Journal of Personality and Social Psychology
1987, Vol. 53, No. 3, 497-509

0022-3514/87/$00.75

Productivity Loss In Brainstorming Groups: Toward the Solution of a Riddle

Michael Diehl and Wolfgang Stroebe
Universität Tübingen, Tübingen, Federal Republic of Germany

We conducted four experiments to investigate free riding, evaluation apprehension, and production blocking as explanations of the difference in brainstorming productivity typically observed between real and nominal groups. In Experiment 1, we manipulated assessment expectations in group and individual brainstorming. Although productivity was higher when subjects worked under personal rather than collective assessment instructions, type of session still had a major impact on brainstorming productivity under conditions that eliminated the temptation to free ride. Experiment 2 demonstrated that inducing evaluation apprehension reduced productivity in individual brainstorming. However, the failure to find an interaction between evaluation apprehension and type of session in Experiment 3 raises doubts about evaluation apprehension as a major explanation of the productivity loss in brainstorming groups. Finally, by manipulating blocking directly, we determined in Experiment 4 that production blocking accounted for most of the productivity loss of real brainstorming groups. The processes underlying production blocking are discussed, and a motivational interpretation of blocking is offered.

In his influential book, Osborn (1957) suggested brainstorming as a method of group problem solving that considerably increases the quality and quantity of ideas produced by group members. Brainstorming groups are traditionally given instructions designed to free the individual members from the inhibiting effects of self-criticism and the criticism by others during the problem-solving session. The rules behind brainstorming are as follows: keep in mind that the more ideas the better and the wilder the ideas the better; improve or combine ideas already suggested; and do not be critical. Osborn (1957) claimed that if these rules are followed "the average person can think up twice as many ideas when working with a group than when working alone" (p. 229).

Taylor, Berry, and Block (1958) were the first to test Osborn's claim in a study in which subjects were asked to brainstorm for a period of 12 min either individually or in 4-person groups. To allow for a statistical comparison between results from individual and group sessions, nominal groups were formed from subjects who had brainstormed individually. For each nominal group the ideas of 4 subjects were combined, eliminating redundant ideas by counting only once any idea that had been suggested several times. Thus, the scores of nominal groups represent the level of productivity one would expect if group interaction neither facilitated nor inhibited group productivity. Contrary to Osborn's claim, Taylor et al. found that nominal groups produced nearly twice as many different ideas as the real groups. This finding has since been frequently replicated. Of the 22 experiments listed in Table 1, 18 reported the performance of nominal groups to be superior to that of real groups, and only 4, all involving 2-person groups (Cohen, Whitmyre, & Funk, 1960; Pape & Bölle, 1984; Torrance, 1970, Experiments 1 and 2), reported no difference.

Results have been more equivocal with regard to quality of ideas. Of the few studies that assessed quality, most have reported a measure of *total quality* (i.e., the sum of the quality ratings of the ideas produced by a given subject or group). Because the total quality is highly related to the number of ideas, some authors have preferred to use *average quality.* However, as brainstorming is assumed to increase the production of good ideas, the number of good ideas appears to be a more appropriate measure of quality. Consequently, in these studies that received a score above a chosen cutoff point on a scale of quality ratings was classified as "good." Finally, some studies have assessed the number of unique or original ideas, having used the frequency with which the idea is suggested as a criterion.

The findings for quality appear to be heavily dependent on the type of measure used: In all six studies that assessed total quality, nominal groups performed better than real groups did. No consistent pattern emerged for the other measures. Among those studies, findings were not only inconsistent between studies but even within the same study, if several topics, subject groups, or experimental conditions had been used.

Theories of Productivity Loss in Brainstorming Groups

In view of the accumulation of evidence for the superior productivity of nominal groups, at least in terms of the quantity of ideas produced, it is surprising that the reasons for their superiority have so far not been explained. The three major interpretations that have been offered to account for the lower productivity of real groups are *production blocking, evaluation apprehension,* and *free riding.* In the first part of this article, we discuss these interpretations in light of existing evidence. In the

The authors are indebted to Thomas Ostrom, Kenneth Gergen, and Margaret Stroebe for helpful comments on an earlier draft of this paper. We are also grateful to Rüdiger Arnscheid, Claudia Brandl, and Christl Fischer for their help in collecting some of the data.

Correspondence concerning this article should be addressed to Michael Diehl, Psychologisches Institut, Universität Tübingen, D-74 Tübingen, Federal Republic of Germany.

Table 1
Results of Studies That Compare Brainstorming Productivity of Real Groups (R) With Productivity of Nominal Groups (N)

Study	Group size	Productivity: Quantity	Productivity: Quality
Taylor, Berry, & Block (1958)	4	$R < N$	TQ:$R < N$ AQ:equivocal[a] NO:$R < N$
Cohen, Whitmyre, & Funk (1960)	2	$R = N$	NO:equivocal
Dunette, Campbell, & Jaastad (1963)	4	$R < N$	TQ:$R < N$ AQ:equivocal
Milton (1965)	4	$R < N$	TQ:$R < N$
Gurman (1968)	3	$R < N$	TQ:$R < N$
Bouchard (1969) Experiment 2	4	$R < N$	TQ:$R < N$ AQ:equivocal NG:$R < N$
Rotter & Portugal (1969)	4	$R < N$	—
Vroom, Grant, & Cotton (1969)	4	$R < N$	TQ:$R < N$ AQ:equivocal NG:$R < N$
Bouchard & Hare (1970)	5, 7, 9	$R < N$	—
Torrance (1970)			
Experiment 1	2	$R = N$	NO:$R < N$
Experiment 2	2	$R = N$	NO:$R < N$
Dillon, Graham, & Aidells (1972)	4	$R < N$	—
Bouchard (1972)			
Experiment 2	4	$R < N$	NG:equivocal
Bouchard, Drauden, & Barsaloux (1974)			
Conditions E and F	4	$R < N$	—
Street (1974)	3	$R < N$	—
Harari & Graham (1975)	4	$R < N$	—
Chatterjea & Mitra (1976)	3	$R < N$	—
Madsen & Finger (1978)	4	$R < N$	—
Maginn & Harris (1980)	4	$R < N$	—
Jablin (1981)	4	$R < N$	—
Barkowski, Lamm, & Schwinger (1982)	2	$R < N$	—
Pape & Bölle (1984)	2	$R = N$	NO:$R = N$

Note. The following quality measures were used: total quality (TQ), average quality (AQ), number of original or unique ideas (NO), and number of good ideas (NG). Dashes indicate that quality was not assessed.
[a] Findings vary across different topics, subjects groups, or experimental conditions.

second part, we present four experiments conducted to evaluate these interpretations.

Production Blocking

In their classic review, Lamm and Trommsdorff (1973) argued that the most important cause of the inferiority of real groups is the rule that only one group member speaks at a time. It is unclear, however, how production blocking operates. Although groups are typically given the same time limit as individuals, researchers have emphasized that there is ample time available and that even group members run out of ideas long before the end of a session. Thus, a lack of speaking time can hardly be the reason for the lower productivity of group members. It seems possible, however, that group members who are prohibited from verbalizing their ideas as they occur, may forget or suppress them because they seem less relevant or less original at a later time. Finally, being forced to listen to the ideas of other group members may prove distractive and interfere with the subjects' own thinking.

Production blocking has never been tested directly and there is only limited indirect evidence. Thus, Bouchard and Hare (1970), who compared the productivity of 5-, 7-, and 9-person brainstorming groups with that of nominal groups, found that the productivity of the two types of groups diverged with increasing group size. This finding is consistent with a blocking interpretation, as the length of delay or the probability of delay is likely to increase with group size. Less relevant is a study by Barkowski, Lamm, and Schwinger (1982), in which the authors claimed to support a blocking interpretation by demonstrating that the advantage of the nominal groups disappeared when, instead of number of ideas, a ratio of numbers of ideas to numbers of words spoken was used as a measure of productivity. Because the number of ideas is likely to be correlated with number of words, it is not clear why this procedure should control for blocking.

