A Process Analysis of Idea Generation and Failure

Hao-Chuan Wang and Carolyn Penstein Rosé (haochuan,cprose@cs.cmu.edu)
Language Technologies Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

This paper presents an experimental study and detailed process analysis investigating the temporal characteristics of idea generation and idea failure during brainstorming. In this investigation, we test the hypothesis that cognitive interference resulting from group interactions has a temporary effect and thus decays as the rate of unique idea generation decays. A further hypothesis is that the effects of cognitive stimulation from external stimuli become more evident over time as the effects of cognitive interference decay. Both of these hypotheses are confirmed by our process analysis.

Keywords: Cognitive stimulation; cognitive interference.

Introduction

The task of idea generation in brainstorming groups has been extensively studied through controlled experiments (Diehl & Stroebe, 1987) and simulation studies (Nijstad & Stroebe, 2006). In the study of group brainstorming for idea generation, empirical work has repeatedly revealed evidence of process losses, in which a group with idea sharing may not always perform better than a collection of non-interacting individuals whose contributions are simply pooled afterwards (i.e., nominal groups), both in terms of the quantity and quality of unique ideas (Hill, 1982; Diehl & Stroebe). A wide range of social and cognitive explanations for process losses have been proposed and evaluated, including social pressure (e.g., evaluation apprehension), social loafing (e.g., "free riding"), production blocking resulting from turn taking conventions (Diehl & Stroebe), and cognitive interference from exposure to similar ideas to the current idea generation focus (Nijstad & Stroebe).

One influential explanation for process losses in group idea generation grows out of the theory of cognitive stimulation and cognitive interference, which models idea generation by individuals as a memory retrieval process (Dugosh, Paulus, Roland & Yang, 2000; Nijstad & Stroebe, 2006). Idea failure related to cognitive interference occurs when one fails to retrieve an image from memory, retrieves an image but fails to generate an idea from it, repeatedly activates the same images, or generates ideas that have already been mentioned. When brainstorming in groups, ideas verbalized by peers may serve as external stimulation potentially leading to one of two extreme results, either process gains or process losses. In the first case, overhearing a peer's ideas may serve as cognitive stimulation for memory retrieval and idea generation when those ideas are significantly different from what an individual was capable of generating alone, while in the later case, this external stimulation becomes a source of cognitive interference leading to process losses. This may occur because the same domain knowledge continues to be activated, which facilitates the generation of similar ideas to those mentioned recently. In cases such as these, brainstorming individuals may believe they have exhausted the number of ideas they are able to generate (Nijstad & Stroebe).

The temporal properties of priming and cognitive stimulation have been investigated (Raaijmakers & Shiffrin, 1981), but a similar investigation of the temporal properties of cognitive interference have not been investigated. Rather, the effects of cognitive interference and related process losses have primarily been evaluated summatively. In order to investigate the temporal properties of cognitive interference and resulting process losses, we present in this paper an experimental study and detailed process analysis. In the remainder of the paper we discuss the theoretical model of idea generation that underlies this work, our experimental study, and findings from our process analysis.

Theoretical Model

Brainstorming may be modeled from a cognitive perspective as a two stage process in which a search is initiated in response to a challenge presented during the first stage, and then an idea is constructed during this search by means of inferences building upon prior knowledge (Nijstad & Stroebe, 2006). Psychological studies in spreading activation have demonstrated that it is more efficient to retrieve a concept from memory when *relevant* information is offered as a *prime* or cue (Anderson, 2005). Based on the cognitive theory of associative memory (Anderson; Brown & Paulus, 2002; Dugosh et al., 2000), idea generation can be viewed as a process of building on the retrieval of information encoded in the stimulated portions of a semantic network stored in one's long-term memory.

