Chapter 9

Why Groups are less Effective than their Members: On Productivity Losses in Idea-generating Groups

Wolfgang Stroebe
University of Utrecht
and
Michael Diehl
University of Mannheim

ABSTRACT

There is pervasive evidence that people produce more ideas and more good ideas when working alone rather than in groups. This chapter will first review the evidence for the productivity loss in brainstorming groups and then evaluate the various theoretical explanations which have been offered to account for these findings in the light of empirical research. This evidence suggests that the productivity loss in idea-generating groups is caused mainly by mutual production blocking due to the constraint on groups that members can talk only in turn. We then discuss various strategies that have been developed to overcome the disruptive effects of production blocking. However, since so far even the most successful strategies have raised the productivity of group members only to the level they would have achieved if they had worked individually, our final section discusses an ‘illusion of group effectivity’ as a reason for the persistence of the belief that group discussions are an effective means of generating ideas.
INTRODUCTION

There is a discrepancy between everyday beliefs and scientific evidence about the effectiveness of group discussion as a method of idea generation. Academics, politicians, and the business community appear to be firmly convinced that groups can stimulate individual creativity, and procedures such as group brainstorming are still widely used in commercial organizations and advertising agencies (e.g. Adriani et al., 1989; Grossman, 1984; Thaima & Woods, 1984). Yet scientific evidence has consistently demonstrated that people produce many more ideas when they work alone rather than in groups (Diehl & Stroebe, 1987, 1991; Lamm & Trommsdorff, 1973; Mullen, Johnson, & Salas, 1991).

GROUP BRAINSTORMING: PROCEDURE AND EMPIRICAL EVIDENCE

Brainstorming was developed by the advertising executive Alex F. Osborn (1953, 1957, 1963) as a technique to increase the effectiveness of group sessions at his advertising agency. He called these sessions ‘brainstorming sessions’, because ‘brainstorming means using the brain to storm a problem’ (Osborn, 1963). Brainstorming is based on two principles which Osborn called ‘deferment of judgement’ and ‘quantity breeds quality’. The principle of deferment of judgment implies a strict separation of idea generation and idea evaluation by having different people or at least different group sessions for idea generation and evaluation. A strict application of this principle should enhance the quantity of ideas produced, and thus, according to the second principle, also the quality. From these two principles, Osborn derived four rules for idea finding:

(1) Criticism is ruled out.
(2) ‘Free-wheeling’ is welcomed.
(3) Quantity is wanted.
(4) Combination and improvements are sought.

Osborn claimed that the adherence to these rules would more than double the ideas of group members.

Empirical studies which compared the productivity of real brainstorming groups with that of ‘nominal groups’ have consistently failed to support this claim. Productivity scores of nominal groups are based on the quantity or quality of the non-redundant ideas of n persons (with n being the number of members of the interaction groups), who work individually under brainstorming rules. The ideas are non-redundant in the sense that ideas that have been suggested several times in a nominal group of isolated individuals have been
eliminated, because in real groups it would be inappropriate for members to voice the same idea several times. Quantity is measured in terms of number of (non-redundant) ideas. Measures of quality are based on ratings of these ideas on dimensions such as originality or feasibility. Most studies that analyze quality use a measure of total quality (i.e. the sum of the quality ratings of the ideas produced by a given subject or group). Because total quality is highly related to number of ideas, some authors have preferred to use average quality. We have argued that the number of good ideas is the most appropriate measure of quality, because brainstorming is assumed to increase the production of good ideas. An idea can be defined as ‘good’ if it received a score above a chosen cutoff point on the quality ratings. The productivity of non-interactive nominal groups presents a baseline of the ideas produced by groups of individuals who are neither helped nor inhibited by group interaction.

In a recent meta-analytic review of this research, Mullen, Johnson, and Salas (1991) drew the following conclusions:

The general combinations reveal that, for both quantitative and qualitative operationalizations, productivity loss in brainstorming groups is highly significant and of strong magnitude. Thus, the quantitative productivity loss engendered by brainstorming groups is not trivially small; also, . . . quantitative productivity loss is not compensated for by an increase in the quality of productivity in brainstorming groups (p. 18).

In fact, measures of quantity and quality are typically highly correlated. For example, Diehl and Stroebe (1987, 1991) reported correlations of approximately 0.80 between number of ideas and number of good ideas.

POTENTIAL CAUSES OF THE PRODUCTIVITY LOSS IN BRAINSTORMING GROUPS

In a seminal discussion of the causes of productivity loss in interacting groups, Steiner (1972) distinguished two types of process losses, namely motivation losses and co-ordination losses. Two of the four interpretations offered to account for the productivity loss of interactive brainstorming groups attribute the inferior performance of these groups to motivational factors, suggesting that some aspect of the group situation lowers the work motivation of group members below that of subjects who work individually (free riding, production matching). The other two interpretations see problems in co-ordinating individual contributions into a group product as the major cause of the lower productivity of brainstorming groups (evaluation apprehension, production blocking). Co-ordination losses occur when knowledge or skills that are available to the individuals who form the group are not co-ordinated optimally into a group product. For example, members of a task group may fail to
suggest potential solutions to the group because of poor communication within the group. Similarly, groups competing in a rope-pulling contest may not reach their optimal performance because members are not pulling at exactly the same time or in exactly the same direction.

**Motivation Losses in Group Brainstorming**

*Free Riding*

This interpretation (e.g. Harkins & Petty, 1982; Stroebe, 1981; Stroebe & Frey, 1982) assumes that members of brainstorming groups are tempted to free ride on the efforts of others for two reasons: first, group members expect their ideas to be pooled to a group product and may therefore perceive their individual contribution as less identifiable. Identifiability is considered important because it is assumed that subjects in brainstorming experiments work hard mainly because they are instructed to do so by the experimenter. Second, group members may also feel that their contributions are less important to the group product (i.e. more dispensable) than individuals who brainstorm individually. The impression that one's ideas are dispensable is the more likely to arise, the more instructions emphasize originality of ideas rather than mere quantity.

The emphasis on quantity turns brainstorming into an additive task (Steiner, 1972). With additive tasks the group product is equal to the sum of individual contributions. Thus, every single contribution counts. An emphasis on originality should change brainstorming into a disjunctive task (Steiner, 1972). With disjunctive tasks the group product is equal to the best performance (i.e. the most original idea). Thus, group members who perceive themselves as less competent might feel that their ideas are dispensable (e.g. Kerr & Bruun, 1983).

*Production Matching*

This explanation has been suggested by Paulus and colleagues (e.g. Paulus & Dzindolet, 1993). They argued that brainstorming is a novel situation for most people. Therefore, individuals are likely to be uncertain about the appropriate level of individual productivity. Subjects will be motivated to reduce

---

1 We use the term 'free riding' here in the sense in which it is employed in the economics literature (e.g. Olson, 1965; cf. Stroebe & Frey, 1982) to refer to motivation losses caused either by a lack of identifiability or by perceived dispensability. Our conceptualization differs from that introduced to the social psychological literature by Kerr and colleagues (e.g. Kerr & Bruun, 1983), who restrict the term 'free riding' to motivation losses caused by the perception that one's contributions are dispensable. Motivation losses caused by the perception that one's contribution cannot be identified are referred to as social loafing by Kerr and colleagues.
uncertainty by comparing their performance to that of other subjects. In case of discrepancies there will be a matching of performance.