Evaluation Apprehension

An interpretation in terms of evaluation apprehension suggests that, despite brainstorming instructions, the fear of negative evaluations from other group members prevents subjects who are working in groups from presenting their more original ideas. Support for this hypothesis comes from a study by Collaros and Anderson (1969) who manipulated perceived expertise of group members in brainstorming groups. The authors reasoned that social inhibition would be greater the more group members perceived other members as experts. In their all-experts condition, each member of a brainstorming group was told that all other members had previously worked in such groups, whereas in a one-expert condition, members were told that only one unidentified member had had such an experience. In a third, a no-expert condition, no such instructions were given. Consistent with predictions, productivity was highest in the no-expert condition and lowest in the all-experts condition. Furthermore, subjects in the expert conditions indicated greater feelings of inhibition and reluctance to offer ideas in a postexperimental questionnaire than did subjects who brainstormed without receiving instructions as to the expertness of other group members.

In contrast, Maginn and Harris (1980), who manipulated evaluation apprehension in subjects working individually, could not demonstrate an effect of social inhibition on productivity. Maginn and Harris told subjects in half of their individual brainstorming conditions that there were three judges on the other side of a one-way mirror "who will be listening to your ideas and rating them for quality and originality" (p. 221). The authors reasoned that if evaluation apprehension accounted for the low productivity of real groups, introducing observers to the individual conditions should lower productivity in these conditions compared with that of the real groups. Contrary to the authors' expectations, individual productivity in the presence of observers was not significantly different from that of individual subjects working without observers. To account for their findings, Maginn and Harris (1980) suggested that their judges

failed to induce evaluation apprehension because they were not in the same room as the subjects. This explanation does not seem very plausible, however, in light of findings from research on the impact of evaluation apprehension on social facilitation. These findings demonstrate "that threat of evaluation in the absence of an audience can produce energizing effects upon performance almost identical to those obtained when experts are observing the individual" (Henchy & Glass, 1968, p. 452).

Free Riding

According to this interpretation, subjects are likely to work hard mainly because the experimenter has instructed them to produce as many ideas as possible. Thus, any factor that reduces the ability of the experimenter to monitor individual productivity is also likely to reduce subjects' motivation to work. Because group members expect their ideas to be pooled and analyzed at the group level only, they may feel tempted to free ride on the efforts of others. Subjects who participate in individual sessions, on the other hand, expect their productivity to be monitored individually and thus see no possibility to evade the control exerted by the experimenter. With physical tasks, the social loafing studies of Latané and co-workers (e.g., Latané, Williams & Harkins, 1979; Williams, Harkins & Latané, 1981) have produced results that are in line with this hypothesis.

A second reason for free riding in brainstorming groups can be derived from the economic theory of public goods (Olson, 1965; Stroebe & Frey, 1982). According to this theory, the temptation to free ride varies with group size not only because increases in size lower the *identifiability* of individual contributions, but also because they decrease the *perceived effectiveness* of individual contributions. Perceived effectiveness refers to members' perception of the difference it would make to the group or to themselves if they decided to contribute. In large groups, not all individual contributions are typically required for the group product; consequently, members may feel that their particular contribution is *dispensable.*

As Kerr and Bruun (1983) emphasized, "the dispensability of members' efforts (and hence, the likelihood of free-rider motivation losses) depends strongly on task features" (p. 80). By using the taxonomy of tasks Steiner (1972) developed, Kerr and Bruun argued that dispensability would be more important with disjunctive tasks (when only the best contribution counts as the group product) than with additive tasks (when the group product is the simple sum or average of individual contribution). Brainstorming is neither a purely additive nor a purely disjunctive task. Whereas the emphasis on quantity contained in the brainstorming rules would make it an additive task, the fact that the obvious purpose of the procedure is to come up with a number of good ideas would also make it a disjunctive task. Thus, the relative importance of dispensability as a source of free-rider motivation will depend to some extent on the subjects' conceptions of the brainstorming task.

The role of dispensability as a mediator of free riding has been examined for physical (e.g., Kerr & Bruun, 1983) as well as cognitive tasks (e.g., Harkins & Petty, 1982). Kerr and Bruun (1983) manipulated dispensability by varying task demands (disjunctive, conjunctive) and subjects' self-perceived ability. Consistent with predictions, low-ability subjects found their contributions dispensable and worked less than high-ability subjects under disjunctive task demands (i.e., when only the best score was counted). The reverse results were observed under conjunctive task demands (i.e., when only the worst score counted).

Similar findings were reported in a series of experiments conducted by Harkins and Petty (1982), who demonstrated with a modified brainstorming task (Experiment 4 of this article) that, even with identifiability held constant, an increase in dispensability lowers brainstorming productivity. Subjects in this study had to brainstorm in groups on the possible uses of an object. Subjects were asked to write their ideas on cards and to drop them into a common hopper, thus eliminating identifiability. Harkins and Petty manipulated dispensability by telling subjects either that all group members worked on the same object or that each worked on a different object. Consistent with predictions, subjects who believed every group member worked on a different object produced more ideas than did subjects who believed that all members had been given the same brainstorming task.

An interpretation in terms of free riding could resolve the apparent discrepancy between the results of Collaros and Anderson (1969) and Maginn and Harris (1980) and also account for the pattern of findings typically reported in the brainstorming literature. As applied to the Collaros and Anderson study, this analysis would suggest that the effect of perceived expertise on productivity was mediated by differences in perceived dispensability rather than in evaluation apprehension. Subjects should expect their own efforts to be more dispensable if they expect some members to be more qualified for the job than if they believe all group members to be equally qualified. The failure of the manipulation Maginn and Harris used to increase brainstorming productivity is consistent with a free-rider interpretation, because their manipulation should have affected neither dispensability nor identifiability, as both are already at a maximum with subjects who brainstorm individually.

The free-rider explanation can be used to interpret other findings in the brainstorming literature that previously could not be explained satisfactorily. For example, in some of his brainstorming groups, Bouchard (1972) required that participants contribute their ideas in a fixed sequence and that they announce a "pass" if they had nothing to say when their turn came. The free-rider interpretation would account for the observed increase in productivity by suggesting that taking turns essentially eliminates the temptation to free ride, by increasing the identifiability of individual contributions. The free-rider hypothesis would also account for the performance decrements Bouchard and Hare (1970) observed with increasing group size. This pattern, which has frequently been observed with physical tasks (e.g., Harkins, Latané, & Williams, 1980; Ingham, Levinger, Graves, & Peckham, 1974; Latané et al., 1979; Williams et al., 1981), could be due to both decreased identifiability and increased dispensability.

There is evidence that free riding does occur with brainstorming tasks (Harkins & Petty, 1982) and that the free-rider interpretation can account for certain inconsistent findings (e.g., Collaros & Anderson, 1969; Maginn & Harris, 1980). Nonetheless, only indirect support has been provided for the hypothesis that free riding is responsible for the difference in the productivity of real and nominal brainstorming groups, because Harkins and Petty (1982) did not compare the productiv-

ity of real and nominal groups within a given study. Therefore, we conducted Experiment 1 to test directly this free-rider interpretation.

Experiment 1

If the difference in productivity between real and nominal groups is due to the fact that members of real groups are tempted to free ride because they know that the ideas of all group members will be pooled, then a direct manipulation of these expectations should eliminate the productivity difference between nominal and real brainstorming groups. To test this hypothesis, we asked subjects to brainstorm individually or in groups under instructions that either stressed personal or collective assessment of ideas. Thus, our major aim in this study was to demonstrate that the productivity difference normally observed between group and individual brainstorming is mediated by the implicit association of type of session and assessment expectations. It was not the purpose of this experiment to determine whether this effect is due to a variation in identifiability or dispensability (or both) as their impact on brainstorming productivity had already been demonstrated by Harkins and Petty (1982).

Method

Subjects

Subjects were 48 male students (aged 15–17) from a Tübingen, Germany, high school who were paid for their participation.

Task

As in all further experiments, brainstorming was conducted according to the usual four brainstorming rules. The issue of how to improve the relationship between the German population and the (foreign) guest workers was chosen, an important topic for these students.

