	True
2	E + E = D	D = 2	True
3	D + D = G	G = 4	True
4	E + D = A	A = 3	True
5	A + G = H	H = 7	True
6	D + H = J	J = 9	True
7	A + D = B	B = 5	True
8	A + A = F	F = 6	True
9	G + G = C	C = 8	True

After feedback on Trial 9, nine of the letters have been identified, and the remaining letter I which is coded as 0 follows by exclusion. Hence the problem is solved in nine trials.

This illustrative two-letter substitution strategy is obviously suboptimal. The hypotheses inefficiently test what is already known from the answers to the proposed equations. The illustrative strategy does not make further inferences beyond the simple substitutions. For example, the identification of E as 1 on Trial 1, D as 2 on Trial 2, and A as 3 on Trial 4 implies that J is 9 on Trial 1.

Although such a substitution strategy will generally correctly solve the problem in the allowed maximum of 10 trials, it is obviously suboptimal. Hence, we predicted that the groups would have a lower proportion of two-letter substitution strategies than each of the best, second-best, third-best, and fourth-best individuals.

3.3. Letters per equation

Equations with more than two letters will generally provide more information than equations with only two letters as in a two-letter substitution strategy. For example, consider the first two trials of the previous illustration of a two-letter substitution strategy. After the *E* is identified as 1 on Trial 1 and the *D* is identified as 2 on Trial 2, the proposed equation and answer ED + DD = AG on Trial 3 would identify the A as 3 and the G as 4. Similarly, the more complex equation and answer EEE + EED + EDD + DDD = BFH on Trial 3 would identify the B as 5, F as 6, and H as 7. Because such equations with more than two letters are demonstrably more efficient than simple substitution equations with two letters, a group member who proposes equations with more than two letters should be able to convince the other group members to use them. Consequently, we predicted that the groups would have more letters per equation than each of the best, second-best, third-best, and fourth-best individuals.

3.4. Known answer equations

In a substitution strategy of any degree of complexity the known letters are proposed on the left-hand side of the equation to identify the unknown letters on the right-hand side of the equation. Another strategy uses the anticipated answer on the right-hand side of the proposed equation to identify letters on the left-hand side of the equation. To illustrate, consider the following strategy, again assuming the correct coding:

A = 3, B = 5, C = 8, D = 2, E = 1, F = 6, G = 4, H = 7,

I = 0, J = 9

Trial	Equation	Hypothesis	Feedback
1	A + B + C + D + E + F + G + H + I + J = GB	A = 7	False
2	A + C + D + E + F + H + I + J = AF	D = 3	False
3	C + D + E + H + I + J = DH	C = 0	False
4	C + E + I + J = EC	I = 9	False

The sum of the integers from 1 to *N* is given by the formula $N(N+1)/2$. Hence the answer to the equation that adds all 10 letters, A + B + C + D + E + F + G + H + I + J = ? will be $9(10)/2 = 45$, as the letter coded as 0 will have no effect. The equation on Trial 1 adds all 10 letters, identifying G as 4 and B as 5. Adding the remaining eight letters in the equation on Trial 2 has a known answer of 36, identifying the A as 3 and F as 6. Adding the remaining six letters in the equation on Trial 3 has a known answer of 27, identifying the D as 2 and H as 7. Similarly, adding the remaining four letters in the equation on Trial 4 has a known answer of 18, identifying the E as 1 and the C as 8. At this point the two letters I and J remain to be coded to the numbers 0 and 9. The hypothesis on Trial 4 tests I = 9. The feedback False therefore specifies the I as 0 and J as 9. The 10 identified letters are then mapped to the 10 numbers on Trial 4, solving the problem in four trials.[4]

In contrast to a substitution strategy where the known letters on the left-hand side of the equation are used to identify the unknown letters on the right-hand side of the equation, the known answer on the right-hand side of the proposed equation identifies the unknown letters on the left-hand side of the equation. We call such equations known answer equations. Since such known answer equations represent an insightful strategy that may be demonstrated to be effective if proposed by a group member, we predicted that more groups would propose the equation that adds all 10 letters (thus identifying the letters coded as 4 and 5) than each of the best, second-best, third-best, and fourth-best individuals. We further predicted that the groups would have a higher proportion of known answer equations (the previous illustration has a proportion of $4/4 = 1.00$ known answer equations) than each of the best, second-best, third-best, and fourth-best individuals.