Without additional assumptions, the matching hypothesis would predict a convergence of individual contributions towards the mean rather than a lower performance of interactive as compared to nominal groups. Paulus and Dzindolet (1993) make two additional assumptions to account for the lower productivity in brainstorming groups. First, in the early phases of a group brainstorming session individual productivity is particularly inhibited through processes of production blocking or evaluation apprehension. Due to production matching, this low performance becomes the group norm and keeps productivity down, even after the influence of some of these inhibiting factors has been weakened. Second, in laboratory settings there is little incentive for performing at a high level. Poorly producing group members should therefore be more influential than highly productive members in setting the performance standards.

Co-ordination Losses in Group Brainstorming

Evaluation Apprehension

This explanation (e.g. Collaros & Anderson, 1969) suggests that the brainstorming instructions are not completely successful in eliminating group members' fear of negative evaluations. Since group members will monitor the quality of their own ideas, they will suppress certain ideas and not share them with other group members. We classify process losses due to evaluation apprehension as co-ordination losses, because this explanation does not assume that group members lack motivation to produce ideas but that they are too critical of some of their own ideas to suggest them in a group session.

Mutual Production Blocking

Blocking (e.g. Lamm & Trommsdorff, 1973) results from the constraint on groups that members of real groups can talk only in turn whereas there are no constraints on members of nominal groups to talk. Mutual production blocking might lower idea generation in interacting groups for a number of reasons. Being prohibited from verbalizing their ideas at the time when they occur, group members might forget them or suppress them because they seem less relevant or less original at a later time. Subjects might also be unable to think of further ideas during the waiting time, either due to limitations of short-term memory (i.e. they may have to rehearse the ideas already produced in order not to forget them) or because exposure to the ideas of others is distracting and interferes with subjects' own thinking. Finally, since brainstorming groups are given the same amount of time as individuals, the blocking effect could be
due to the fact that subjects in real groups have less time to express their ideas than those who work individually.

Conclusions

It should be emphasized that these different interpretations are not mutually exclusive and that several of the potential causes of productivity loss discussed in this section could be jointly responsible for the productivity loss in brainstorming groups. The influence of different inhibiting factors on brainstorming productivity could be sequential as suggested by the production-matching hypothesis, but it could be simultaneous (e.g. production blocking and evaluation apprehension acting simultaneously). However, as we will see from the following review of the research evidence, mutual production blocking emerges as the predominant cause of the productivity loss in brainstorming groups.

TRACKING DOWN THE CAUSES OF PRODUCTIVITY LOSS: A REVIEW OF THE EVIDENCE

Free Riding

Our own interest in brainstorming dates back to the development of a theory of group productivity (Stroebe & Frey, 1982) which derived from the economic theory of public goods (Buchanan & Tullock, 1962; Olson, 1965; Samuelson, 1954). This theory allowed us to account for a wide range of motivation losses in situations where group products were distributed independently of individual contributions (Stroebe & Frey, 1982). The interpretation of the low productivity of brainstorming groups in terms of free riding was originally offered by us to account for discrepancies in the findings of research which examined evaluation apprehension as a cause of productivity loss in brainstorming groups (Stroebe, 1981; Stroebe & Diehl, 1983).

Apparent support for evaluation apprehension came from a study by Collaros and Anderson (1969) who manipulated perceived expertise of group members in brainstorming groups. These authors reasoned that social inhibition would be greater, the more group members perceived other members as experts. In their all-expert condition, each member of a brainstorming group was told that all other members had previous experience with such groups, whereas in a one-expert condition members were told that only one unidentified member had previously worked in such a group. In a third, no-expert condition, no such instructions were given. Consistent with predictions, productivity was highest in the no-expert condition and lowest in the all-expert one. Furthermore, in a post-experimental questionnaire subjects in the expert conditions indicated
greater feelings of inhibition and reluctance to offer ideas than did those who brainstormed without receiving this kind of instruction.

However, results of a later study by Maginn and Harris (1980), who manipulated evaluation apprehension in subjects working individually, appeared to be inconsistent with these findings. 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 were listening to their ideas and rating them in terms of quality and originality. The authors reasoned that if evaluation apprehension accounted for the low productivity of interacting groups, then introducing evaluation apprehension in nominal groups should lower the productivity of subjects who brainstorm individually to that shown by subjects who brainstorm in groups. In contrast to these expectations, the presence of observers did not result in a significant decrease in individual productivity.

An interpretation in terms of free riding could resolve the apparent discrepancy between the results of Collaros and Anderson and Maggin and Harris and could also account for other findings reported in the brainstorming literature. This analysis would suggest that the effect of the manipulation of perceived expertise on productivity in the study by Collaros and Anderson 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 other members to be more qualified for the job than if they believe that the other members possess the same qualifications as themselves. The failure of the manipulation used by Maginn and Harris to influence brainstorming productivity is consistent with a free-rider interpretation, because their manipulation should have affected neither dispensability nor identifiability.

The free-rider interpretation could also account for other findings of brainstorming research. For example, the rule introduced in a study by Bouchard (1972) that members have to give their ideas in a fixed sequence and have to announce a ‘pass’ if they have nothing to say when their turn comes should increase identifiability of individual contributions. Thus, the finding that the introduction of such a rule increases group productivity is consistent with an interpretation in terms of free riding (Bouchard, 1972). Increases in group size should increase dispensability and decrease identifiability. Thus, the finding that the productivity loss of brainstorming groups increases monotonically with increasing group size would be consistent with an interpretation in terms of free riding (Bouchard & Hare, 1970).

There was evidence claiming that free riding occurs with brainstorming tasks (e.g. Harkins & Petty, 1982). However, since this research did not compare the productivity of real and nominal groups within the same study, these

---

2 If the expectation to be evaluated by a group of experts had increased identifiability (or decreased dispensability), economic theory would have predicted an increase in productivity rather than a productivity loss.
findings were not conclusive with regard to the role of free riding as an inhibitor of productivity in group brainstorming. Therefore, our first experiment was conducted to test the free-rider hypothesis (Diehl & Stroebe, 1987, Experiment 1). Subjects who worked either in four-person groups or individually had to produce ideas on how to improve the relationship between German nationals and (foreign) guest workers, a very important topic not only for student subjects. In group as well as in individual sessions, subjects were led to believe either that their ideas would be pooled (collective assessment) or that they would be assessed individually (personal assessment). We reasoned that if the difference in productivity between groups and individuals was due to differences in assessment expectations, then the explicit manipulation of assessment should eliminate the impact of type of session. At that time, we were so convinced that free riding was the sole cause of the productivity loss that we did not think it necessary to include control groups without assessment instructions in our design.

Results were not supportive of our hypothesis (see Table 9.1). Although subjects produced more ideas and marginally more good ideas under personal rather than collective assessment instructions, the explicit manipulation of assessment instructions did not eliminate the impact of type of session. Nominal groups produced many more ideas and more good ideas than real groups. Furthermore, type of session explained 83% of the variance of idea production whereas assessment instructions accounted only for 8%. This finding suggested that free riding was, at best, of secondary importance as a cause of productivity loss in brainstorming groups.3

Table 9.1 Average number and quality of ideas suggested by real and nominal four-person brainstorming groups working under personal- versus collective-assessment instructions (from Diehl & Stroebe, 1987, Experiment 1; © APA 1987)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of Ideas</th>
<th>Number of Good Ideas</th>
<th>Average Originality</th>
<th>Average Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal</td>
<td>32.33</td>
<td>3.00</td>
<td>2.52</td>
<td>2.90</td>
</tr>
<tr>
<td>Collective</td>
<td>23.66</td>
<td>2.00</td>
<td>2.49</td>
<td>3.07</td>
</tr>
<tr>
<td>Nominal group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal</td>
<td>84.33</td>
<td>13.33</td>
<td>2.46</td>
<td>2.60</td>
</tr>
<tr>
<td>Collective</td>
<td>64.66</td>
<td>5.66</td>
<td>2.43</td>
<td>2.70</td>
</tr>
</tbody>
</table>

Note: Lower numbers indicate higher originality and feasibility.