Independent Variables

Type of session (individual vs. group). Subjects brainstormed either individually or in 4-person groups.

Assessment instructions (personal vs. collective). Subjects were instructed that the purpose of the experiment was to compare the productivity of persons working individually with that of individuals working in groups. Subjects assigned to group sessions were then told that their performance would be compared with that of subjects working alone. Under personal-assessment instructions group members were informed that each member's individual performance would be compared with that of a subject working alone. Those working under collective-assessment instructions were told that the group's performance would be compared with that of a nominal group, that is, the combined output of 4 subjects working alone. In individual sessions subjects working under personal-assessment instructions were told that their performance would be compared with that of somebody working in a group, whereas subjects following collective-assessment instructions were informed that their ideas would be pooled with that of three other individuals to compare the productivity of this nominal group with that of real groups.

Procedure

Subjects were enrolled in 4-person groups. On arrival they were given their topic and informed of the brainstorming rules. Subjects were then either seated alone in small rooms (individual condition) or led into a somewhat larger room (group condition). Subjects were assigned to conditions on a predetermined random basis.

For group sessions, subjects were seated around a table and given clip-on microphones. They were instructed that they were in a group condition and should make suggestions on the guest worker topic. They were then given the assessment instructions and informed that they had 15 min for the brainstorming and that their responses would be tape-recorded. The experimenter then left the room to switch on the tape recorder and stayed in the control room until the end of the session. On returning, the experimenter handed out the postexperimental questionnaire.

For individual sessions, subjects were seated individually in small rooms and given a clip-on microphone. They were told that they were in an individual condition and should make suggestions about their topic. These suggestions were tape-recorded. They were then given the assessment instructions and told that they had 15 min for the brain-storming. The experimenter then left the room to switch on the tape recorder, stayed in the control room until the end of the session, and later returned to hand out the postexperimental questionnaire.

Dependent Variables

The major dependent variables were the number of nonredundant ideas and the quality of ideas produced by real or nominal groups. The postexperimental questionnaire assessed subjects' understanding of the experimental instructions and also asked how at ease they felt in the brainstorming situation, how satisfied they were with their performance, whether they had suggested all the ideas that had occurred to them, and whether they had as much time as they wanted.

Scoring

Quantity. Ideas were transcribed from the tape recording by a research assistant, who was instructed to write each separate idea on a separate card. To test scorer reliability, a second research assistant repeated this task on the tapes of two nominal and two real groups. The correlation computed on the number of ideas per individual or group for the two scorers was $r = .94$. These cards were then compiled in sets that reflected either the performance of a real or a nominal group. An assistant who was blind as to whether a set constituted the work of a real or a nominal group was instructed to review these cards and eliminate any idea that had been suggested more than once within a given set. To assess the reliability of this decision, a second assistant repeated the procedure for a subsample of four sets. By relating the number of choices in which both raters agreed to the total number of possible pairs, we found that the raters agreed in 99.64% of the total number of possible pairs.[1]

Quality. Ideas were rated for originality and feasibility on two 5-point scales. The second assistant assessed reliability by performing these ratings on a subset of 190 items. Defining the two ratings as in agreement whenever both fell within one point of each other, the two raters agreed in 94.79% of the originality ratings and 94.27% of the feasibility ratings although the variance of the ratings was fairly high. The two measures of quality were not correlated.

[1] The following formula was used to assess the degree of agreement in decisions to eliminate redundant ideas:

$$100\left(1 - \frac{2d}{n(n-1)}\right)$$

where n = number of ideas and d = number of pairs of ideas for which raters arrive at discrepant decisions.

Table 2
Average Number and Quality of Ideas Suggested By Real and Nominal 4-Person Brainstorming Groups Working Under Personal- Versus Collective-Assessment Instructions

Condition	Measure			
	Number of ideas	Number of good ideas	Average originality	Average feasibility
Real group				
Personal	32.33	3.00	2.52	2.90
Collective	23.66	2.00	2.49	3.07
Nominal group				
Personal	84.33	13.33	2.46	2.60
Collective	64.66	5.66	2.43	2.70

Note. Lower numbers indicate higher originality and feasibility.

Results

Table 2 presents the average number and quality of ideas suggested by nominal and real 4-person groups in the two assessment conditions. The quantity measure reflects the number of nonoverlapping ideas per group. Three measures of quality were used: the average for each of the two quality ratings (*originality, feasibility*), and the number of *good ideas.* A good idea was defined as one that received a rating of 1 on one rating scale and no worse than a rating of 2 on the other.

An analysis of variance (ANOVA) with a two-factor design (Type of Session × Type of Assessment) conducted on quantity scores resulted in two main effects. Nominal groups produced significantly more ideas than real groups, $F(1, 8) = 87.56$, $p < .01$ and subjects working under personal-assessment instructions produced more ideas than subjects working under collective-assessment instructions, $F(1, 8) = 8.13$, $p < .05$. There was no indication of an interaction.

Whereas the same two-factor ANOVA conducted on the measures of average originality and average feasibility did not yield any significant effects, results for number of good ideas tended to parallel those for number of ideas. Subjects produced more good ideas in individual sessions rather than in group sessions, $F(1, 8) = 10.38$, $p < .05$, and there was a tendency to produce more good ideas under personal-assessment instructions than under collective-assessment instructions, $F(1, 8) = 3.98$, $p < .10$. Even though the pattern of means (Table 2) suggests that assessment instructions had their major impact in the individual session rather than in the group session, the Type of Session × Type of Assessment interaction did not reach an acceptable level of significance.

The items used to evaluate the understanding of the experimental manipulation indicated only a few misunderstandings. Thus, 5 subjects (two from the group session and three from the individual session) who had received personal-assessment instructions thought that they were following collective-assessment instructions, whereas 1 subject made the opposite mistake. The only two significant effects were main effects for type of session: Subjects reported that they had suppressed more ideas in group sessions than in individual sessions, $F(1, 44) = 4.64$, $p < .05$. Subjects also felt that there was slightly less time in group sessions than in individual sessions, $F(1, 44) = 6.53$, $p < .05$. However, the average difference between conditions was only half a unit on a 5-point scale. Furthermore, the correlation computed between rated availability of time and individual productivity for the group condition was not significant ($r = -.21$). Finally, an analysis of the distribution of ideas over time indicated that more than 90% of the ideas had been expressed at the end of 10 to 12 min in group sessions and even earlier in individual sessions.

Discussion

The finding that subjects produced more ideas when working under personal-assessment as opposed to collective-assessment instructions is consistent with the assumption that subjects' expectations about the assessment of their contributions could account for some of the difference between real and nominal groups. However, the fact that the type of session affected productivity even though assessment expectations had been manipulated independently and that descriptively, the impact of the type of session was much larger (accounting for 83.46% of the total variance) than that of assessment instructions (accounting for only 7.75%), suggests that assessment expectations are at best responsible for a small proportion of the total productivity loss observed in real groups.[2]

The quality ratings did not add a great deal to this evaluation. Our finding that despite their impact on quantity, the experimental manipulations did not affect average quality suggests that the increase in quantity was associated with an increase in poor as well as in good ideas. However, when the number of good ideas was analyzed, the pattern of effects was similar to those observed with quantity ratings. Because the number of good ideas showed a correlation of $r = .82$ with the number of ideas, the additional information gained by looking at these quality ratings did not seem to justify the costly effort involved in conducting quality ratings. Therefore, we decided not to conduct quality ratings in further studies.