3.5. Letters identified per equation

The obtained protocols may be scored for conclusive identification of the coding of a given letter to a number by experimenter feedback on hypotheses or numerical and logical inference from the answers to the proposed equations. We may then determine the average number of letters identified per trial by dividing the total number of identified letters (none to 10) by the number of trials. Because we predicted that the groups would propose more complex equations, use more known answer equations, demonstrate correct inferences to each other, and reject erroneous inferences, we predicted that the groups would identify more letters per trial than each of the best, second-best, third-best, and fourth-best individuals.

3.6. Weighted letter identification

We may also weigh each identified letter inversely by the trial on which it is identified, i.e., 10 for letters identified on the first trial, 9 for letters identified on the second trial, ..., 1 for letters identified on the 10th (final) trial, sum these weighted letter identifications, and divide by the number of trials. This provides a further

[4] To further realize the possibilities of known answer strategies, again assume the coding:
A = 3, B = 5, C = 8, D = 2, E = 1, F = 6, G = 4, H = 7, I = 0, J = 9

Trial	Equation
1	A + B + C + D + E + F + G + H + I + J = GB
2	GG + GG + GG = EAD
3	EDAG + BBBB = FHCJ

The equation on Trial 1 adds all 10 letters, identifying G as 4 and B as 5. Using this known value of G = 4 the answer of the proposed equation GG + GG + GG = ? on Trial 2 will be $44 + 44 + 44 = 132$. This identifies E as 1, A as 3, and D as 2. Using these known values for 1, 2, 3, 4, and 5 the answer of the proposed equation EDAG + BBBB = ? on Trial 3 will be $1234 + 5555 = 6789$. This identifies F as 6, H as 7, C as 8, and J as 9. The letter coded as 0 (here I) follows by exclusion. Hence the problem is solved in three trials.

A known answer equation of the form $N + NN + NNN + NNNN + NNNNN + NNNNNN + NNNNNNN + NNNNNNNN + NNNNNNNNN = ?$, where N is the integers 1, 2, 4, 5, 7, or 8, will give a unique nine-letter answer with no duplications of letters and thus solve the problem. To illustrate, assume E is known to be 1. E + EE + EEE + EEEE + EEEEE + EEEEEE + EEEEEEE + EEEEEEEE + EEEEEEEEE = ? = EDAGBFHCJ = 123456789.

These considerations of known answer strategies indicate that individuals or groups could be trained in effective strategies in research on individual-to-group transfer, group-to-individual transfer, and group-to-group transfer.

measure of problem-solving efficiency. We predicted that the groups would have a higher weighted letter identification score than each of the best, second-best, third-best, and fourth-best individuals.

3.7. Summary of predictions

In summary, we predicted that the groups would solve the problems in fewer trials to solution, have a lower proportion of two-letter substitution strategies, propose more letters per equation, propose more known answer equations, identify more letters per trial, and have a higher weighted letter identification score than each of the best, second-best, third-best, and fourth-best individuals.

4. Method

The participants were 656 students at the University of Illinois at Urbana-Champaign. Three hundred and twenty-eight participants solved a Letters-to-Numbers problem as a four-person cooperative group, and 328 as individuals. Twenty groups and 80 individuals were paid recruits and 62 groups and 248 individuals received course credit for participation.[5]

The experiment was run as a randomized blocks design with each of 82 blocks consisting of eight participants who were randomly assigned to solve the problems as a four-person group or four independent individuals with a given coding of the 10 letters $A, B, \ldots, J$ to the 10 numbers $0, 1, \ldots, 9$. There were 20 different random codings of the 10 letters to numbers. Each of four successive sets of 20 replications (1–20, 21–40, 41–60, 61–80) solved one of these 20 random codings in order. The 81st replication solved the first coding, and the 82nd replication solved the second coding.