The process of generating a new idea is said to involve a constructive process analogous to that of self-explanation (Chi, De Leeuw, Chiu & Lavancher, 1994) in which bridging inferences link instances of known domain facts on the way towards generating ideas (Brown & Paulus, 2002). For example, someone may have access to the following two domain facts: (1) Debris flow refers to the mass movement of rocks and sedimentary materials in a fluid like manner. And, (2) There are many typhoons, or hurricanes, in Taiwan in the summer time. That person may then make the following two bridging inferences: (1) Heavy rain implies the presence of a massive amount of water. And, (2) The presence of a massive amount of water may lead to erosion or the movement of rocks in a fluid like manner. That person may then generate the following idea: "Typhoons may be a factor leading to the occurrence of a debris flow hazard." It is the middle stage, where bridging inferences are generated, that makes this process similar to that of self-explanation. And because of those bridging inferences we have hypothesized that the process of idea-generation may lead to learning (Wang et al., 2007).

The first hypothesis we investigate in this paper is that cognitive interference has a temporary effect, since the activation of knowledge from idea contributions is temporary. If this is true, we expect to see that the magnitude of important effects like process losses of group work might differ during different periods of time. In particular, across all conditions, we see that the rate at which ideas are contributed decays exponentially over time. If our first hypothesis is true, then we further hypothesize that there will be more cognitive interference from peers early in a group brainstorming session because of the higher frequency of contributed ideas. It should decay over time since the interference from earlier ideas will decay over time, and there will be fewer ideas contributed later to cause new interference later. Thus, over time we expect the effect of the external stimuli to begin to overtake that of the cognitive interference.

Method

Experimental Design

In order to test our hypotheses, we conducted an experiment in which participants engaged in a brainstorming task. The selected task, namely the Debris Flow Hazard (DFH) task, has been designed by science educators to engage students in scientific inquiry in the area of Earth sciences (Chang & Tsai, 2005). The learning objective of this task is to make concepts related to geology, agriculture, and urban development concrete for students as they grapple with the manner in which these very different types of factors interact in real world scenarios. We manipulated whether brainstorming took place as an individual or pair activity and whether feedback designed to provide cognitive stimulation was offered or not, both as between subjects factors. Thus, the experiment was a 2 (individual brainstorming vs. pair brainstorming) X 2 (no system-generated feedback vs. system-generated feedback) factorial design resulting in four experimental conditions, which are referred to in the remainder of the paper as IN (Individual-No feedback), IF (Individual-Feedback), PN (Pair-No feedback), and PF (Pair-Feedback).

Experimental Infrastructure

In order to implement the four conditions in a way that maintains maximal consistency across conditions, we built our experimental infrastructure on top of a well known instant messaging (IM) service over the Internet, MSN messenger. Due to the popularity of this IM service with the target user population, using an MSN-based client also lessens potential concerns of software difficulty or novelty effects. More importantly, because strict turn taking is not enforced in this environment, there is a potential for supporting a group interaction that avoids process losses.

A brainstorming feedback agent called VIBRANT (Wang, Rosé, Cui, Chang, Huang, Li, 2007) was used to provide prompts in response to conversational behavior in the two feedback-enabled conditions. In order to be adapted to a specific task, VIBRANT must be provided with an idea hierarchy at multiple levels of abstraction. In our domain idea hierarchy, the top node representing the entire DFH task is first broken down into 5 general topic areas including geology (e.g., shale rock area), agriculture (e.g., having shallow-rooted economic plants which cannot solidify the soil mass as much as original forests), influences caused by other natural phenomena (e.g., typhoons which break the hydraulic balance), urban development (e.g., building houses on a potentially dangerous slope), and social factors (e.g., improper environmental policies). Each subtopic is divided into specific idea nodes. A total of 19 specific idea nodes are included. Feedback messages are attached to the nodes of the idea hierarchy both at the general topic level and the specific idea level. Similarly, at the specific idea level, prototype expressions of the related idea collected in previous studies involving the DFH task are attached to idea nodes. In this way, participant conversational contributions can be matched to nodes in the hierarchy by matching the text of their contribution to the associated prototype texts using a simple semantic similarity measure.