Although the results of this study give some support to the assumption that differences in free riding are partly responsible for the productivity difference typically observed between individual and group brainstorming, the results also suggest that there must be other processes contributing to the productivity loss. That free riding is not a major cause of productivity loss in brainstorming groups may be due to the fact that, unlike the physical tasks used in much of the social loafing research (e.g., Latané et al., 1979; Williams et al., 1981), brainstorming does not require a great deal of effort. According to the economic

[2] The proportion of variance accounted for by each of the factors manipulated in an experiment depends on the relative strength of these manipulations. Therefore, statements about the amount of variance accounted for have little meaning in most studies that use social psychological variables, because there is typically no basis for determining on what points on the dimensional scale the manipulations are set. In the case of brainstorming, the situation is different, however, as standard instructions have been used in most of this research. Because our studies were designed to explore potential causes for the productivity difference between group and individual sessions, it is quite instructive to know that type of session accounted for 70% to 80% of the total variance in brainstorming productivity observed in our experiments, even when other variables assumed to mediate this difference were controlled.

model (e.g., Olson, 1965; Stroebe & Frey, 1982), the temptation to free ride should vary as a function of the cost of contributing; consequently, there should be little temptation to free ride with a task that is practically effortless and that involves no time costs (i.e., subjects had to stay for a given period of time no matter what they did).

With free riding ruled out as a major explanation of productivity loss, we were left to consider evaluation apprehension and blocking as additional processes. Because social inhibition is such a pervasive feature of social life, Maginn and Harris's (1980) finding that brainstorming should somehow be exempted had always seemed puzzling. To clarify this issue, we tried to identify differences between our Experiment 1 and the Maginn and Harris study that might have minimized the impact of evaluation apprehension in their experiment.

One such feature could have been the nature of Maginn and Harris's problems. It seems plausible that individuals will be most likely to censor their responses if they fear that certain answers might reveal socially undesirable or even embarrassing aspects of themselves (e.g., lack of knowledge, ideological biases). It can be argued that none of these reasons for self-censure applied to the brainstorming topics Maginn and Harris (1980) used. Their "thumbs" problem ("What would happen if everyone after a certain date had an extra thumb on each hand?") was so obviously irrelevant that subjects even competed in producing silly ideas. The "energy problem" ("How can we reduce gasoline consumption?"), though more involving, was a problem that was uncontroversial and one that had been discussed extensively in the news media. Our topic, on the other hand, was highly controversial. Many potential solutions (e.g., that the guest workers should try harder to learn German and to adapt to German culture; that the government should offer financial inducements to motivate guest workers to return to their own countries) are considered "right wing" and cannot be mentioned in student circles.

If evaluation apprehension in our brainstorming groups was indeed partly caused by the fear of producing ideas that may reveal an unacceptable ideological position, it seems possible that a person's expectation of his or her performance being evaluated by peers would create more evaluation apprehension in this particular situation than would the anticipated evaluation of some unknown assessor. There are two reasons for this: First, the source of apprehension is not one's lack of knowledge but the risk of violating norms concerning political positions shared by one's peers (as opposed to some anonymous evaluator) and second, whereas one expects further interaction with peers, one might never again meet the evaluator.

Experiment 2

In Experiment 2, we examined these hypotheses by manipulating the nature of the brainstorming topics, the level of evaluation apprehension, and the source inducing this apprehension. Thus, subjects were to brainstorm individually on topics that were either rather uncontroversial or controversial and under low- or high-evaluation apprehension. We induced high-evaluation apprehension by telling subjects either that their ideas would be evaluated by some judges (who were not described further) or by their fellow students. If our hypotheses outlined previously were valid, the evaluation apprehension manipulation should affect brainstorming productivity for controversial topics but not for uncontroversial topics. Furthermore, on the controversial topics, we expected high-evaluation apprehension to lead to lower productivity if induced by a threat of peer rather than anonymous assessor evaluation.

Method

Subjects

Subjects were 36 male psychology students of the University of Tübingen who participated in the study as part of their course requirement.

Independent Variables

Evaluation apprehension. Two different procedures were used to induce high-evaluation apprehension. In one condition (judges), modeled after Maginn and Harris (1980), a room with a one-way mirror was used and subjects were told that there were judges sitting on the other side of the mirror who would be listening to their ideas and rating them for quality and originality. In a second condition (peers), a videocamera was mounted in a corner and subjects were told that their performance would be videotaped and that these tapes would be used for demonstration purposes in social psychology classes.[3] To induce low-evaluation apprehension, there was no one-way mirror or videocamera and no instructions were given regarding experts or peers.

Type of topic (uncontroversial vs. controversial). Two rather uncontroversial ("How can entertainment programs on television be improved?"; "How can life quality be improved in the suburbs?") and two controversial topics ("How can the number of guest workers be reduced?"; "How can economic growth be increased in West Germany?") had been selected for brainstorming on the basis of pretests.[4] A controversial topic was defined as one that subjects were forced to argue, not only against their own private opinion but also against a position widely shared by their fellow students. Subjects brainstormed on either the two controversial or the two uncontroversial topics, with the order counterbalanced within each condition.

Procedure

Subjects who were randomly assigned to one of the six experimental conditions were seated individually in small rooms. They were given

[3] This variation in the immediacy of assessment had to be accepted to make the peer manipulation plausible. This was not considered crucial, however, because there is empirical evidence that delay of evaluation has no significant effect on impact (Henchy & Glass, 1968; Maginn & Harris, 1980). Furthermore, even if delay had somewhat reduced the impact of peer evaluation, this would not have interfered with the purpose of the study: that is, (a) to establish that the findings of Maginn and Harris could be replicated with the type of uncontroversial topics used in their study; (b) to establish that it would also hold for controversial topics; and (c) to test whether an even stronger manipulation of evaluation apprehension (i.e., the use of the film in classroom demonstrations) would reduce productivity.

[4] The opposition to economic growth is part of the "green" ideology that connects economic growth to the destruction of our natural environment. Thus, pretest subjects rejected the idea that economic growth should be increased in West Germany and they believed that their fellow students shared this position. Pretest subjects also rejected the notion that the number of guest workers should be reduced (and attributed the same position to their fellow students). With regard to the uncontroversial topics, subjects were in favor of improving entertainment programs and life quality and attributed the same position to their fellow students.

Table 3
Average Number of Ideas Suggested By Individuals Brainstorming Under High- or Low-Evaluation Apprehension Discussing Controversial or Uncontroversial Topics

Type of topic	Low-evaluation apprehension	High-evaluation apprehension	
		Peers	Judges
Controversial			
Economic growth	45.67	24.00	29.00
Guest workers	32.67	14.50	20.50
Uncontroversial			
Suburbs	62.67	36.50	31.33
Entertainment	45.17	28.33	21.83

taped instructions establishing one of three experimental conditions of evaluation apprehension (low-evaluation apprehension, high-evaluation apprehension/peers, high-evaluation apprehension/judges). They were then given the brainstorming rules and either an uncontroversial or a controversial topic and were told that they should speak their suggestions into the microphone. They were informed that they had 15 min for the brainstorming. At the end of the session the experimenter returned and gave the second topic to the subjects. After subjects worked for another 15 min the experimenter returned to hand out the postexperimental questionnaire.

Dependent Variables

The major dependent variable was the number of nonredundant ideas produced by subjects working under the three experimental conditions. As in Experiment 1, the postexperimental questionnaire contained questions that checked the effectiveness of the experimental manipulations. To investigate the controversialness of topics, subjects were asked the extent to which they perceived a need for an improvement in entertainment programs and life quality or whether they believed that the number of guest workers should be reduced and economic growth be increased. They were also asked to indicate the opinions they would attribute to their fellow students on these issues. Finally, subjects had to rate how at ease they felt in the brainstorming situation and whether they had verbalized all the ideas that had occurred to them.

Scoring

The scoring of the number of ideas was performed according to the same procedure as in Experiment 1. Interrater agreement was $r = .87$ for number of ideas transcribed. Although not strictly necessary when only comparisons between individual conditions are planned, it was nevertheless decided to eliminate redundant ideas (within subjects) to increase the comparability between experiments. When this procedure was repeated on a subset of 299 ideas by a second scorer, the scorers agreed in their decision on 99.91% of the possible pairs.

Results

Table 3 presents the scores for the number of ideas individual subjects suggested under each of the three conditions of evaluation apprehension and for the two controversial and the two uncontroversial topics.