The instructions were as follows: This is an experiment in problem solving. The objective is to figure out a code in as few trials as possible. The numbers 0–9 have been coded as the letters A–J in some random order. You will be trying to find out which letter corresponds to which number. It is important to remember that all we are doing is changing the characters used to represent the numbers. We are not changing the way that the number system works. That is, we are still using the same decimal number system you have been using all of your life. Below is an example of a random code:

$$A = 3, \quad B = 5, \quad C = 8, \quad D = 2, \quad E = 1, \quad F = 6, \quad G = 4, \quad H = 7,$$
$$I = 0, \quad J = 9$$

First you will come up with addition or subtraction equations using the letters A–J that will be solved by the experimenter who will give you the answer in letter form. Then you will make a guess as to what one of the letters represents, and we will tell you whether or not the guess is correct. Then you will propose the full mapping of the 10 letters to the 10 numbers. When you have proposed the full correct mapping of the 10 letters to the 10 numbers you will have solved the problem. The objective is to solve

[5] The experiment was run in four sets of replications (20, 20, 20, and 22 replications, respectively) over a Summer Session (paid participants) and three subsequent semesters (unpaid participants who received course credit). Neither the paid nor unpaid participants knew the nature of the experiment prior to participation. An ANOVA for number of trials to solution indicated a significant effect of the four sets of replications, $F(3, 390) = 6.64$, $p < .001$. Post hoc comparisons indicated significantly fewer trials to solution for the fourth (unpaid) set than each of the first, second, and third sets, which did not differ significantly from each other. As the replications $\times$ persons (groups, best, second-best, third-best, fourth-best individuals) interaction was nonsignificant, $F(12, 360) < 1$, paid and unpaid participants were aggregated and are not further considered.

the problem in as few trials as possible. Here are four example trials using the random code above. Note that underlined letters represent experimenter feedback.

Trial	Equation	Hypothesis	Feedback
1	A + B = C	A = 1	False
2	B + C = EA	A = 8	False
3	F + A + D = EE	E = 1	True
4	H − J = −D	I = 0	True

On the first trial the person chooses the equation A + B = and the experimenter tells the person that the solution to this equation is C. This is because $A = 3$ and $B = 5$, which sums to 8, the letter represented by C. The person then guesses that A represents the number 1 and the experimenter indicates that this is not the case, or False. On the second trial the person asks the solution to the equation B + C = and is told that the answer is EA. This is because B = 5 and C = 8, which sums to 13 or EA. Note that on the third trial the person chooses to add three letters together. You may use as many letters as you desire in your equations. Note that on the fourth trial the person chooses to use a subtraction equation. You may use either addition or subtraction equations as you see fit.

Each group member had a response sheet and a randomly selected member had a second group response sheet. On each trial each group member first wrote a proposed equation (e.g., A + D = ?) on his/her response sheet and the group then discussed and proposed an equation, which the group recorder wrote on the group response sheet. Each member then wrote a proposed single hypothesis (e.g., A = 7) on his/her response sheet, and the group discussed and proposed a single hypothesis, which the recorder wrote on the group response sheet. The experimenter stated whether or not the group hypothesis was correct, and the recorder circled Y (for Yes it is correct) or N (for No it is not correct) on a column on the response sheet. Each group member then filled out a proposed full coding of the 10 letters to the 10 numbers (e.g., $\mathrm{A} = 4, \mathrm{B} = 9, \ldots, \mathrm{J} = 7$) on his/her response sheet, and the group discussed and proposed the full coding, which the recorder filled out on the group response sheet. After each trial the experimenter checked the coding of the 10 letters to numbers by the group and stated whether or not it was completely correct. The full correct group coding solved the problem, whereas an incorrect coding required another trial, up to a maximum of 10 trials. Individuals followed the same procedure of proposed equations, one hypothesis, and full coding of the 10 letters to the 10 numbers. All group members and individuals had additional scratch paper for notes.

After the problem was over the experimenter explained the purpose of the research to the participants, gave them a written debriefing with a reference for further reading, thanked them, and asked them not to discuss the experiment with other potential participants.