The feedback provided by VIBRANT consists of two parts. The initial portion, which we refer to as the comment, acknowledges the idea that was matched, and how it fits or doesn't fit into the hierarchy. The second portion, which we refer to as the tutorial, offers a hint for thinking about a new contribution. Feedback messages are constructed by concatenating a selected comment with a selected tutorial. For example, if the participant has contributed the idea "deforestation", the system will acknowledge this with the following comment, "Good, you seem familiar with the effects of excessive urban development." A next focus for brainstorming that follows coherently from this would be more discussion related to urban development, for example "Can you think of a farming practice motivated by economic concerns that may increase the risk of a debris flow hazard?" VIBRANT never offers participants specific example ideas. Instead, the hints offered by VIBRANT are more similar to the "category label" stimuli (such as "improve parking" for the task of "how can your university be improved?") demonstrated to enhance idea production in previous studies of group and individual brainstorming behavior (Dugosh et al., 2000; Nijstad & Stroebe, 2006). VIBRANT's built in strategy for selecting a next focus was designed to balance breadth and depth across the idea hierarchy while maintaining coherence. This design is motivated by prior findings that brainstorming is more efficient when successive ideas are clustered so that semantically related ideas are contributed in close proximity, and transitions between general idea categories are relatively rare (Nijstad & Stroebe).

For the IF condition, VIBRANT offered feedback in response to each contribution of the participant. For the PF condition, in order to avoid having the feedback from the agent prohibit interaction between participants, the system collected and evaluated the two participants' contributions during a fixed period of time, and then gave feedback based on the accumulated text rather than contributing a response to every contribution. In this study, the time delay parameter was set to 30 seconds, which was observed during a pilot experiment to allow participants enough time to interact with one another. No feedback from the system was offered to participants in the two no feedback conditions. Thus, in contrast to the two feedback-enabled conditions just described, for the IN and PN conditions, a simple computer agent did nothing but simply recorded participants' contributions. In the IN condition, where participants worked alone, they were simply instructed to use the IM program as a text input buffer.

Participants

The study was conducted in a computer classroom of a public high school located in central Taiwan. Four sessions were scheduled on the same day, two in the morning and two in the afternoon. In each session, the classroom accommodated at most 16 participants. Every participant worked at a computer assigned to him or her. Participating students were allowed to choose the session they attended, and were randomly assigned to experimental conditions within that session. For experimental conditions PN and PF, participants were paired into dyads randomly. The version of the MSN based software used as our experimental infrastructure was configured so that in the pair conditions, the other participant assigned to the same dyad appeared in the "buddy list" of that participant. Additionally, in the feedback-enabled conditions, the computer agent that provides feedback also appeared in the "buddy list". Thus, when the participants launched the application during the experimental manipulation, their MSN based client was configured to support a conversation between all of the relevant parties. Altogether, there were 7 participants in the IN condition, 7 participants in IF, 14 participants in PN (i.e., 7 pairs), and 14 participants in PF (i.e., 7 pairs). During the study, all participants were blind to the experimental design, and unaware of the existence of other conditions.

Experimental Procedure

The experimental procedure can be divided into five phases, namely (1) background readings, (2) pretest, (3) brainstorming 1, (4) brainstorming 2, and (5) the post test. The experimental manipulation took place during phase (3), which is the first brainstorming phase. All other phases were conducted the same way for all participants, and in all phases other than the experimental manipulation phase, all participants worked alone. The pre/post tests were administered in order to investigate the relationship between idea generation and learning, which is addressed in a separate paper (Wang et al., 2007). Thus, the learning results will not be discussed in the remainder of the paper. The purpose of the second brainstorming phase is to test whether the experimental manipulation that takes place in phase 3 has a lasting effect on brainstorming behavior beyond the duration of the manipulation that can be detected within a new brainstorming task. Again, the transfer task is part of the learning assessment, which is not a focus of this paper. We strictly controlled for time in all phases. Here we describe the whole procedure in detail.