A two-factor ANOVA (Evaluation Apprehension × Type of Topic) was conducted collapsing across the two topics within each of the conditions. This analysis resulted in a marginal effect for type of topic, $F(1, 30) = 3.65$, $p < .10$. Thus, fewer ideas were produced when topics were controversial as opposed to uncontroversial. Two orthogonal contrasts were used to examine evaluation apprehension. Whereas an F value of less than 1 did not indicate a differential impact of the two conditions of high-evaluation apprehension (judges vs. peers), the contrast comparing the combined conditions of high-evaluation apprehension with low-evaluation apprehension was highly significant, $F(1, 30) = 14.25$, $p < .01$. Thus, regardless of the type of manipulation, the induction of high-evaluation apprehension resulted in a significant drop in productivity. There was no indication of the predicted interaction between evaluation apprehension and type of topic ($F < 1$). Because we modeled one of our evaluation apprehension conditions (judges) after the procedure Maginn and Harris (1980) used, our findings also confirmed that the evaluation apprehension main effect could be replicated when this condition only was used to represent high-evaluation apprehension.

The manipulation checks indicated that the controversialness manipulation was successful. Thus, subjects themselves agreed less, $F(1, 30) = 85.72$, $p < .01$, and also believed that their fellow subjects would agree less, $F(1, 30) = 56.30$, $p < .01$, with the controversial topic (i.e., need to reduce number of guest workers or need to increase economic growth) than with the uncontroversial topic (i.e., need to improve television programs; need to improve life quality). Subjects working under high-evaluation apprehension did not indicate less ease with the brainstorming situations than did those working under low-evaluation apprehension ($F < 1$).

Discussion

These findings are clearly inconsistent with those of Maginn and Harris (1980), even for the conditions that closely replicate those used in their study. Thus, in contrast to the findings of Maginn and Harris, our induction of evaluation apprehension reduced the quantity of ideas produced in this experiment, and this effect was independent of both the procedure by which evaluation apprehension had been induced and the controversialness of the topic.

These findings suggest that despite brainstorming instructions, the presence of outside observers motivates subjects to censor their own ideas (even when ratings in terms of originality are not made explicit). Because the identifiability of individual contributions is a necessary condition for inducing evaluation apprehension, the findings of this experiment could be considered inconsistent with those of Experiment 1, in which an increase in identifiability was associated with an increase in productivity. This apparent inconsistency becomes even more salient if one considers the findings of Harkins and Jackson (1985), who demonstrated that for identifiability to reduce the temptation to free ride, subjects must believe that their individual outputs could be evaluated by comparison with the outputs of other group members. When subjects believed that individual outputs were not comparable, and thus could not be evaluated, the identifiability induction did not reduce productivity.

The differential impact of the two manipulations was due to the different standards attributed to the evaluators. Our brainstorming instructions clearly indicated that the experimenter was mainly interested in quantity of ideas. Thus, subjects whose

individual contributions were identifiable produced more than those whose contributions were not identifiable. The induction of evaluation apprehension, on the other hand, raised concerns about the quality of ideas and thus resulted in a decrease in the quantity of ideas produced due to self-censoring processes. The emphasis on quality was made explicit in the judges condition, and the fact that there was no difference in impact between judges and peers supports our contention that the anticipation of an evaluation through peers induces the same concerns.

Whereas our findings suggest the possibility that evaluation apprehension might be responsible for part of the productivity loss observed in brainstorming groups, they do not permit any definite conclusions, as the impact of evaluation apprehension was demonstrated only for individual brainstorming.

Experiment 3

To clarify whether the productivity loss in brainstorming groups is due to evaluation apprehension, an experiment was needed to manipulate both evaluation apprehension and type of session. We hypothesized that if the productivity loss in real brainstorming groups is partially or fully caused by the high level of evaluation apprehension induced by the group setting, then inducing evaluation apprehension should have a greater impact on individual brainstormers (who normally have low evaluation apprehension) than on subjects interacting in real groups (where apprehension is always at a high level).[5] We conducted Experiment 3 to test this hypothesis. However, to determine whether the two variables in combination could account for the productivity difference between nominal and real groups, we varied assessment instructions in addition to the type of session and evaluation apprehension.

Method

Subjects

Subjects were 64 psychology students (32 men, 32 women) from Tübingen University who participated in the study as part of their course requirements.

Task

Brainstorming on the guest worker topic used in Experiment 1 was performed in group sessions as well as in individual sessions.

Independent Variables

Type of session (group vs. individual). Subjects were assigned to brainstorm either individually or in 4-person groups. All groups were homogeneous with regard to sex composition, and an equal number of men and women were assigned to each of the experimental conditions.

Evaluation apprehension. High-evaluation apprehension was created by combining the two procedures used in Experiment 2. In the low-evaluation apprehension conditions, we told subjects that their suggestions would be recorded on audiotape. Evaluation by judges or for use in class demonstrations was not mentioned.

Assessment instructions (personal vs. collective). The same instructions were used as in Experiment 1.

Procedure

The procedure for this experiment was almost identical to that of Experiment 1, with the following exceptions: Although 4 subjects were enrolled for sessions at any given time, sometimes fewer than 4 showed up for what had been planned as a group session. These subjects were then tested individually. The instructions used to induce evaluation apprehension were given after the information on type of session and assessment instructions.

Dependent Variables

The major dependent variable was the number of nonredundant ideas produced by real or nominal groups. In addition, subjects had to complete a postexperimental questionnaire that evaluated their understanding of the experimental instructions and also asked how at ease they felt in the brainstorming situation and whether they had been given as much time as they wanted.

Scoring

The scoring of ideas was performed according to the same procedure as in Experiment 1. The interrater agreement was $r = .87$ for the ideas transcribed and 99.83% for the number of redundant ideas.

Results

Table 4 presents the group scores for each of the eight experimental conditions. A three-factor ANOVA (Type of Session × Assessment Instructions × Evaluation Apprehension) performed on these group scores yielded significant main effects for type of session and evaluation apprehension. Subjects who brainstormed individually produced significantly more ideas than did subjects who participated in group sessions, $F(1, 8) = 74.08$, $p < .01$. Similarly, subjects produced significantly more ideas when working under conditions of low- rather than high-evaluation apprehension, $F(1, 8) = 7.08$, $p < .05$. There was no main effect for assessment instructions but only an Assessment Instructions × Evaluation Apprehension interaction, $F(1, 8) = 5.76$, $p < .05$. Orthogonal contrasts that compared the impact of high- versus low-evaluation apprehension (across group and individual conditions) indicated a significant difference under personal assessment instructions, $F(1, 8) = 12.89$, $p < .01$, but not collective assessment instructions ($F < 1$). Thus, inducing high-evaluation apprehension lowered productivity only when personal-assessment instructions had been given. When collective-assessment instructions were given, the evaluation apprehension manipulation had little effect.

Of the subjects, 93.75% correctly answered a multiple-choice item that checked recall of assessment instructions. The question asking how at ease subjects felt in the brainstorming session resulted in a weak main effect for type of session. Subjects felt somewhat less at ease under individual as opposed to group conditions, $F(1, 8) = 4.97$, $p < .10$. Similar to Experiment 2, evaluation apprehension did not affect these ratings. There was also a significant difference between individual and group brainstormers in their assessment of the time available for the task, $\chi^2(2, N = 64) = 14.35$, $p < .01$. Although the majority of responses for both conditions fell into the category "sufficient time," the distribution of the remaining responses differed between the two types of sessions. A sizable minority of individual

[5] This prediction assumes that there is a functional ceiling on felt evaluation apprehension. However, if no such ceiling exists, then the predicted interaction effect need not occur.

Table 4
Average Number of Ideas Suggested By Real and Nominal 4-Person Brainstorming Groups Working Under High- Versus Low-Evaluation Apprehension and Collective- Versus Personal-Assessment Instructions

Condition	Type of assessment	
	Collective assessment	Personal assessment
Real group		
Low-evaluation apprehension	34.50	52.50
High-evaluation apprehension	36.00	40.00
Nominal group		
Low-evaluation apprehension	82.00	102.00
High-evaluation apprehension	78.00	66.00

subjects indicated that there was "too much time," yet subjects who had performed in groups felt that there was "too little time." There were again indications from the group discussion, however, that group members typically ran out of ideas some time before the end of a session.