5. Results

The best, second-best, third-best, and fourth-best individuals were defined by the number of trials to solution for each set of four individuals for each of the 82 replications. If two or more individuals had the same number of trials to solution they were further distinguished by the total number of correct codings of letters to numbers over trials. Nonsolvers in the allotted 10 trials were considered to require 11 trials. A randomized blocks analysis of variance (ANOVA) on the number of trials to solution indicated a significant main effect of the best, second-best, third-best, and fourth-best individuals, $F(3, 243) = 310.15,\ p < .001$. Tukey post hoc tests indicated that all six pairwise comparisons of the four means were significant at $p < .001$. Thus, the best, second-best, third-best, and fourth-best individuals differed significantly from each other on the basic performance measure of trials to solution, which is necessary for a meaningful comparison of the groups and best, second-best, third-best, and fourth-best individuals.

We first conducted a randomized blocks multivariate analysis of variance for persons (groups, best, second-best, third-best, fourth-best individuals) for the dependent variables: (1) trials to solution, (2) letters per equation, (3) known answer equations, (4) letters identified per trial, and (5) weighted letter identification. The effect of persons was significant, $F(20, 1062) = 30.31$, $p < .001$, Wilks' Lambda = .2237. We then conducted univariate analyses of variance and planned contrasts for each of these dependent variables.

5.1. Trials to solution

Table 1 gives the mean trials to solution and standard deviations for persons (groups, best, second-best, third-best, and fourth-best individuals). A randomized blocks analysis of variance (ANOVA) indicated a significant main effect of groups and individuals, $F(4, 324) = 227.33$, $p < .001$. Planned contrasts with one-tailed tests indicated that the groups had significantly fewer trials to solution than each of the best individuals, $t(1, 162) = 1.84$, $p < .05$, standard effect size (Abelson, 1995, p. 46, subsequently abbreviated SES) = .22; second-best individuals, $t(1, 162) = 8.77$, $p < .001$, SES = .97; third-best individuals, $t(1, 162) = 16.95$, $p < .001$, SES = 1.55; and fourth-best individuals, $t(1, 162) = 25.50$, $p < .001$, SES = 1.69.

5.2. Two-letter substitution strategy

In a two-letter substitution strategy all proposed equations either add two letters (e.g. A + B = ?) or subtract one letter from another (e.g., A − B = ?). Forty-nine percent of the groups, 65% of the best individuals, 72% of the second-best individuals, 77% of the third-best individuals, and 71% of the fourth-best individuals used a two-letter substitution strategy exclusively. There was a lower proportion of two-letter substitution strategies for the groups than each of the best individuals, $z = 2.05$, $p < .025$; one-tailed test; second-best individuals, $z = 3.03$, $p < .001$; third-best individuals, $z = 3.72$, $p < .001$; and fourth-best individuals, $z = 2.87$, $p < .001$.

5.3. Letters per equation

Table 1 gives the mean letters per equation and standard deviations for the groups, best, second-best, third-best, and fourth-best individuals. A randomized blocks ANOVA indicated a significant main effect of groups and individuals, $F(4, 324) = 15.99$, $p < .001$. Planned contrasts indicated that the groups had

Table 1
Means and standard deviations for trials to solution, letters per equation, and proportion of known answer equations

Condition	Trials to solution		Letters per equation		Known answer equations	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Group	5.60	1.33	3.12	2.11	.08	.20
Best	5.90	1.11	2.25	.71	.02	.12
Second-best	7.05	1.29	2.10	.26	.00	.02
Third-best	8.40	1.49	2.10	.32	.00	.02
Fourth-best	9.82	1.34	2.08	.20	.00	.00

significantly more letters per equation than each of the best individuals, $t(1, 162) = 5.53$, $p < .001$, SES = .55; second-best individuals, $t(1, 162) = 6.44$, $p < .001$, SES = .65; third-best individuals, $t(1, 162) = 6.43$, $p < .001$, SES = 1.85; and fourth-best individuals, $t(1, 162) = 6.61$, $p < .001$, SES = .66.

5.4. Known answer equations

Sixteen of the groups, five of the best individuals, one of the second-best individuals, one of the third-best individuals, and none of the fourth-best individuals proposed the known answer equation that added all 10 letters, thus identifying the letters coded as 4 and 5. The groups had a higher proportion of these equations ($16/82 = .195$) than the best individuals ($5/82 = .061$), $z = 2.57$, $p < .01$, and all individuals combined, $z = 6.12$, $p < .001$.