3 This study manipulated the extent to which the experimenter could identify individual contributions because we assumed that subjects worked hard in these experiments mainly because they were instructed to do so by the experimenter.
WHY GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS

In retrospect, the failure of free riding to account for the productivity loss in brainstorming groups is not all that surprising. People free ride on public goods to save the costs of a contribution for a good which they expect to be able to consume without making a contribution. The production of ideas in a brainstorming group involves few costs, especially for student subjects who enjoy idea generation. There is little effort and there are no time costs because subjects are committed for the whole group session.

Production Matching

The hypothesis that processes of production matching are responsible for the low productivity of brainstorming groups has been examined in a series of studies conducted by Paulus and Dzindolet (1993). Their research focused mainly on evidence for performance matching. Assuming that individuals will tend to compare their performance with that of others and will be motivated to match their performance with that of others, one would expect greater similarity among members of interactive rather than nominal groups (Paulus & Dzindolet, 1993). Consistent with this hypothesis, a comparison of the number of ideas produced by nominal and interacting two-person groups resulted in fairly high correlations for the interacting but not the nominal groups in Experiment 1. Experiment 2 provided further evidence for a matching process in dyadic brainstorming. However, both experiments failed to reveal significant differences in the productivity of interacting and nominal groups.

The third study, in which the authors attempted to demonstrate performance matching in groups of four, yielded some evidence supporting performance matching but again failed to replicate the performance advantage of real over nominal groups. Whereas a failure to find a significant productivity loss in interactive groups is not unusual with dyads, a fact which we have attributed to the virtual absence of blocking effects under these conditions, the absence of significant productivity losses with four-person groups is quite unusual in this area. Thus, even though their first three studies provide persuasive evidence for performance matching, in none of these studies was the productivity of the interacting groups significantly lower than that of the nominal groups. The authors were therefore unable to test whether the degree of matching was in any way related to the extent of productivity losses in their interacting groups. Such evidence would be necessary, even though not sufficient, to establish that performance matching contributes to the lower productivity of brainstorming groups.

In a further study, Paulus and Dzindolet tested the hypothesis that the low level of performance of interacting groups could be counteracted if the groups were provided with high normative standards. For half of their nominal and interacting groups standards were influenced by giving individual subjects or
groups information about the number of ideas supposedly generated by the typical subject (information condition). In fact, these standards were two and a half times greater than their typical performance. In the 'no information' condition no such information was given. These manipulations resulted in significant main effects for type of session and information condition. Nominal groups generated more ideas than interactive groups and the information manipulation led to the generation of a larger number of ideas. However, there was no interaction between the two factors.

To support the assumption that differences in goal levels are responsible for the lower productivity of brainstorming groups it would have been necessary to demonstrate that goal setting was *more* effective in groups than in individual sessions. If individual brainstormers set themselves higher performance goals than groups and *therefore* produce more, imposing substantially higher goals for individuals and groups should result in less performance increase for individuals, who already perform at their highest level than for groups who are assumed to work far below their best effort. In contrast, the results suggest that both groups and individuals work below their best effort and that therefore performance can be increased by setting higher goals. That goal setting can increase performance in brainstorming groups is, however, well known from earlier studies (e.g. Hyams & Graham, 1984). This experiment also provided some evidence which is inconsistent with one of the key hypotheses of production matching. According to the authors:

> the inhibitory aspects of interactive brainstorming ensure that group members will have an initially low level of performance. The matching process helps to make this a normative level of performance for the group, and *this level is maintained throughout the task session* (p. 580; italics ours).

To test this hypothesis, Paulus and Dzindolet coded the brainstorming productivity into 5-minute blocks. They found a main effect for time blocks and an interaction between type of session and time blocks. (The effects for goal setting need not be discussed here.) Performance decreased over time and this decrease was greater for nominal than for interacting groups. However, although there was less decrease in productivity for interacting than for nominal groups, the productivity for interacting groups also decreased substantially from 12.9 ideas during the first 5 minutes to 5.5 ideas at the end. This decrease in performance seems inconsistent with the assumption that interacting groups adopt a low performance standard at the beginning of the group session which then becomes normative for the remainder of the session. If members of interacting groups initially adopted as normative a standard which was far below their level of ability, it would seem unlikely that within the course of a 25-minute group session they would run so short of ideas as to be unable to conform to this group norm.
WHY GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS

In conclusion, production matching presents an intriguing and psychologically plausible explanation for the productivity loss of brainstorming groups. However, the evidence presented so far by Paulus and Dzindolet (1993) is insufficient to support their hypothesis that production matching is responsible for, and even contributes to, the productivity loss of brainstorming groups.

Evaluation Apprehension

Social inhibition is such a pervasive feature of social life that the findings of Maginn and Harris (1980) that brainstorming should somehow be exempted had always been puzzling. We suspected that Maggin and Harris might have inadvertently minimized evaluation apprehension by the nature of the problems used in their study. It seems plausible that individuals will be most likely to censor their responses if they fear that certain answers might reveal socially undesirable attitudes or embarrassing gaps in knowledge. It was unlikely that the 'thumbs' problem ('What would happen if everyone after a certain date had an extra thumb on each hand?') and the 'energy' problem ('How can we reduce gasoline consumption?') used by Maginn and Harris would require such self-revelation from their subjects.

To test this assumption, an experiment was conducted in which individuals brainstormed either on topics which were highly involving and controversial (e.g. 'How can the number of guest workers be reduced?') or uncontroversial (e.g. 'How can entertainment programs on television be improved?' (Diehl & Stroebe, 1987, Experiment 2). A controversial topic was defined as one that forced subjects to argue not only against their own private opinion but also against a position widely shared by other subjects. For example, a subject who suggested that guest workers should be paid a premium contingent on their willingness to return to their home countries might fear that other subjects would suspect him or her of right-wing attitudes. Evaluation apprehension was manipulated by leading subjects either to believe that their ideas would be evaluated by other people (e.g. peers, judges) or by not creating this impression.

To account for the null findings of Maginn and Harris, we expected an interaction of evaluation apprehension and the controversial nature of the brainstorming topics. The manipulation of evaluation apprehension should have greater impact on brainstorming productivity for controversial rather than uncontroversial topics. Instead of the expected interaction, our manipulations resulted in two main effects. Subjects produced fewer ideas with controversial rather than uncontroversial topics and under high rather than low evaluation apprehension.

That the expectation of being evaluated significantly reduced idea generation even for the uncontroversial topic is inconsistent with the results published by Maginn and Harris. Since we used the same manipulation of
evaluation apprehension employed in their study it is difficult to understand why this manipulation had no impact on their subjects. Furthermore, the finding that evaluation apprehension, in contrast to the conclusions drawn by Maginn and Harris (1980), did reduce productivity under individual brainstorming conditions raised the possibility that evaluation apprehension might also contribute to the productivity loss in brainstorming groups.