Discussion

Although the three experimental manipulations had significant effects on brainstorming productivity, the pattern of findings is only partly consistent with our predictions. Type of session resulted in the usual main effect, with real groups producing less than nominal groups. When low-evaluation apprehension was used, the impact of the assessment instructions on productivity replicated the findings from Experiment 1. Again productivity was higher for subjects working under personal as opposed to collective instructions but, as in Experiment 1, assessment instructions did not eliminate the productivity difference between nominal and real groups.

The findings for evaluation apprehension are somewhat unexpected. The impact of evaluation apprehension seems to have been restricted to subjects who expected personal assessment. With personal-assessment instructions, subjects working under high-evaluation apprehension produced significantly fewer ideas than did those working under low-evaluation apprehension. However, the fact that evaluation apprehension did not affect performance of those following collective-assessment instructions suggests that the threat of evaluation does not seem to raise apprehension when the target of the evaluation is perceived as the group and not the individual. This pattern is similar to that Harkins and Jackson (1985) reported.

Furthermore, our results do not confirm the evaluation apprehension interpretation of the productivity loss in real groups. If evaluation apprehension were at least partially responsible for the low productivity of real groups, the evaluation apprehension manipulation would be expected to have less effect in groups than in individual sessions. Although there is a tendency for evaluation apprehension to be more effective in individual than in group conditions, this interaction did not even approach an acceptable level of significance. However, the failure of this interaction to reach statistical significance could be due to the small size of our sample. Also, as noted earlier (see Footnote 5), the interaction prediction requires that evaluation apprehension already approaches its maximum possible value in brainstorming groups; the results of Experiment 4 tend to challenge this assumption.

Although assessment expectations and evaluation apprehension were manipulated in this experiment, type of session still accounted for more than 70% of the total variance in brainstorming productivity. Because our manipulation of evaluation apprehension used an external source of evaluation (judges and nonparticipating peers), it is possible that the evaluation apprehension could be stronger when aroused by members of one's own group hearing and evaluating arguments than by nonparticipating peers. Although we have no data on this issue, this assumption does not seem plausible because, unlike the group members, the external evaluators are not under instructions to be uncritical. Thus, although assessment expectations and evaluation apprehension have been shown to affect brainstorming productivity and can thus be assumed to contribute to the productivity loss in brainstorming groups, their impact has been minor when compared with that of type of session. This would suggest that there are still other powerful causes of the productivity loss that have not been identified in our experiments.

Experiment 4

We designed Experiment 4 to investigate production blocking as a factor contributing to the low productivity of brainstorming groups. Because production blocking cannot be eliminated in real groups, we examined its role by introducing blocking into individual sessions. If blocking were fully responsible for the low productivity of real brainstorming groups, the introduction of comparable blocking into individual sessions would be expected to lower the productivity of subjects brainstorming individually to that of subjects working in groups.

Design Overview

The experiment contained five conditions: In addition to real group and individual brainstorming conditions included as baseline controls, there were three conditions in which blocking was manipulated for subjects working in separate rooms. This was done by signal displays connected to each subject's microphone through a voice-activated sensor. The display consisted of four lights (one green and three red), each light representing a subject. When a subject talked, the sensor switched on the green light for this subject and the red lights for the other subjects. The subject's light was switched off when he or she stopped talking. Because members of real groups usually speak only when others are silent, subjects in Conditions 1 and 2 were asked to talk only when all lights were off. Whereas in Condition 1 subjects could hear the ideas of the other "group members," this was not possible in Condition 2. The fact that subjects' ideas could not be heard in this latter condition eliminated evaluation apprehension within the group. In Condition 3 subjects were informed of the function of the lights (the same explanation as in Conditions 1 and 2) but were told to disregard them and to talk whenever they had anything to say. Thus, no blocking was expected in this condition. To keep expectations regarding assessment constant across conditions, all subjects were given personal-assessment instructions.

Method

Subjects

Subjects were 60 psychology students (20 men, 40 women) at the University of Tübingen who participated in the experiment as part of their course requirement.

Task

The brainstorming topic was: "How can unemployment be reduced in Germany?"

Procedure

Subjects were enrolled in same-sex, 4-person groups. On arrival at the laboratory, each group was assigned to an experimental condition according to a predetermined random order that assured that the proportion of male to female groups remained the same for all experimental conditions. All subjects were then given tape-recorded instructions explaining the brainstorming rules and telling them that the experiment was designed to compare the productivity of individuals who brainstormed individually or in groups. The instructions that follow varied according to conditions.

Group control. The instructions used in this condition were identical to those of the group sessions with personal assessment in Experiments 1 and 3.

Experimental Condition 1 (blocking, communication). These subjects were also told that they were in a group condition, that their suggestions would be tape-recorded, and that their individual performance would be compared with that of another subject working individually. They were then informed that they would each work in a separate room and that they had to communicate with the other group members using intercoms that were connected to a signal system. The function of this system was explained as follows:

> There will be a display of four lights, three red and one green. Each of the lights represents one group member sitting in one of the rooms. Thus Room 1 is represented by Light 1, Room 2 by Light 2, etcetera. In each room, the member's own light is green; that of the other group members red. As soon as one person begins to talk, an acoustic sensor will switch on the light of this person. If a person stops talking for 1.5 seconds, his (her) light will be switched off automatically. Thus, everybody can see who is speaking at the moment. In addition you will hear through your earphones what is being said. Since in a group discussion only one person can talk at any moment, you should make your own contribution only when no other light is on.

Experimental Condition 2 (blocking, no communication). The first part of the instructions for this condition was identical to that of the Condition 1. Subjects were then told that they could not hear each other's ideas and that their terminals had lights and microphones but no earphones. After the function of the lights had been explained, subjects were told that "since this is a simulation of a group discussion, you are only allowed to speak when no other light is on."

Experimental Condition 3 (no blocking, no communication). These subjects were told that they were in an individual condition, that their suggestions would be tape-recorded, and that their performance was to be compared with that of a person working in a group. Their terminals had lights and a microphone but no earphone. After the function of the lights had been explained, subjects were told that "everybody could talk whenever they wanted and that they need not pay any attention to the lights." However, the lights were left operating throughout the session.

Individual control. These subjects received the same instructions used in Experiments 1 and 3 for individual brainstorming sessions with personal assessment.

After instructions had been specified, all subjects were given a first brainstorming topic (discussed for 8 min only) to familiarize subjects in the experimental conditions with the signal system. These discussions were not recorded. At the end of this period, subjects were given the unemployment topic and told that they would now have 15 min to make suggestions.

Table 5
Average Number of Ideas Suggested By Nominal Versus Real 4-Person Groups Working Under Blocking and Nonblocking Conditions

Condition	Number of ideas
Group control	55.67
Condition 1 (blocking, communication)	37.67
Condition 2 (blocking, no communication)	45.67
Condition 3 (no blocking, no communication)	102.67
Individual control	106.00

Dependent Measures

The dependent variable was the number of nonredundant ideas produced by real or nominal groups. The postexperimental questionnaire checked the subjects' understanding of the personal assessment instruction and also asked how at ease they were and whether they had verbalized all the ideas that had occurred to them. In addition, subjects had to rate the adequacy of the time available to them on a 5-point scale.

Scoring

Scoring of the ideas was performed according to the procedure described in Experiment 1. The interrater agreement was $r = .97$ for ideas transcribed and 99.72% for number of redundant ideas.

Results

Table 5 presents the group scores for each of the five experimental conditions. A one-way ANOVA indicated a significant difference between conditions, $F(4, 10) = 10.99$, $p < .01$. A planned contrast comparing the three conditions with blocking (group control, Conditions 1 and 2) against the two conditions without blocking (Condition 3, individual control) resulted in a highly significant effect, $F(1, 10) = 42.22$, $p < .01$. This analysis further revealed that 96% of the variance due to experimental conditions could be attributed to this comparison. A post hoc comparison of all means by Newman–Keuls confirmed that significant differences only occurred between (but not within) blocking and no-blocking conditions.