Table 1 gives the mean proportion of known answer equations (number of known answer equations divided by number of proposed equations) and standard deviations for the groups, best, second-best, third-best, and fourth-best individuals. A randomized blocks ANOVA on arcsin transformations of the proportions indicated a significant main effect of groups and individuals, $F(4, 324) = 9.91$, $p < .001$. Planned contrasts indicated that the groups had a higher proportion of known answer equations than each of the best individuals, $t(1, 162) = 3.70$, $p < .001$, SES = .37; second-best individuals, $t(1, 162) = 5.07$, $p < .001$, SES = .57; third-best individuals, $t(1, 162) = 5.08$, $p < .001$, SES = .73; and fourth-best individuals, $t(1, 162) = 5.35$, $p < .001$, SES = .60.

5.5. Letters identified per trial

The protocols were scored for the trial on which a given letter was identified by experimenter feedback that a proposed hypothesis (e.g., A = 3) was True, numerical and logical inference, or exclusion when only one unidentified letter remained. Table 2 gives the mean letters identified per trial and standard deviations for the groups, best, second-best, third-best, and fourth-best individuals. A randomized blocks ANOVA indicated a significant main effect of groups and individuals, $F(4, 324) = 24.84$, $p < .001$. Planned contrasts indicated that the groups had significantly more letters identified per trial than each of the best individuals, $t(1, 162) = 2.71$, $p < .001$, SES = .32; second-best individuals, $t(1, 162) = 4.63$, $p < .001$, SES = .56, third-best individuals, $t(1, 162) = 6.20$, $p < .001$, SES = .95; and fourth-best individuals, $t(1, 162) = 9.33$, $p < .001$, SES = 1.03.

5.6. Weighted letter identification

Each identified letter was weighted inversely by the trial on which it was identified, i.e., 10 for letters identified on the first trial, 9 for letters identified on the second trial, ..., 1 for letters identified on the tenth (final) trial, These weighted letter

Table 2
Means and standard deviations for letters identified per trial and weighted letter identification

Condition	Letters identified per trial		Weighted letter identification	
	Mean	*SD*	Mean	*SD*
Group	1.34	.61	9.71	5.04
Best	1.17	.45	8.25	3.58
Second-best	1.04	.38	6.49	2.72
Third-best	.94	.33	5.20	2.23
Fourth-best	.75	.35	3.62	1.82

identifications were then summed and divided by the number of trials. Table 2 gives the mean weighted letter identification scores and standard deviations for the groups, best, second-best, third-best, and fourth-best individuals. A randomized blocks ANOVA indicated a significant main effect of groups and individuals, $F(4, 324) = 53.79,\ p < .001$. Planned contrasts indicated that the groups had a significantly higher weighted letter identification score than each of the best individuals, $t(1, 162) = 3.15,\ p < .001$, SES = .33; second-best individuals, $t(1, 162) = 6.94$, $p < .001$, SES = .74, third-best individuals, $t(1, 162) = 9.70,\ p < .001$, SES = 1.35, and fourth-best individuals, $t(1, 162) = 13.11,\ p < .001$, SES = 1.25.

5.7. Correlations of response measures

Table 3 gives the correlation matrix for the five dependent variables: (1) trials to solution, (2) letters per equation, (3) known answer equations, (4) letters identified per trial, and (5) weighted letter identification. There were negative correlations between the number of trials to solution and each of the four process measures, indicating fewer trials to solution for increasing letters per equation, increasing use of known answer equations, increasing number of letters identified per trial, and increasing weighted letters identified per trial.

6. Discussion

The four-person groups had significantly fewer trials to solution than each of the best, second-best, third-best, and fourth-best of an equivalent number of individuals. The groups used fewer two-letter substitution strategies, proposed more letters per equation, proposed more known answer equations, identified more letters per trial, and had higher weighed letter identification scores than each of the best, second-best, third-best, and fourth-best individuals. Although a large amount of research over seven decades supports the robust generalization that cooperative groups perform better than the average individual in problem solving, this is the first reported evidence that cooperative problem-solving groups perform better than the best of an equivalent number of independent individuals. We attribute this superiority of the groups over the best of an equivalent number of individuals to the highly intellective nature of Letters-to-Numbers problems, which entail demonstrable recognition of correct answers, demonstrable rejection of erroneous answers, and effective collective information processing by proposing informative equations and reasoning from the answers to the equations.