Phase 1. Background Reading (10 minutes)

During phase 1, the participants read a 3 page packet of background reading materials on the climate, geology, and development of Taiwan as well as some information about natural disasters but no specific information about debris flow hazards. This packet was compiled by domain experts working in a science education center at National Taiwan Normal University. The readings were designed to offer participants a wide range of background material related to the topics contained within the idea hierarchy discussed above, however it did not contain the direct answers to any questions on the test nor did it directly express the ideas participants were required to contribute in the brainstorming task. The reading material itself does not explicitly introduce the factors underlying DFH occurrences. The purpose of the reading materials was to prepare participants for the brainstorming task. Participants in all conditions were instructed to read the material for 10 minutes, and to learn as much as possible from the material.

Phase 2. Pre-brainstorming Test (15 minutes)

In phase 2, participants took an on-line pretest assessing their conceptual knowledge and reasoning about debris flow hazards.

Phase 3. Brainstorming Activity 1 (30 minutes)

After the pretest, the participants participated in the first brainstorming phase, which is where the experimental manipulation took place. Participants were instructed to launch the MSN program and to start working on the DFH brainstorming task. Specific instructions for the task appeared as the first prompt in the MSN messenger window. Participants were given a scenario about a specific debris flow hazard and then asked to generate as many thoughts as possible in answer to the question, "*what are the possible factors that may cause a debris flow hazard to happen?*" During this activity, participants were invited to use the reading materials from Phase 1 as a resource. The duration of the brainstorming session was limited to 30 minutes.

Phase 4. Brainstorming Activity 2 (10 minutes)

Upon the completion of the brainstorming task, participants regardless of experimental condition were then instructed to do individual brainstorming on a second brainstorming task. In this idea generation task, participants were requested to offer preventive solutions for DFH. The prompt for this solution-finding brainstorming activity was "*what facilities or solutions may prevent a debris flow hazard from happening?*" No system-generated feedback, reading material, or peer interaction was provided when doing this transfer task.

Phase 5. Post-brainstorming Test (15 minutes)

Finally, participants took an on-line post-test identical to the one used as a pretest.

Verbal Protocol Analysis

Logs of all IM behavior in all conditions were saved for analysis. Altogether we collected 28 logs, 7 in each condition. Note that in the pairs conditions, there is only one log per pair rather than one log per participant.

To derive appropriate quantitative measures of idea generation for analyses, including task performance (number of unique ideas in the main idea generation task) and transfer performance (number of unique ideas the solution-finding transfer task), data collected in the main brainstorming phase (i.e., phase 3) and the transfer task phase (i.e., phase 4) have been coded and inventoried.

For the main idea generation task, participant IM logs were first segmented into idea units, since during IM conversations, participants may contribute more than one idea per turn. The inter-rater reliability between two independent coders over 10% of the data for segmentation was satisfactory (Kappa= .7). Each unit contribution was then classified into one of the 19 domain concepts in the aforementioned idea hierarchy. If an idea would count as an acceptable idea for the transfer task, as defined in the next paragraph, it received the label of "off-task idea", since it was a valid idea, but not for the current task. Otherwise, if there was no feasible label for a particular contribution, the label of "junky ideas" was given, since the idea was not appropriate for either of the two related tasks. The inter-rater reliability for the concept coding over 10% of the data was sufficiently high (Kappa=.84). From this coding we computed how many unique high quality ideas were generated within a session. In the pairs conditions, an idea only counted as a unique idea the first time it was mentioned by either participant in that session.

For the second brainstorming task, participants' responses were coded according to a similar coding scheme developed by domain experts for this task based on the distribution of observed participant responses in prior studies. The categories in that coding scheme represent what the experts consider the 15 most common "valuable" ideas from this distribution. The inter-rater reliability of this coding of two independent coders over 10% of the data was Kappa=.74, which is satisfactory.