None of the postexperimental questions resulted in significant effects. Ratings of the availability of time in the various conditions were all above the midpoint on a scale ranging from *very short* to *very long*. Subjects seemed to feel that there was sufficient time available to discuss the problem; this was true under blocking as well as no-blocking conditions. A contrast that compared time ratings under blocking and no-blocking conditions resulted in an F of less than 1, indicating that there was no significant difference in the rated availability of time. That there was sufficient time for discussion even under blocking conditions was further supported by an analysis of the time

spent on discussion in the group condition. It was found that talk (all comments) filled only 73% of the total time available.

Discussion

The findings of Experiment 4 provide strong support for the blocking interpretation. Working under conditions that allowed them to verbalize their ideas as they occurred, subjects produced approximately twice as many ideas as they did when working under conditions in which subjects had to wait their turn. However, the processes by which blocking reduces production are still somewhat unclear. In our earlier discussion of the blocking hypothesis we suggested that group members who are prevented from verbalizing their ideas as soon as they occur may reevaluate them in light of points made by other subjects or that they might simply forget some of these ideas as a result of the distraction provided by the group discussion. Both of these processes would require exposure to the group discussion. Thus, the finding that preventing subjects from overhearing each other's ideas did not significantly increase brainstorming productivity as long as blocking rules were imposed raises doubts about the validity of this interpretation. Because the knowledge that other group members will not be able to overhear one's ideas should eliminate perceptions of the other group members as a source of social inhibition, this last finding is also inconsistent with the evaluation apprehension interpretation of the productivity loss in brainstorming groups.

To control for a variation in assessment expectations, we tested all experimental conditions under personal-assessment instructions. This eliminated differences in assessment expectations as a confounding factor that could have contributed to the productivity difference observed between individual and group sessions. It should be remembered, however, that because assessment expectations were controlled, the findings of this study cannot rule out the possibility that free riding may contribute to the productivity loss in brainstorming groups.

General Discussion

The findings of these four experiments indicate that the type of session has a powerful effect on brainstorming productivity, accounting for more than two-thirds of the variance in the number of nonredundant ideas produced. Therefore, our evaluation of the three interpretations of the productivity loss not only addresses the question of whether evaluation apprehension, free riding, and blocking have a demonstrable effect on brainstorming productivity but also whether any of these processes is likely to be a major cause of the productivity loss. Obviously, our answers to the second concern have to be approximated for two reasons: First, we never manipulated all three variables within the same design and second, we cannot be sure whether the potency of our experimental manipulations is representative of the strength of these variables in real group settings. However, in view of the facts that the potency issue does not arise for the manipulation of type of session, that the type of session was pitted against the other two factors, and that, in each of these comparisons, the type of session always accounted for an overwhelming share of the variance, we argue that our findings do permit statements about the relative importance of the three processes.

Evaluation Apprehensions

In contrast to Maginn and Harris (1980), who failed to find an impact of evaluation apprehension on brainstorming productivity, the results of our Experiment 2 clearly demonstrated that the knowledge that peers or judges are evaluating one's performance significantly reduced the number of ideas produced in individual brainstorming sessions. However, the fact that our attempts to experimentally induce evaluation apprehension in both individual and group sessions in Experiment 3 did not significantly reduce the difference in productivity between the two types of sessions raises some doubt as to whether evaluation apprehension is a major cause of the productivity loss in brainstorming groups.[6] This doubt was strengthened by the finding of Experiment 4, that whether or not subjects could overhear each other's ideas made little difference to brainstorming productivity. Because brainstorming instructions are designed specifically to free group members from the inhibiting effects of criticism by other group members, it is plausible that members of brainstorming groups do not constitute powerful sources of social inhibition for each other.

Free Riding

Although the free-rider interpretation fared somewhat better than the evaluation apprehension explanation, our findings raise considerable doubt as to whether free riding should be considered a major cause of the productivity loss in brainstorming groups. Thus, despite the explicit manipulation of assessment expectations in Experiments 1 and 3, the type of session was still shown to account for more than two-thirds of the total variance in brainstorming productivity. This finding, which suggests that there is little temptation to free ride in brainstorming groups, is quite consistent with the economic analysis (e.g., Olson, 1965; Stroebe & Frey, 1982). According to this model, the costs of the individual contribution (e.g., effort, money, time) are an important determinant of the motivation to free ride. Other things being equal, the temptation to free ride should be greater, the greater the cost of the contribution expected of the individual. Because suggesting solutions to the types of problems used in brainstorming research does not require great effort or concentration and because subjects are already committed to spending a set period of time in the laboratory, the costs of contributing are very low in brainstorming situations. The results of our experiments are, therefore, not inconsistent with the fact that studies using strenuous physical tasks like rope pulling (e.g., Ingham et al., 1974) and clapping and shouting (e.g., Harkins et al., 1980; Latané et al., 1979; Williams et al., 1981) typically reported sizable free-rider effects.

Blocking

The one interpretation our data supported most strongly was production blocking. The presence or absence of blocking accounted for most of the variance created by the experimental manipulations in Experiment 4. Because the impact of the blocking manipulation remained unaffected by a variation in

[6] But again, see Footnote 5.

group members' exposure to the content of the group discussion, the blocking effect does not seem to be due to a reevaluation of ideas in light of the points made by other group members or to a forgetting of ideas due to the distractive activity of listening to the discussion.

However, although subjects may not be forgetting ideas, the periods of delay due to blocking may prevent them from developing new thoughts. After all, storage space in short term memory is fairly limited, and individuals will only be able to store a small number of ideas at a given time. This would suggest that productivity in brainstorming groups could be improved if group members were allowed to write down their own ideas. Tentative support for this hypothesis comes from Street (1974), who found that 3-person groups increased their productivity if each subject wrote down his or her own ideas instead of dictating them to a group member who acted as a recorder.

A second factor contributing to the blocking effect could be the difference in the length of time available for suggesting ideas. Brainstorming research makes use of what has been called *equal man-hour comparisons* (i.e., participants under individual and group conditions are given the same amount of time), which allows more time for individual brainstorming.[7] However, if one defines productivity as the number of ideas developed per time unit (as the emphasis on quantity contained in the brainstorming instructions suggests), then time should be kept constant in a comparison of individual and group productivity. Furthermore, there seems to be little indication that group members lack sufficient time to express their ideas. Even in group sessions, subjects tend to run out of ideas toward the end of the allotted time period.

How, then, does the time limit influence brainstorming productivity? We would like to suggest an interpretation of blocking in terms of situational demands. This interpretation, which seems to be consistent with most of the evidence reported earlier, conceptualizes blocking as a source of motivation rather than as a coodination loss (Steiner, 1972). If individuals are given a time period in which to suggest ideas on some topic, they are likely to feel obliged to continue this activity for most of this time. Obviously, this task is less difficult for group members than for participants in individual sessions. Whereas group members can relax and let others do the talking, subjects in individual sessions have to fill all of the time by themselves. Even periods of silence might be more embarrassing in individual sessions, where subjects have nobody to blame but themselves, than in group sessions, where the responsibility is shared.

General Implications

In view of the considerable effort that has been invested in developing and testing models that allow a comparison of individual and group productivity, it is surprising that solution time has received relatively little attention as a variable in group problem solving. Most individual and group comparisons relate group size only with the proportion of solvers, although it has been found that the analysis of solution times resulted in conclusions at variance with those that were based on the proportion of solutions. Thus Taylor and Faust (1952), who compared group and individual performance for the game of "Twenty Questions," concluded that although "group performances were superior to individual performance in terms of . . . elapsed time per problem . . . in terms of man-minutes required for solution, the performance of individuals was superior to that of groups" (p. 367).

Bray, Kerr, and Atkin (1978), who assessed performance of groups of different sizes in terms of both the proportion of solvers and the time of solution, found that although groups often worked up to their potential productivity with regard to the proportion index, those same groups fell below potential on the latency index. Bray et al. suggested the concept of *functional size* to explain this pattern. According to this notion, the number of nonparticipators in a group increases with group size, resulting in a functional group size smaller than the actual size. Bray et al. found that an analysis that was based on the assumption that groups function at the rate of the fastest member of the functional group resulted in a more accurate prediction of solution times than the assumption implied in the equalitarian model of Restle and Davis (1962). The latter model assumes that group functioning results in a slow-down of each solver's efficiency that is proportional to the number of nonsolvers in the group.