This hypothesis was tested in an experiment which manipulated evaluation apprehension under individual and group conditions (Diehl & Stroebe, 1987, Experiment 3). Subjects who were asked to brainstorm on the guest worker topic either individually or in groups were led to believe that either their performance would be videotaped to be shown to their fellow students in a class demonstration (high evaluation apprehension) or merely recorded on audiotape to be analyzed by the experimenter. We hypothesized that if the productivity loss in real brainstorming groups is partially or fully caused by the high level of evaluation apprehension induced in the group session, the induction of evaluation apprehension should have greater impact on individual brainstormers (who normally have low evaluation apprehension) than on subjects interacting in real groups (where apprehension is already at a high level).4

Instead of the expected interaction between evaluation apprehension and type of session on productivity, the manipulations resulted in two main effects. Subjects produced fewer ideas in group rather than individual sessions and under high rather than low evaluation apprehension. The interaction did not even approach significance. Furthermore, type of session accounted for more than 70% of the variance in brainstorming productivity, suggesting that even if evaluation apprehension accounted for some part of the productivity loss in real brainstorming groups the impact of this variable is minimal when compared to that of type of session. This would suggest that there are still other powerful causes of the productivity loss which have not been identified in our experiments.

**Production Blocking**

Since interaction is a defining feature of brainstorming groups production blocking cannot be eliminated in real face-to-face groups. However, if blocking was a major cause of the productivity loss in real brainstorming groups, introducing blocking into nominal groups should substantially reduce their productivity. We therefore conducted an experiment in which blocking was manipulated in individual brainstorming sessions by a specially constructed communication apparatus (Diehl & Stroebe, 1987, Experiment 4).

---

4 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.
Subjects in the experimental conditions worked in individual rooms and interacted via a communication apparatus which acted like traffic lights at an intersection, allowing only one party to pass. They were seated in front of a display of four lights, each light assigned to one of the group members. The subject’s own light was green and the other three lights were red. As soon as one subject talked into the microphone a voice-activated sensor switched on the (green) light of the subject who was talking and a red light on the displays of the other subjects. When the subject stopped talking for more than 1.5 seconds, his or her light was switched off. This communication apparatus was used in three experimental conditions. In two of these conditions, subjects were asked to speak only when none of the red lights was on (i.e. when no-one else was speaking), thus simulating the interdependence of contributions typical of real groups. One of these conditions (EC I) replicated the usual brainstorming conditions in all important aspects, except that subjects sat in separate rooms and communicated via an intercom. In the second condition (EC II), subjects were not provided with earphones. Thus, even though they were mutually interdependent through the light system, they could not overhear each others’ ideas which should have minimized in-group evaluation apprehension. The third condition (EC III) was identical to EC II, except that after subjects had been informed about the function of the lights they were instructed to disregard them and to talk whenever they had anything to say. Thus, there should have been no blocking in this condition. In addition to these experimental conditions, there were also two control conditions: a real group of subjects who brainstormed jointly in the same room (RGC) and a nominal group condition (NGC).

It can be seen from Figure 9.1 that brainstorming productivity was affected only by the presence or absence of blocking. The planned comparison between the three conditions with blocking (EC I, EC II, RGC) and the two conditions without blocking (EC III, NGC) resulted in a highly significant effect. This analysis further revealed that 96% of the variance due to experimental conditions could be attributed to this comparison. A post hoc comparison (Newman–Keuls test) resulted in significant differences between but not within blocking and non-blocking conditions. Thus, the findings of this experiment provide strong support for an interpretation of productivity losses in terms of blocking, leaving very little room (i.e. variance) for other differences between conditions.

That productivity in the real groups did not differ significantly from that of the groups who communicated via the communication apparatus suggests that the communication apparatus succeeded in replicating the situation of interacting subjects. Furthermore, the finding that, as long as subjects had to talk in turn, productivity under conditions where subjects could overhear each other’s ideas via the intercom was not substantially lower than under those where subjects could not hear each other supports the conclusion from the
previous study that evaluation apprehension is not a major cause of productivity loss in brainstorming groups.

**Mediators of the Blocking Effect**

To have demonstrated that the productivity loss in brainstorming groups is due to blocking still leaves the question about the psychological processes underlying this effect unanswered. One potential interpretation in terms of thought distraction assumes that hearing other subjects present their ideas is distracting and prevents subjects from developing their own thoughts. If this were true we should have observed a major difference in our blocking experiment between the productivity of the experimental groups who were blocked by the lights and could overhear other subjects and those who were blocked but could not overhear other subjects. That there is no significant difference between blocking conditions with communication and those without either in the experiment reported above or in two further studies makes this kind of thought distraction an unlikely candidate.

The finding that the presence or absence of communication did not significantly affect productivity also eliminates an interpretation of the blocking effect in terms of a re-evaluation of ideas. It would seem plausible that the exposure to ideas of other group members during the waiting period might result in a re-evaluation of the subject's own ideas. Subjects might find some of their own ideas no longer worth mentioning after hearing another group member expressing a similar idea. Obviously this
WHY GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS

A third plausible cause for the productivity loss in brainstorming groups is the difference in the *speaking time* available to subjects who brainstorm in group as compared to individual sessions. 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). However, if a subject who brainstorms individually is given the same amount of time as four subjects who brainstorm in a group, subjects in individual sessions have approximately four times as much time to express their ideas than those who work in groups. To investigate this possibility, we varied the length of time available to groups and individuals in the next study.

Assuming that the pool of ideas on any given topic is limited, subjects are likely to exhaust these ideas towards the end of a session. Therefore, if competition for speaking time reduces the productivity of interacting groups, we would expect the productivity gap between nominal and real groups to narrow over the course of a session. Given enough time, groups might even become more productive than individuals who may have exhausted their pool of ideas towards the end of a session. As a consequence of such a decrease in productivity over time, one would also predict that the difference in brainstorming productivity of individuals and groups should be smaller, the greater the length of the sessions in a given study.

These hypotheses were tested in an experiment which varied length of session from 10 to 20 minutes for subjects who participated in either group or individual sessions. Since it has often been argued that four-person groups should have four times the time of individual sessions, we also introduced a nominal group control condition of 5 minutes to control for real groups of 20 minutes and a real group control condition of 40 minutes to control for the nominal group of 10 minutes.

Table 9.2 presents the average number of ideas and number of good ideas suggested by real and nominal 4-person groups under the different time conditions. Analyses of variance for quantity and quality resulted only in main effects for type of session and length of session. Nominal groups produced significantly more ideas and more good ideas than real groups and quantity and quality of output were also greater in longer rather than shorter sessions. In contrast to predictions there was no indication of an interaction. The productivity gap between individual and group sessions did not seem to vary with length of session. Similarly, a breakdown of the 20-minute session into 10 two-minute intervals did not provide any evidence for a significant narrowing of the productivity gap over time.