To achieve a fair comparison between individual idea generation and group idea generation, it is standard to form "nominal pairs" by randomly selecting people from experimental conditions in which participants worked individually, and then pool their ideas collectively for a comparison with real pairs. As in the pairs conditions, each of the 19 ideas is only counted at most once per session. Thus, in the remainder of the paper, when we compare individuals to pairs, what we are really comparing is productivity in "nominal pairs" with productivity in real pairs.

Results

Exponential Decay of Idea Productivity Over Time

In all conditions, we observed an exponential decay of idea productivity over time, which we argue below affects the magnitude of cognitive interference in the pairs conditions over time.

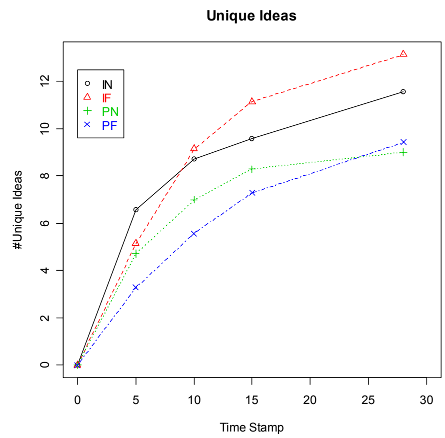

Figure 1: Accumulation of unique ideas per condition.

In particular, roughly half of the unique ideas contributed during the 30 minute brainstorming session were generated within the first 5 minutes, as depicted in Figure 1.

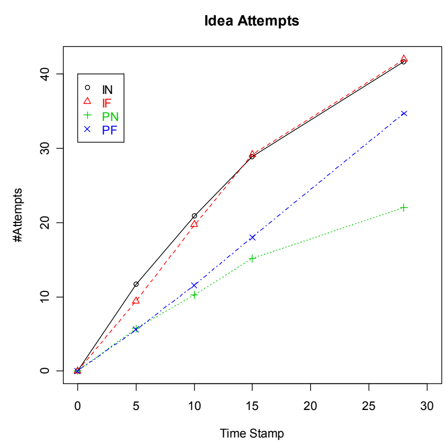

Figure 2: Accumulation of idea attempts per condition.

However, as displayed in Figure 2, the decay of idea attempts over time was much more gradual. By the end of the 30 minute brainstorming session, idea attempts have only begun to level off in the pairs with no feedback condition. In the pairs condition with feedback, idea attempts continue at a constant rate. And in the two individual conditions, a slow down in the rate of idea contributions only begins to become apparent in the second half of the brainstorming session. Thus, over time participants in all conditions experience more and more failed attempts to generate unique ideas. In the remainder of our analysis we will examine the process of idea failure, how cognitive interference affects that process, and how system generated feedback mitigates that effect. Note that this consistent pattern of decay across conditions is not what we refer to as process losses. Process losses are evident in the comparison between real pairs and nominal pairs, examined in the Cognitive Interference section below.

High Level View of the Process of Idea Failure

In order to examine the process of idea failure, we classified idea attempts into four categories based on their classification from the verbal protocol analysis discussed earlier in the paper. Above we introduced these four categories as unique ideas, repetitions, off-task ideas, and junky ideas. In Figures 3 and 4 it is possible to observe how these types of contributions accumulate over time in the pairs with no feedback (PN) and individuals with no feedback (IN) conditions respectively.

First, let us look qualitatively at the patterns that emerge from these graphs. It is clear that the proportion of successful idea attempts is highest in the first 5 minutes. The gap between the line representing number of attempts and the line representing unique ideas widens over time.