6.1. Group versus individual problem solving and task demonstrability

Previous research clearly indicates that the superiority of groups over the average individual increases as the task becomes increasingly intellective, from estimation, to vocabulary and world knowledge, to problem solving within a logical or quantitative conceptual system. The Letters-to-Numbers problems of the current research strongly fulfilled the four conditions of demonstrability of Laughlin and Ellis (1986).

Table 3
Correlations of performance measures

	Trials	Letters	Known	Identified
Letters per equation	−.32			
Known answer equations	−.30	.79		
Letters identified per trial	−.52	.48	.43	
Weighted letter identification	−.70	.55	.50	.91

The instructions and previous knowledge of the number system, elementary arithmetic, and elementary logic comprised group member agreement on the underlying conceptual system (Condition 1). The group members were able to propose informative equations and reason appropriately from them to conclusively identify the correct mappings of letters to numbers (Condition 2). The members who had not realized that equations with more than two letters were more informative than simple two-letter substitution equations were able to recognize that such equations were more informative when they were proposed by another group member, and the groups used effective known answer equations when proposed (Condition 3). The members who proposed more effective equations with more than two letters, known answer equations, and correct inferences and codings had sufficient ability, motivation, and time to demonstrate these effective equations, correct inferences, and codings to the incorrect members (Condition 4).

In the two previous comparisons of groups and an equivalent number of individuals (Laughlin et al., 1998, 1991, Experiment 2) four-person groups performed at the level of the best of an equivalent number of individuals on rule induction problems. In experimental research on rule induction one of a set of plausible hypotheses is designated as correct by an experimenter who gives error-free feedback on the status of selected or inspected evidence as an example or nonexample of the correct hypothesis (Klayman, 1995; Klayman & Ha, 1987). As formalized by Postulate 6 of Laughlin (1999), rule induction is both intellective and judgmental: nonplausible hypotheses may be demonstrated to be nonplausible by a failure to fit the available evidence (intellective) but correct hypotheses may not be demonstrated to be uniquely correct relative to other plausible hypotheses that also fit the evidence (judgmental). In contrast, in Letters-to-Numbers correct codings and inferences may be demonstrated to be uniquely correct by the operations of arithmetic, logic and the number system and by experimenter feedback that a proposed hypothesis is correct. Because Letters-to-Numbers is more intellective than rule induction, groups are able to perform better than the best of an equivalent number of individuals on Letters-to-Numbers problems but only at the level of the best of an equivalent number of individuals on rule induction problems.

In summary, both previous research and the current experiment support a fundamental generalization: groups perform increasingly better than individuals as the task becomes increasingly intellective with demonstrable recognition of correct answers, demonstrable rejection of erroneous answers, and demonstrably effective collective information processing strategies.

6.2. Group process and productivity

In his classic book *Group Process and Productivity*, Steiner (1972) proposed an elegant theory that has been widely influential in organizational and social psychology. Steiner's basic formulation is expressed in the equation: Actual Productivity = Potential Productivity minus Process Loss. Process Loss includes either or both of coordination loss and motivation loss. On discretionary problem-solving tasks the group selects from the proposed member responses in making a single group response. Such discretionary problem-solving tasks may be conjunctive, where all group members must be correct for the group to be correct, or disjunctive, where the group will be correct if any group member is correct and the group accepts the correct answer. Hence the maximum Potential Productivity of the group is set by the productivity of the best member of the group. Failure of the group to perform as well as the best member of the group necessarily indicates Process Loss. This aspect of Steiner's theory has been widely influential in organizational and social psychology: group process is seen as entailing coordination and/or motivation loss, and group productivity is seen as bounded by the best group member.

In the current experiment an equal number of participants were randomly assigned to work as either a cooperative four-person group or four independent individuals. The best member of the group would be comparable to the best independent individuals, so that group performance at the level of the best individuals would indicate that Actual Productivity was equal to Potential Productivity, with minimal or no Process Loss. But, the groups performed better than the best independent individuals, which is not possible according to Steiner's theory for discretionary tasks where the group selects from proposed member responses. However, it is completely possible and predictable on a consideration of an earlier but less cited article of Steiner (1966).