One reason for our failure to find the expected interaction of time and type of session on brainstorming productivity could have been that extending the length of time to 20 minutes did not provide individual brainstormers with
Table 9.2 Average number of ideas and good ideas suggested by real and nominal brainstorming groups working for different lengths of time (from Diehl & Stroebe, 1991, Experiment 1; © APA 1991)

| Length of Session | Number of Ideas | | Number of Good Ideas | |
|-------------------|----------------|----------------|-------------------|
|                   | Real Group | Nominal Group | Real Group | Nominal Group | |
| **Experimental Condition** | | | | |
| 10 min            |          |              |          |              | |
| M                 | 38.00        | 57.25        | 8.00      | 32.25        | |
| SD                | 7.14         | 12.07        | 3.60      | 7.59         | |
| 20 min            |          |              |          |              | |
| M                 | 60.75        | 84.25        | 37.00     | 51.50        | |
| SD                | 17.61        | 21.00        | 14.93     | 25.31        | |
| **Control Condition** | | | | |
| 5 min             |          |              |          |              | |
| M                 | —           | 47.50        | —         | 17.50        | |
| SD                | —           | 4.15         | —         | 6.18         | |
| 40 min            |          |              |          |              | |
| M                 | 81.25       | —            | 38.75     | —            | |
| SD                | 19.00       | —            | 15.69     | —            | |

*Note: Dashes indicate that no subjects were run under these conditions.*

sufficient time to exhaust their pool of ideas. However, this explanation seems less plausible in view of the fact that the analysis of idea production over time showed little indication of a narrowing of the gap. Furthermore, Kanekar and Rosenbaum (1972), who varied the time limit of groups and individuals working on an anagram task from 15 to 105 minutes, also failed to find any interaction between length and type of session. However, as reported earlier, Paulus and Dzindolet (1993) did find a greater decrease in productivity over time for interacting rather than nominal groups. Since their session was only 5 minutes longer than ours, this discrepancy is difficult to explain.

Our study resulted in one finding which could have been interpreted as supportive of an interpretation in terms of speaking time. Interacting groups who had been given 20 and 40 minutes produced significantly more ideas than nominal groups who had been given only a quarter of that time (i.e. 5 or 10 minutes). This might suggest that groups can be more effective than individuals when given the same amount of time per member to report their ideas. However, extension of the length of a session extends not only the speaking time available for subjects but also the time subjects have to think. Thus, even though the four subjects who participate in a 40-minute group session have, on average, only 10 minutes during which to report their ideas, they have 40
WHEN GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS

minutes to generate ideas as compared to the 10 minutes available to subjects in individual sessions.

To eliminate this ambiguity, speaking time was manipulated directly in a further study (Diehl & Stroebe, 1991, Experiment 2). Subjects who brainstormed individually for 20 minutes were either allowed to report their ideas during the whole period or reporting time was limited to 5 out of 20 minutes. When speaking time was limited in this way subjects could talk at any time they wanted during a session. But the time they used for speaking was measured with a voice-controlled clock and subjects, who were informed of the amount of time used on a monitor, were not allowed to report further ideas after they had used up their 5 minutes. The findings of this study were quite clear. Restriction of speaking time to 5 minutes did not reduce the productivity of nominal groups, which was significantly higher than that of a control group who brainstormed in a 20-minute group session.

With speaking time limited to 5 minutes per person in the individual condition, the major difference between the situation in the real as compared to the nominal group is the fact that subjects who brainstorm individually under limited speaking time are allowed to present their ideas whenever they want to. Subjects who brainstorm in groups, on the other hand, usually have to wait until some other group member finishes talking. The findings of this study therefore point to the waiting time as the factor responsible for the productivity loss in brainstorming groups. More specifically, these findings suggest that group members are unable to use their waiting time productively.

A Meta-analytic Critique

In 1991 Mullen, Johnson, and Salas (1991) published a meta-analytic review of brainstorming research. On the basis of this review they rejected our conclusion that production blocking was the major cause of the productivity loss in brainstorming groups. This disagreement resulted in a discussion of the relationship between primary and meta-analytic evidence (Bond & van Leeuwen, 1991; Stroebe & Diehl, 1991). Since this discussion illustrates a number of important points about the utility of meta-analysis, we will briefly review the main findings of the meta-analysis of Mullen and colleagues as well as the main arguments of the ensuing discussion.

The meta-analysis conducted by Mullen, Johnson, and Salas (1991) examined the impact of group size, experimenter presence (whether the experimenter was present or absent during the procedure), response mode (reporting ideas by talking into a microphone versus writing), type of groups (e.g. 'standard nominal groups' with individuals brainstorming in separate rooms versus 'together nominal groups' where individuals work in the same room) on brainstorming productivity. The impact of these variables on productivity was assessed through a series of independent comparisons of the
effect sizes for each of the predictors. Mullen and colleagues concluded that productivity losses are related to (1) group size (greater losses in larger groups) and experimenter presence (greater loss when the experimenter was present rather than absent), (2) response mode (greater loss with tape-recorded than with written responses), and (3) type of nominal group (greater losses when nominal group is of the Alone rather than the Together type).

On the basis of these results Mullen and colleagues drew the following sweeping conclusions:

These results contradict the interpretation of productivity loss suggested by Diehl and Stroebe (1987) as being largely attributable to the mechanics of parsing up a given amount of time amongst a certain number of group members. Certainly, when blocking is experimentally manipulated (as in Diehl & Stroebe, 1987, Experiment 4), it can yield a productivity loss. However, simply because productivity loss can occur after a manipulation of blocking, this does not mean that all instances of productivity loss must result from production blocking. Our results indicate that procedural mechanisms might contribute to productivity loss in brainstorming groups, but that their influence is likely to be secondary to that of the more predictive social psychological mechanisms (p. 19).

In our reply to Mullen and colleagues we raised three points of criticism. First, we argued that although the variables assessed in their analysis had the advantage of being easily inferred from research reports, they had the disadvantage of being 'only remotely related to the theoretical mechanisms that they [the authors] hope to examine' (Stroebe & Diehl, 1991, p. 30). Since it is very unclear which psychological processes underlie the various effects reported by Mullen and colleagues it is difficult to draw any conclusions from their findings other than that many variables affect brainstorming productivity.

Second, we also questioned the appropriateness of the procedure employed by Mullen and colleagues to assess the impact of the predictor variables on the grounds that some of the variables they examined were confounded across experiments: 'For example, nominal groups of the Together type have to use a written response mode, whereas nominal groups of the Alone type typically use tape-recorded sessions' (Stroebe & Diehl, 1991, p. 30). With the type of independent statistical tests conducted by Mullen and colleagues it would be impossible to disentangle the correlation between their predictor variables.

That this latter criticism was justified was borne out by the findings of Bond and van Leeuwen (1991), who re-analyzed the data of Mullen and colleagues with a regression analysis which used effect size as a criterion and the various experimental variables as predictors. This method allows one to assess the independent impact of the various factors and is thus superior to the procedure employed by Mullen and colleagues. Bond and van Leeuwen found
that brainstorming losses have no independent relationship to subjects’ response mode whereas type of nominal group did. In addition, they also failed to find an effect of experimenter presence, which suggests that this variable may also have been confounded with some other factor in the primary literature.

Third, our most serious objection was to the negative conclusion of Mullen and colleagues regarding blocking, a conclusion which is clearly discrepant with the results of our experimental studies. Mullen and colleagues justified this conclusion with the argument that while all their findings could be explained by social psychological mechanisms, blocking could account only for the impact of response mode and group size. Indeed, if we had argued that blocking was the only cause of productivity losses in brainstorming groups, then the finding that variables unrelated to blocking lower brainstorming productivity would have forced us to retract this claim. Since we never had made such a claim, there was no need for such a retraction.

On what basis would a meta-analytic review be justified in contradicting the findings of an experimental study? Meta-analysis is a procedure which allows one to integrate results of large sets of studies and summarize their findings in terms of a common metric, the effect size. Thus, if there were a great number of studies which manipulated blocking and which resulted in findings discrepant with our own study (Diehl & Stroebe, 1987, Experiment 4), then a meta-analysis would rightly allow us to conclude that our own study was an outlier and that, overall, the effect size of blocking was not significant. In an ideal world, the meta-analysis would also identify factors which might have been responsible for the discrepancy between our own study and those conducted by others. However, as we pointed out in our reply to Mullen and colleagues: (1) our study is the only study to date, which manipulated blocking and (2) in all other studies reported in the literature blocking is confounded with the difference between interacting groups and nominal groups. Therefore, a meta-analytic review does not allow one to draw valid conclusions about the impact of blocking on brainstorming productivity.