In addition to the theoretical importance of the study of solution times, this type of analysis also has practical implications (e.g., for the decision to have individuals work in groups rather than alone). As Taylor and Faust (1952) pointed out:

> It appears probable that there are many kinds of problems which a group will solve more quickly than an individual. If elapsed time in hours, weeks, or months is the primary consideration, then such problems should be undertaken by groups. However, it appears equally probable that few of those same problems will be solved more efficiently in terms of man-minutes or man-hours by groups than by individuals. (pp. 522–523).

Our findings suggest even more specific implications for the use of group discussion procedures. Because blocking slows down the generation of ideas in groups, it might be more effective to ask subjects first to develop their ideas in individual sessions and next have these ideas discussed and evaluated in a group session. The task of the group would then consist of evaluation rather than production of ideas. This procedure might

---

[7] As an alternative solution, it has been suggested that group members be given the same time as individuals who work on a problem by themselves (i.e., to give n-person groups $n \times$ the time period allotted to individuals). Thus with persons working for 15 min in our experiments, 4-person groups should be given 1 hr. However, even if members of groups were slightly more productive than persons working alone, they would have to be more than four times as productive to compensate for the greater need for time. Furthermore, because subjects might be able to think about the issue when others are talking, providing four times as much time to members of 4-person groups may give them an unfair advantage. Alternatively, one might think of making time discretionary (i.e., persons brainstorming in groups or alone would be asked to continue until they had no more ideas). This would give subjects a chance to exhaust their repertory of responses regardless of context. However, this solution has the drawback of introducing time costs into the problem-solving situation. Whereas subjects working under set times had nothing to lose if they attended to the problem, subjects working under discretionary time limits would have no opportunity costs in terms of the activities they could have performed if they had left the laboratory earlier. Among other things, this should increase the temptation to free ride.

combine the advantage of group and individual sessions without making unnecessary demands on individual time.

References

Barkowski, D., Lamm, H., & Schwinger, T. (1982). Einfallsproduktion von Individuen und Dyaden unter "Brainstorming"-Bedingungen [Brainstorming productivity of individuals and dyads]. *Psychologische Beiträge, 24,* 39–46.

Bouchard, T. J. (1969). Personality, problem-solving procedure and performance in small groups. *Journal of Applied Psychology Monograph, 53* (1, Pt. 2).

Bouchard, T. J. (1972). A comparison of two group brainstorming procedures. *Journal of Applied Psychology, 56,* 418–421.

Bouchard, T. J., Drauden, G., & Barsaloux, J. (1974). A comparison of individual subgroup and total group methods of problem solving. *Journal of Applied Psychology, 59,* 226–227.

Bouchard, T. J., & Hare, M. (1970). Size, performance, and potential in brainstorming groups. *Journal of Applied Psychology, 54,* 51–55.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology, 36,* 1224–1241.

Chatterjea, R. G., & Mitra, A. (1976). A study of brainstorming. *Manas, 23,* 23–28.

Cohen, D., Whitmyre, J. W., & Funk, D. W. (1960). Effect of group cohesiveness and training upon creative thinking. *Journal of Applied Psychology, 44,* 319–322.

Collaros, P. A., & Anderson, L. R. (1969). Effect of perceived expertness upon creativity of members of brainstorming groups. *Journal of Applied Psychology, 53,* 159–163.

Dillon, P. C., Graham, W. K., & Aidells, A. L. (1972). Brainstorming on a "hot" problem: Effects of training and practice on individual and group performance. *Journal of Applied Psychology, 56,* 487–490.

Dunette, M. D., Campbell, J., & Jaastad, K. (1963). The effect of group participation on brainstorming effectiveness for two industrial samples. *Journal of Applied Psychology, 47,* 30–37.

Gurman, E. B. (1968). Creativity as a function of orientation and group participation. *Psychological Reports, 22,* 471–478.

Harari, O., & Graham, W. K. (1975). Tasks and task consequences as factors in individual and group brainstorming. *The Journal of Social Psychology, 95,* 61–65.

Harkins, S. G., & Jackson, J. M. (1985). The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin, 11,* 457–465.

Harkins, S. G., Latané, B., & Williams, K. (1980). Social loafing: Allocating effort or taking it easy? *Journal of Experimental Social Psychology, 16,* 457–465.

Harkins, S. G., & Petty, R. E. (1982). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology, 43,* 1214–1229.

Henchy, R. & Glass, D. C. (1968). Evaluation apprehension and the social facilitation of dominant and subordinate responses. *Journal of Personality and Social Psychology, 10,* 446–454.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology, 10,* 371–384.

Jablin, F. M. (1981). Cultivating imagination: Factors that enhance and inhibit creativity in brainstorming groups. *Human Communication Research, 7,* 245–258.

Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology, 44,* 78–94.

Lamm, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideational proficiency (brainstorming). *European Journal of Social Psychology, 3,* 361–387.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37,* 822–832.

Madsen, D. B., & Finger, J. R. (1978). Comparison of a written feedback procedure, group brainstorming, and individual brainstorming. *Journal of Applied Psychology, 63,* 120–123.

Maginn, B. K., & Harris, R. J. (1980). Effects of anticipated evaluation on individual brainstorming performance. *Journal of Applied Psychology, 65,* 219–225.

Milton, G. A. (1965). Enthusiasm vs. effectiveness in group and individual problem-solving. *Psychological Reports, 16,* 1197–1202.

Olson, M. (1965). *The logic of collective action.* Cambridge, MA: Harvard University Press.

Osborn, A. F. (1957). *Applied imagination* (rev. ed.) New York: Scribner.

Pape, T., & Bölle, I. (1984). Einfallsproduktion von Individuen und Dyaden unter "Brainstorming"-Bedingungen [Brainstorming productivity of individuals and dyads]. *Psychologische Beiträge, 26,* 459–468.

Restle, F., & Davis, J. H. (1962). Success and speed of problem solving by individuals and groups. *Psychological Review, 69,* 520–536.

Rotter, G. S., & Portugal, S. M. (1969). Group and individual effects in problem solving. *Journal of Applied Psychology, 53,* 338–341.

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

Street, W. R. (1974). Brainstorming by individuals, coacting and interacting groups. *Journal of Applied Psychology, 59,* 433–436.

Stroebe, W., & Frey, B. S. (1982). Self-interest and collective action: The economics and psychology of public goods. *British Journal of Social Psychology, 21,* 121–137.

Taylor, D. W., Berry, P. C., & Block, C. H. (1958). Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly, 3,* 23–47.

Taylor, D. W., & Faust, W. L. (1952). Twenty questions: Efficiency in problem solving as a function of size of group. *Journal of Experimental Psychology, 44,* 360–368.

Torrance, E. P. (1970). Influence of dyadic interaction on creative functioning. *Psychological Reports, 26,* 391–394.

Vroom, V. H., Grant, L. D., & Cotton, T. W. (1969). The consequences of social interaction in group problem-solving. *Organizational Behavior and Human Performance, 4,* 77–95.

Williams, K., Harkins, S., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology, 40,* 303–311.

Received October 27, 1986
Revision received March 31, 1987
Accepted April 15, 1987 ■

TABLE 8-2 Performance Data of Brainstorming and Solitary Groups

	Face-to-Face Brainstorming Group	*Same Number of People Working Independently (Solitary Brainstorming)*
Quantity: The number of ideas generated	28	74.5
Quality: Percentage of "good ideas" as judged by independent experts who did not know whose ideas they were evaluating	20.8%	79.2%

Source: Diehl, M., & Stroebe, W. 1987. "Productivity Loss in Brainstorming Groups: Toward a Solution of a Riddle." *Journal of Personality and Social Psychology, 53,* 497–509.