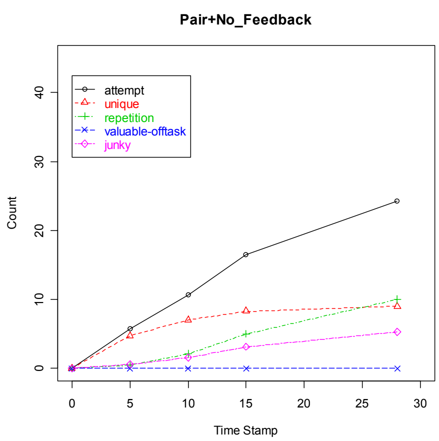

Figure 3: Accumulation of instances of idea categories over time in the PN condition.

During the first five minutes, repetitions, off-task, and junky ideas are almost non-existent in the pairs condition. Repetitions occur almost half the time during this time interval in the individual condition, although junky and off-task ideas are almost non-existent. Moving on from the first 5 minute interval, the picture begins to change in both the IN and PN conditions. In the PN condition, the proportion of repetitions and junky ideas begins to increase at the 5 minute mark, with repetitions increasing at a higher rate. By the end of the brainstorming session, unique ideas and repetitions are in equal proportion, with junky and off-task ideas remaining a smaller proportion. In the IN condition the picture is similar. In both cases, repetitions are a more common, and possibly earlier, form of idea failure, with junky ideas being less frequent and possibly later.

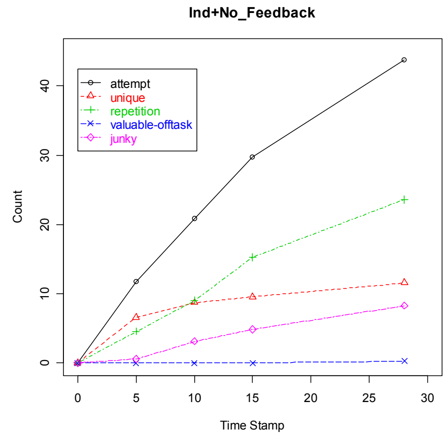

Figure 4: Accumulation of instances of idea categories over time in the IN condition.

This qualitative image is confirmed in our statistical analysis. By using the elapsed time since the start of the session for each idea attempt as the dependent variable and idea type as a four level independent variable (with unique, repetition, off-task, and junky as the four levels) in an ANOVA analysis, it is found that the difference in elapsed time across idea types was statistically significant ($F(3, 655)=52.85$, $p<.001$). From a post-hoc analysis using a Tukey-test, it is found that the unique ideas occurred significantly earlier than the other three types of idea contributions (effect size Cohen's $d=0.9$).

Cognitive Interference

Process losses are hypothesized to result from cognitive interference experienced because of exposure to a partner's similar idea contributions. Because the frequency of idea contributions is highest in the first five minutes, we hypothesize that cognitive interference has the greatest effect during that time. If we consider only brainstorming behavior during that period, we already see significant evidence of process losses by the five minute mark when we compare unique ideas generated by "nominal pairs" with unique ideas generated by real pairs, $F(1, 24)=12.22$, $p<.005$. The effect size of the advantage for "nominal pairs" is 1 standard deviation. If we then look just at brainstorming behavior after the five minute mark, we still see significant process losses due to working in pairs, however the effect size is only .61, $F(1, 24)=4.61$, $p<.05$. This is consistent with earlier conjectures that process losses may be worse during the early stage of a brainstorming session, although this was not investigated directly in prior work (Diehl & Stroebe, 1991; Nijstad & Stroebe, 2006). One possible explanation is that the cognitive interference experienced as a result of the exposure to similar ideas is only a temporary effect. Thus, cognitive interference decays over time as the frequency of unique idea contributions decays. We see evidence of this in the way the effect of feedback changes over time, as examined in the next section.