Bond and van Leeuwen (1991), who were asked by the editor of the journal to comment on this exchange, did not accept our argument. They maintained that if Mullen and colleagues had made a more complete search of the literature or conducted a better analysis, their meta-analysis would have allowed them to explain the productivity loss in brainstorming groups. Yet Bond and van Leeuwen’s (1991) own attempt at resolving the brainstorming riddle meta-analytically left them somewhat bemused:

Having modeled the brainstorming literature, we do not know how to explain brainstorming losses. All of the explanations offered by Mullen et al. enjoy some support, but none can account for all of the published research findings to date (p. 39).
It is gratifying that Eagly and Wood (in press) in a chapter which discusses the relationship between meta-analysis and primary research fully agreed with our position. Commenting on our reply to the conclusions of Mullen and colleagues, Eagly and Wood stated their position as follows:

Stroebe and Diehl (1991) pointed out that their direct manipulations of moderators in primary research yielded findings that conflict with the between-studies correlational results of their meta-analysis [i.e. the meta-analysis of Mullen and colleagues]. Given the higher internal validity associated with experimental findings than between-studies analyses, greater certainty should be attached to the experimental results, as Stroebe and Diehl maintained.

Conclusions

We have reviewed a great deal of evidence which supports the hypothesis that the productivity loss in brainstorming groups is mainly (but not exclusively) caused by the fact that only one member of a group can speak at any given time. Thus, the feature which Osborn (1957) considered the main strength of the brainstorming technique, namely that members are exposed to and stimulated by the ideas of others, appears to be more of a disadvantage. Our findings suggest that, rather than being stimulated by the exposure to other ideas, group members appear to be unable to use the waiting time productively.

In designing his brainstorming technique, Osborn was under the impression that the creative potential of groups could be freed if it were possible to suppress evaluation apprehension. Hence, his instructions not to be critical. Whether this testifies to the effectiveness of his instructions or to an overestimation of the problem of criticism, our own research suggests that evaluation apprehension is only a minor cause of the productivity loss in brainstorming groups. Perhaps there is no creative potential of the group to be freed. But if there is, it will be necessary to learn to reduce or eliminate the impact of production blocking on idea generation in brainstorming groups. The following studies were conducted to achieve this goal.

OVERCOMING THE IMPACT OF PRODUCTION BLOCKING

Regulating the Waiting Time

The British are admired for the patience they display when waiting in queues. However, even foreigners would enjoy waiting in queues in Britain because there are clearcut and generally accepted norms which regulate waiting behavior. While Germans take great pride in beating others who have waited
WHY GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS

much longer to the goal post, queue jumping is discouraged in Britain. Thus, individuals can relax and simply wait until their turn comes.

Since turn taking in the type of leaderless discussion groups created in our laboratory is not well regulated, individuals wishing to report their own ideas have to be attentive to make use of gaps in the conversation. We wondered whether the need to expend this kind of cognitive effort might not contribute to the apparent inability of group members to utilize the waiting period. To test this hypothesis, we conducted a study in which norms were introduced to make the waiting time predictable or even controllable (Diehl & Stroebe, 1991, Experiment 3).

Subjects in this experiment brainstormed via the communication apparatus under one of three conditions:

1. ‘Unorganized waiting’ corresponded to the usual arrangement in brainstorming groups when members have to wait for a break in the discussion to suggest their own ideas.
2. With ‘Predictable waiting’ waiting time was made predictable by asking members to report their ideas in a fixed sequence. Thus, each member knew precisely when his or her turn would come.
3. With ‘Controlled waiting’ controllability was achieved by establishing a ‘speakers’ list’. Entering their number on the keyboard, subjects could reserve a place on this list at any time during the discussion. This condition is probably closest to the usual conditions in discussion groups, where members wishing to speak raise their hands and discussion leaders keep track of the order of speakers.

To our own surprise, the attempt to organize the waiting not only failed to improve brainstorming productivity but under some conditions it even had a negative impact. While groups produced, on average, 68 ideas when waiting was unorganized, they produced only 49 ideas when waiting was made predictable and 28 when it was controllable. Thus subjects seemed to be unable to utilize the waiting periods for idea production, even when they knew that their own turn would come soon. We speculated that the significant decrease in idea production under conditions where subjects could enter their name into a waiting list may have been caused by the fact that having to attend to the speakers’ list displayed on a video monitor requires further cognitive resources from our subjects and is thus distracting rather than helpful.

Providing External Storage for Ideas

One reason why subjects are unable to use the waiting time to produce ideas may be limitations of short-term memory. It seems possible that subjects who have thought of ideas which they are waiting to report have to rehearse them
in order not to forget them. If this assumption were correct, productivity could be improved by providing subjects with a means to store their ideas externally during the waiting period. Since even in the computer age, note pads are still the most common means for external storage of ideas, we manipulated the availability of external storage by either providing or not providing subjects with note pads to write down their ideas during a waiting period. However, since it seems possible that subjects are unable to make effective use of these note pads if they are distracted by group discussion, this study once more manipulated the availability of interpersonal communication. Finally, in order to preclude the uncontrolled variation of assessment expectations, we also manipulated assessment instructions in a three-factorial design.

These manipulations resulted in a complex pattern of findings. Subjects produced significantly more ideas when communication was unavailable rather than available and marginally more ideas when note-taking was permitted than when there was no opportunity to take notes. However, these two main effects were moderated by a marginally significant communication by note-taking interaction. As expected, note-taking tended to improve performance mainly when communication was unavailable rather than available. There was further indication that the positive impact of note-taking was most marked under personal assessment expectations without communication.

These findings indicate that being able to make a note of one's ideas while waiting for the previous contributor to finish talking did not lead to a general improvement in the quantity of ideas produced. Note-taking improved performance mainly when there was no communication and when subjects expected to be assessed personally. Under these conditions, when subjects were both motivated and able to make use of the note pad, they also judged the opportunity to take notes as most helpful in the post-experimental questionnaire. Since real groups are characterized by communication and collective assessment expectations, the provision of note pads to subjects who brainstorm in groups is unlikely to improve productivity.

**Electronic Brainstorming**

Another way to overcome the detrimental consequences of mutual production blocking without giving up interpersonal communication would be to call up information from other group members only at times when it does not inhibit one's own productivity. Such a possibility is provided by electronic brainstorming, a technique for generating ideas in groups developed by Nunamaker, Applegate, and Konsynski (1987). In a study by Gallupe, Bastianutti, and Cooper (1991) an electronic brainstorming system was introduced in real

---

5 Without note-taking, the availability of communication did not have a significant effect on brainstorming productivity.
and nominal groups and compared to the conventional brainstorming procedure. With electronic brainstorming each group member has a computer terminal connected to all other terminals. The brainstorming software allows individuals to enter their ideas as the ideas occur to them. Every time an individual enters an idea, a random set of the group’s ideas is presented on the individual’s screen. The individual can continue to see new random sets of ideas at will by pressing the appropriate key (p. 138).

Results showed that with electronic brainstorming the usual productivity loss of real groups compared to nominal groups did not occur. Interacting and non-interacting brainstorming groups using electronic brainstorming did not differ in the number of nonredundant ideas. However, both produced significantly more ideas than the interacting nonelectronic brainstorming groups that adopted the conventional verbal face-to-face brainstorming technique. The authors claim that their pattern of findings indicates that electronic brainstorming prevents production blocking in ‘interacting’ groups.