In this paper Steiner (1966) also proposed the basic Actual Productivity = Potential Productivity minus Process Loss Equation. However, in addition to models of Potential Productivity for the additive, disjunctive, and conjunctive tasks considered in Steiner (1972), he presented models of potential productivity for compensatory tasks and complementary tasks. On compensatory tasks the group does not interact and all members make independent judgments. The group judgment is the mean or median of the independent member judgments, and Potential Productivity is modeled by elementary statistics.

In contrast to additive, disjunctive, conjunctive, and compensatory tasks, where each member performs all aspects of the task, complementary tasks may be divided into subtasks and assigned to different members, enabling a division of labor: "Complementary models are designed to deal with cases in which the single individual performs only a part of a total task, while other persons, possessing different kinds of resources, perform the remaining parts." (Steiner, 1966, p. 280). Steiner presents models that indicate that the potential productivity of groups may be greater than the potential productivity of their best member on complementary tasks. Because Steiner did not consider complementary tasks and present these insightful models in his 1972 book, they have had less impact on theory and research in organizational and social psychology than his considerations of group coordination loss and process loss on additive, disjunctive, and conjunctive tasks.

The current Letters-to-Numbers problems are clearly a Steiner complementary task. Different members may suggest demonstrably effective known answer equations, suggest demonstrably more effective equations than a simple two-letter substitution strategy, reason effectively from the answers to the equations, recognize correct codings, and reject incorrect codings. Because the problems are a highly intellective complementary task, the groups were able to perform better than the best of an equivalent number of individuals.

6.3. Groups and the best individuals in organizations

We have proposed that the groups performed better than the best individuals because of the highly intellective nature of Numbers-to-Letters, a complementary task that entails demonstrable recognition of correct answers, demonstrable rejection of erroneous answers, and multiple insights into effective collective information processing strategies. How could these results apply to organizations? A contemporary illustration comes from planetary astronomy. The past five years have seen a dramatic surge in the number of extrasolar planets which have been detected in orbit around other sunlike stars. To date, nearly 100 extrasolar planets have been discovered, whereas in 1995 only a single example was known. For a description of the recent discovery of two planets orbiting 47 Ursa Majoris see Fischer, Marcy, Butler, Laughlin, and Vogt (2002); and for a review see Marcy, Cochran, and Mayer (2000). One important factor in this increase has been the emergence of highly cooperative efforts among teams of four or five researchers. The impression of researchers in the field is that the pooled application of effort among teams of this size is leading to a more rapid advancement than any one of these team members could achieve alone (Personal Communication, Gregory Laughlin, Department of Astrophysics and Astronomy, University of California at Santa Cruz, August 14, 2001).

More generally, many scientific, business, educational, and administrative organizations depend upon groups and teams that incorporate distributed expertise in the search for problem solutions (Hollenbeck et al., 1995). Many of these problems are complementary intellective tasks which entail demonstrable recognition of correct answers, demonstrable rejection of erroneous answers, and demonstrably effective strategies. It is probably rare to have as many independent individuals as team members working on the same problem. The current results suggest that these problem-solving teams that incorporate distributed expertise on complementary tasks are performing better than the best of an equivalent number of independent individuals in the organization could perform alone.

Acknowledgments

We thank Celia M. Gonzalez and Denise Sommer for assistance in conducting this research.

References

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.

Baron, R. S., Kerr, N. L., & Miller, N. (1992). *Group process, group decision, group action*. Buckingham, England; Bristol, PA: Open University Press.

Bartlett, F. C. (1958). *Thinking*. New York: Basic Books.

Bottger, P. C., & Yetton, P. W. (1988). An integration of process and decision scheme explanations of group problem solving performance. *Organizational Behavior and Human Decision Processes., 69*, 214–221.

Brown, R. (2000). *Group processes* (2nd ed.). Oxford, UK: Blackwell.

Davis, J. H. (1969). *Group performance*. Reading, MA: Addison-Wesley.

Davis, J. H. (1992). Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950–1990. *Organizational Behavior and Human Decision Processes, 52*, 3–38.

Fischer, D. A., Marcy, G. W., Butler, R. P., Laughlin, G., & Vogt, S. S. (2002). Two planets orbiting 47 Ursa Majoris. *Astrophysical Journal, 564*, 1028–1034.