The Effect of Feedback

Based on summative measures, we do not see a significant benefit of feedback on idea productivity. As mentioned, there is a main negative effect of working in pairs on idea productivity, with no interaction with feedback. If we take a process level view, however, we see a different picture. We saw above that it is reasonable to conjecture that cognitive interference is highest during the first five minutes of the brainstorming session where the frequency of idea contributions is highest. If the inhibitory effect of exposure to a similar idea is only temporary, then we would expect the level of cognitive interference to decay over time with the frequency of new idea contributions, and the pattern we saw was consistent with this. It is possible that during the time when cognitive interference is the strongest, the effect of the feedback, which may be smaller in comparison, may not be visible. Thus, we might expect the positive effect of feedback to become more noticeable as the negative effect of cognitive interference decays. If the positive effect of the feedback is not evident during the first five minutes where cognitive interference is the strongest, then the individual conditions would have the opportunity to gain an enduring advantage during that time that might not be compensated for later even if cognitive interference could be completely removed as a result of the feedback and pairs could experience equal success at idea generation as individuals. If that was the case, we would expect to see process losses evidenced by summative measures computed over the entire brainstorming session, but possibly not by summative measures computed over the remaining segment of time following the initial five minutes.

In line with this, we see in Figure 1, that the gap between individuals with no feedback (IN) and pairs with feedback (PF) narrows slightly over time past the five minute mark rather than widens. The difference between IN and PF during the first five minutes is significant, $t(12) = 3.49$, $p<.005$, effect size Cohen's $d = 1.5$. After the first five minutes to the end of the session, PF actually slightly outperformed IN during this period of time, although the difference is not significant, $t(12) = -1.55$, effect size Cohen's $d = 0.65$. This confirms that the feedback was effective in mitigating process losses except during the first five minutes where cognitive interference was the strongest.

Conclusions and Current Directions

In this paper, we have presented an experimental study and detailed process analysis investigating the temporal characteristics of idea generation and idea failure during brainstorming. While summative measures of idea generation productivity show strong evidence of process losses in brainstorming pairs even with the support of feedback, a detailed process analysis demonstrates that after the first five minutes of the brainstorming session, the that process loss is removed so that productivity over the remainder of the session for pairs with the support of feedback is nearly identical to that of individuals without feedback. However, since feedback also increased the productivity of individuals, we do see process losses in pairs with feedback in comparison with individuals supported with the same feedback. The results from our process analysis confirm conjectures put forward in earlier work that process losses may mainly occur early in brainstorming sessions. These results also explain why earlier studies found process losses with interacting groups in comparison with "nominal groups" even when the interacting groups were given more time (Diehl & Stroebe, 1991).

This work was funded in part by the National Science Council (NSC), Republic of China grant number NSC 94-2511-S-003-028 and National Science Foundation grant number SBE0354420.

References

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6 ed.). New York: Worth.

Brown, V. R., & Paulus, P. B. (2002). Making group brainstorming more effective: recommendations from an associative memory perspective. *Current Directions in Psychological Science, 11*(6), 208-212.

Chang C., & Tsai, C. (2005). The interplay between different forms of CAI and students' preferences of learning environment in the secondary science class. *Science Education, 89*, 707-724.

Chi, M. H., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-488.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. *Journal of Personality and Social Psychology*. *53*(3), 497-509.

Diehl, M. & Stroebe, W. (1991). Productivity loss in idea-generating groups: tracking down the blocking effect. *Journal of Personality and Social Psychology*, *61*(3), 392-403.

Dugosh, K. L., Paulus, P. B., Roland, E. J., & Yang, H. (2000). Cognitive stimulation in brainstorming. *Journal of Personality and Social Psychology. 79*(5), 722-735.

Hill, G. W. (1982). Group versus individual performance: are *N*+1 heads better than one? *Psychological Bulletin*, *91*(3), 517-539.

Nijstad, B. A., Stroebe, W., & Lodewijkx. F. M. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of Experimental Social Psychology*, *38*, 535-544.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: a cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *10*(3), 186-213.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93.134.

Wang, H. C., Rosé, C.P., Cui, Y., Chang, C. Y, Huang, C. C., Li, T. Y. (2007). Thinking Hard Together: The Long and Short of Collaborative Idea Generation for Scientific Inquiry, *Proceedings of CSCL 2007*.