Two additional experiments by Gallupe et al. (1992), in which electronic and conventional group brainstorming were compared in groups of varying size, appeared to corroborate an interpretation of the effects of electronic brainstorming on group productivity in terms of production blocking. There were two-, four-, and six-person groups in the first and six- and 12-person groups in the second experiment. While the number of nonredundant ideas per person decreased with increasing group size for the conventional brainstorming technique there was no effect of group size on per person performance for the electronic brainstorming groups. The pattern of results for the number of high-quality ideas resembles the one for the number of nonredundant ideas, thus corroborating the high correlations of quantity and quality of ideas observed in our own studies (Diehl & Stroebe, 1987, 1991).

Unfortunately, Gallupe et al. fail to provide evidence of the extent to which members of electronic brainstorming sessions really attended to and made use of the random set of the group’s ideas presented on their screens. Although the interacting electronic brainstorming groups differ from nominal groups in numerous responses in the post-experimental questionnaire, this does not rule out the possibility that the interacting electronic brainstorming groups outperformed the face-to-face groups merely because it was easier to disregard written than verbal communications. If participants in electronic brainstorming sessions did not pay any attention to the ideas of other group members, it would not be surprising that their level of productivity paralleled that of nominal groups.

Group Composition: Increasing the Heterogeneity of Groups

Practitioners often criticize the fact that the groups used in studies on brainstorming are often very similar in their knowledge about the brainstorming
WOLFGANG STROEBE AND MICHAEL DIEHL

topic and argue that the stimulating impact of group brainstorming can only be achieved if groups are heterogeneous with regard to their knowledge. Since it is plausible that idea generation in groups should be less effective the more individuals overlap in their knowledge, a study was conducted in which the homogeneity or heterogeneity of the knowledge structure of the nominal or interacting group on the brainstorming topic was manipulated (Diehl, 1992).

Since environmental issues are very important to young people in Germany, the issue of the protection of our natural environment was used as a brainstorming topic. On the basis of an association test on the topic of environmental protection which subjects had to fill out several weeks before the brainstorming session, subjects were assigned to either eight homogeneous or eight heterogeneous mixed-sex four-person groups (two male, two female). A further set of eight mixed-sex groups was formed from male and female subjects who were randomly assigned to these groups. The formation of homogeneous versus heterogeneous groups was based on the similarity or dissimilarity of the categories which were dominant in these individuals (i.e. highly accessible in the association test). A second criterion for the assignment of subjects to groups was the minimization of performance differences between group members. Thus, all groups were of equal sex-composition and group members were matched with regard to their performance level on the association task.

Subjects were signed up in four-person groups for a study on problem solving in small groups. These groups, who had been matched to be either homogeneous or heterogeneous or had been randomly formed with regard to the content of their association concerning environmental protection, brainstormed either individually or in groups. In addition to assessing the number of nonredundant ideas produced by individuals and groups, an attempt was also made to assess the flexibility of performance in this study (i.e. the variability and diversity of idea production by an individual or a group). Flexibility is usually operationalized as the number of different categories needed to categorize the ideas. In order to assess the flexibility of idea production, a classification system had to be developed for categories of ideas about protection of the environment. Proceeding on the assumption that an elaborated idea suggests a specific action that could be seen as a path to achieve a specific sub-goal of the global goal 'environmental protection', a matrix was constructed that combined specific paths of actions with specific goals of actions. These path–goal matrices consisted of 10 goals (e.g. reduction of air pollution, saving of raw materials) and five paths by which these goals could be reached (influencing consumption, giving information). In order to be able to categorize less elaborated ideas, the categories 'path not specified' and 'goal not specified' were added to the classification system. The final classification system consisted of 66 categories.
A comparison of the number of ideas produced by nominal and interacting groups indicated the usual productivity loss for the real groups of homogeneous or random composition. However, no significant productivity loss occurred for real groups composed of members who were heterogeneous in their knowledge structure (Figure 9.2). The analysis of the flexibility of idea production resulted in the same pattern. The number of different categories of ideas used in a session can be analyzed at the individual and the group level. Figure 9.3 shows the average number of categories used by individuals in real and nominal groups. Again, the flexibility of idea production in real groups compared to nominal groups is much lower for homogeneous and randomly composed groups. In contrast, interacting groups composed of members who were heterogeneous with regard to their knowledge produced almost the same number of categories as nominal groups. The same pattern of results emerged for the flexibility of idea production at the group level, shown in Figure 9.4.

How can we explain this pattern? With an issue of such general interest as the protection of our environment, students are likely to have great similarity in their knowledge structures. Therefore, there should not have been much difference in knowledge structure between the homogeneously and the randomly formed groups. As a consequence, all members of homogeneous or randomly formed groups had gaps in the same areas (i.e. the same categories were less accessible to all group members) and thus subjects would not complement each other. Therefore, there were no processes that would
counteract the usual reduction in productivity due to blocking. In heterogeneous groups, on the other hand, the knowledge structures of different group members would complement each other. In these groups, a reduction in individual flexibility was prevented by an interpersonal stimulation of those
WHY GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS

categories that were low in accessibility in one person but high in accessibility in another. Thus, despite the inhibiting effects of interpersonal blocking demonstrated in our previous brainstorming studies, no productivity loss was observed for interacting groups composed of members who were heterogeneous in their knowledge structure, either for the number of ideas or for the number of different categories of ideas used.

This result gives rise to the hope that a heterogeneous group composition regarding the individuals' knowledge structures will lead to an interpersonal stimulation that makes more of the available knowledge accessible to an individual than a homogeneous group composition. Given a brainstorming technique that reduces or avoids interpersonal blocking (e.g. electronic brainstorming), heterogeneous real groups might even outperform heterogeneous nominal groups.

Conclusions

With traditional methods of group brainstorming, however, the best result we appear to be able to achieve is the avoidance of a productivity loss in interacting brainstorming groups. There appears to be a ceiling for the productivity of interacting groups at the level of the productivity of the nominal groups. Thus, there is not only no evidence so far to support Osborn's contention that individuals will be substantially more creative when working in groups rather than individually. On the contrary, all the evidence suggests that under normal circumstances individuals produce fewer ideas and fewer good ideas when working in groups rather than individually.

THE ILLUSION OF GROUP EFFECTIVITY

Despite the consistent failure of brainstorming groups to reach the level of productivity achieved by nominal groups, brainstorming is still widely used in business organizations and advertising agencies. Thus, academics, politicians, and the business community remain unshakable in their conviction that groups can stimulate creativity. We have called this the illusion of group effectivity (Diehl & Stroebe, 1991). In this last section we attempt to explain how this illusion can persist despite consistent evidence to the contrary.

We are all members of a variety of groups and, as such, are frequently involved in idea generation. Therefore, the illusion of group effectivity could not have persisted unless group members typically experienced that they had been more creative in groups then they would have been on their own. Indeed, subjects in our studies overwhelmingly reported that they thought that people produced more ideas when working in groups than when working individually. It did not make any difference whether subjects themselves had been in a group
session or had worked individually. Group members also felt facilitated by the presence of other group members. They were also likely to be more satisfied with their performance than were subjects who worked individually.