Forsyth, D. R. (1999). *Group dynamics* (3rd ed.). New York: Brooks/Cole.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgment. *Psychological Bulletin, 121*, 149–167.

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8) (pp. 45–99). New York: Academic Press.

Hastie, R. (1986). Review essay: Experimental evidence on group accuracy. In G. Owen, & B. Grofman (Eds.), *Information pooling and group accuracy* (pp. 129–157). Westport, CT: JAI.

Hill, G. W. (1982). Group versus individual performance: Are $N + 1$ heads better than one? *Psychological Bulletin, 91*, 517–539.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*, 43–64.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision-making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*, 292–316.

Kelley, H. H., & Thibaut, J. W. (1969). Group problem solving. In G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 4) (pp. 1–101). Reading, MA: Addison-Wesley.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996a). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103*, 687–719.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996b). "When are N heads better (or worse) than one?": Biased judgment in individuals versus groups. Vol. 1: Consensual action by small groups. In E. Witte, & J. H. Davis (Eds.), *Understanding group behavior* (pp. 105–136). Mahwah, NJ: Lawrence Erlbaum Associates.

Klayman, J. (1995). Varieties of confirmation bias. In J. R. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision making from the perspective of cognitive psychology* (pp. 385–418). New York: Academic Press.

Klayman, J., & Ha, Y.-M. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Laughlin, P. R. (1980). Social combination processes of cooperative problem-solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 127–155). Hillsdale, NJ: Erlbaum.

Laughlin, P. R. (1999). Collective induction: Twelve postulates. *Organizational Behavior and Human Decision Processes, 80*, 50–69.

Laughlin, P. R., & Adamopoulos, J. A. (1980). Social combination processes and individual learning for six-person cooperative groups on an intellective task. *Journal of Personality and Social Psychology, 38*, 941–947.

Laughlin, P. R., Bonner, B. L., & Altermatt, T. W. (1998). Collective versus individual induction with single versus multiple hypotheses. *Journal of Personality and Social Psychology, 75*, 1481–1489.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22*, 177–189.

Laughlin, P. R., Kerr, N. L., Davis, J. H., Halff, H. M., & Marciniak, K. A. (1975). Group size, member ability, and social decision schemes on an intellective task. *Journal of Personality and Social Psychology, 33*, 80–88.

Laughlin, P. R., Kerr, N. L., Munch, M. M., & Haggarty, C. A. (1976). Social decision schemes of the same four-person groups on two different intellective tasks. *Journal of Personality and Social Psychology, 33*, 80–88.

Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology, 61*, 50–67.

Levine, J., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2) (pp. 415–467). New York: McGraw-Hill.

Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input–process–output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology, 69*, 877–889.

Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance: 1920–1957. *Psychological Bulletin, 55*, 337–372.

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of Eureka-type problems. *Psychometrika, 29*, 139–148.

Marcy, G. W., Cochran, W. D., & Mayer, N. (2000). Extrasolar planets around main sequence stars. In V. Mannings, A. P. Ross, & S. S. Russell (Eds.), *Protostars and Planets IV* (pp. 1285–1301). Tucson: University of Arizona Press.

McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice-Hall.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Reagan-Cirincione, P. (1994). Improving the accuracy of group judgment: A process intervention combining group facilitation, social judgment analysis, and information technology. *Organizational Behavior and Human Decision Processes, 58*, 246–270.

Shaw, M. E. (1932). Comparison of individuals and small groups in the rational solution of complex problems. *American Journal of Psychology, 44*, 491–504.

Stasser, G., & Dietz-Uhler, B. (2001). Collective choice, judgment and problem solving. Vol. 3: Group processes. In M. A. Hogg, & R. S. Tindale (Eds.), *Blackwell handbook of social psychology* (pp. 31–55). Oxford, UK: Blackwell.

Stasson, M. F., Kameda, T., Parks, C. D., Zimmerman, S. K., & Davis, J. H. (1991). Effects of assigned group consensus requirement on group problem solving and group members' learning. *Social Psychology Quarterly, 54*, 25–35.

Steiner, I. D. (1966). Models for inferring relationships between group size and potential group productivity. *Behavioral Science, 11*, 273–283.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.

Received 14 June 2001