How can one explain the persistence of this illusion, despite the performance deterioration associated with idea generation in groups? One reason could be a baseline fallacy. In brainstorming sessions more ideas are produced by a four-person group than by an individual. Members of such groups could therefore correctly be under the impression that they achieved much more as group members than they could have achieved on their own. This baseline fallacy, which is reflected by such sayings as ‘Two heads are better than one’, led early group researchers such as Shaw (1932) to infer group superiority from the observation that more groups than individuals solved their problems.

A somewhat different interpretation assumes that after a session group members are unable to differentiate between the ideas they have had themselves and those that were suggested by other group members. Because individuals are usually motivated to view their own performance in a positive light (i.e. self-enhancement), the inability to differentiate between own and others’ ideas should result in a tendency among group members to attribute some of the ideas of other group members to themselves. This tendency could result in an overestimation of ideas that subjects claim to have reported and/or ideas that they recall having had at the session without actually verbalizing them. It would explain why people who brainstorm in groups rate their productivity higher than those who generate ideas in individual sessions. It could also be the reason people avoid dividing group products according to individual contributions (Mikula, 1980).

To test these hypotheses, we conducted an experiment which consisted of two parts (Stroebe, Diehl, & Abakoumkin, 1992). During a first session, subjects were asked to brainstorm either individually or in four-person groups. When subjects returned two weeks later for a second session, they had to identify their own ideas from a set consisting of their own ideas and those produced by the members of their real or nominal groups. Subjects were also asked to identify ideas that had occurred to them during the first session even though they had not verbalized them.

Findings after the first session replicated the typical productivity loss of interacting groups. The average number of nonredundant ideas produced by real groups was 84 and thus markedly lower than the 110 ideas produced by nominal groups. Nevertheless, group members were more satisfied with their performance than were subjects who participated in individual sessions. Both types of subjects thought that people produce more ideas in groups than individually. Under both conditions subjects stated that they had reported practically all the ideas that had occurred to them during the brainstorming session.

In the second session subjects were individually presented with the total set of nonredundant ideas produced by their nominal or real group, each idea
written on a card. Their task was to assign each idea to one of three categories: (1) suggested by me; (2) suggested by another group member but had also occurred to me; (3) suggested by another group member and had not occurred to me.

Table 9.3 presents the percentage of the total number of nonredundant ideas produced by a group that was assigned to each of the three categories. There were significant differences for all three categories. Compared with subjects in nominal groups, members of real groups assigned a significantly lower percentage of ideas to the categories 'Suggested by me' and 'Suggested by another group member and had not occurred to me' but a higher percentage to the category 'Suggested by another group member but had also occurred to me'. In fact, when the first two categories are combined into a category that implies the claim that an idea had occurred to a subject during the first session, subjects in real four-person groups claimed that 61% of the ideas had occurred to them at the first session, whereas those in nominal groups made such a claim for only 47%.

<table>
<thead>
<tr>
<th>Category</th>
<th>Nominal Group</th>
<th>Real Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suggested by me</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>Suggested by another group member but had also occurred to me</td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>Suggested by another group member and had not occurred to me</td>
<td>53</td>
<td>39</td>
</tr>
</tbody>
</table>

To test the assumption that members of real groups were less accurate than those of nominal groups in identifying the ideas they had suggested, a comparison was made on the hits and misses for ideas subjects identified as having been suggested by themselves (Table 9.4). The first row of Table 9.4 presents the percentage of ideas that subjects had actually suggested and identified as suggested (hits) and the percentage of ideas that subjects had actually suggested but did not identify accordingly (misses). The second row presents the percentage of ideas that subjects erroneously identified as not suggested by themselves (misses) and the percentage that they correctly identified as not suggested by themselves (hits). This analysis indicated a significant difference between real and nominal groups in terms of correct identifications of suggested ideas. Members of real groups identified a significantly lower percentage of ideas that they had actually suggested than members of nominal groups and also felt significantly less confident of the validity of their decisions.
Table 9.4 Percentages of ideas correctly and incorrectly identified by members of nominal and real groups as suggested by themselves or not suggested by themselves (reproduced by permission from Stroebe, Diehl & Abakoumkin, 1992)

<table>
<thead>
<tr>
<th>Actual Source of Idea</th>
<th>Identified as 'Suggested by Me'</th>
<th>Identified as 'Not Suggested by Me'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal</td>
<td>Real</td>
</tr>
<tr>
<td>Suggested by Subject</td>
<td>76*</td>
<td>57*</td>
</tr>
<tr>
<td>Not Suggested by Subject</td>
<td>20</td>
<td>18</td>
</tr>
</tbody>
</table>

* Hits (correct identifications).

Thus, results of this experiment are generally consistent with our hypotheses. There is clear evidence that members of real groups were not only much less accurate than subjects who brainstormed individually in differentiating their own ideas from those of their colleagues but also felt less confident in their decisions. This confusion resulted in a marked overestimation of the ideas group members claimed to have had during the first session but not of the ideas they claimed to have suggested.

We would like to argue that the first claim is much more relevant to our hypothesis than the second. First, there is reason to believe that members of real groups might have been reluctant to make such claims because they were affected by fairness norms which did not apply to subjects who brainstormed individually. When members of real groups claim to have suggested an idea they imply at the same time that other members did not suggest it. For members of nominal groups such claims are much less exclusive, because in this situation the same idea can be suggested by several subjects. More importantly, however, the illusion of group effectivity is based on assumptions about creativity rather than on claims about actual contributions. Thus, the fact that group members were under the illusion that more than half the ideas suggested by the group had also occurred to them at the time (i.e. that they were highly creative in the group setting) is sufficient to account for the illusion of group effectivity.

Conclusions

The findings of our last study suggest a solution to the puzzle of how people maintain their belief in the effectivity of group discussion as a means for stimulating individual creativity despite pervasive evidence to the contrary. The fact that people vastly overestimate the ideas that occur to them in group settings as well as the lack of baseline information from nominal groups goes some way towards explaining the persistence of the illusion of group effectivity.
GENERAL CONCLUSIONS

This chapter has presented evidence to support three assumptions. First, individuals who brainstorm alone produce a greater number of ideas and a greater number of good ideas than individuals who brainstorm in groups. Second, the productivity loss in brainstorming groups is mainly (though not exclusively) due to mutual production blocking, that is, the fact that only one member of a group can speak at any given time. Third, attempts at overcoming the impact of production blocking have been largely unsuccessful. The best we have been able to achieve is the avoidance of a productivity loss in interacting groups. Thus, there is no evidence to support Osborn's contention that individuals will be more creative in groups.

We have little doubt, however, that even though our findings suggest that group members should generate their ideas individually and use the group merely for purposes of evaluation, the evidence presented in this chapter will have little impact on reducing the popularity of group brainstorming in commercial and academic settings. People want to believe that groups are more effective because they enjoy group brainstorming much more than individual brainstorming. A further attraction of group brainstorming which also has little to do with the alleged freeing of creative potential in the group is that ideas produced by a group are more likely to be acceptable to the group members than those produced by some outside individuals. However, if people insist on generating ideas in groups rather than individually, we would recommend at least to keep the size of these groups small. With small groups one can still have the advantage of the group setting in terms of motivation and enjoyment without paying too high a price in terms of efficiency.

ACKNOWLEDGEMENTS

Some of the research reported in this chapter has been financially supported by the Deutsche Forschungsgemeinschaft. We are grateful to Steve Harkins, Chester Insko, Norbert Kerr, and Miles Hewstone for helpful suggestions on an earlier draft of this manuscript.

REFERENCES


WHY GROUPS ARE LESS EFFECTIVE THAN THEIR MEMBERS