Idea Generation, Creativity, and Incentives

Olivier Toubia

July 5th, 2005

Olivier Toubia is an assistant professor of marketing, Columbia Business School, Uris 522, 3022 Broadway, New York , NY 10027, ot2107@columbia.edu.

I am indebted to John Hauser for his generous advice and encouragement. This work has also benefited from the suggestions of Duncan Simester, Shane Frederick, Dražen Prelec, Dan Ariely, Birger Wernerfelt, Brian Gibbs, Ray Weaver, Kristina Shampagn'er and Oded Netzer. I would also like to thank John Hauser and Dan Ariely for funding experiment 1, Rosa Blackwood for her generous help with the logistics of this experiment, as well as Robert Klein and his team at Applied Marketing Science, Inc. for their cooperation and suggestions.

Idea Generation, Creativity, and Incentives

Abstract

Idea generation (ideation) is critical to the design and marketing of new products, to marketing strategy, and to the creation of effective advertising copy. However, there has been relatively little formal research on the underlying incentives with which to encourage participants to focus their energies on relevant and novel ideas. Several problems have been identified with traditional ideation methods. For example, participants often free ride on other participants' efforts because rewards are typically based on the group-level output of ideation sessions.

This paper examines whether carefully tailored ideation incentives can improve creative output. I begin by studying the influence of incentives on idea generation using a formal model of the ideation process. This model illustrates the effect of rewarding participants for their impact on the group, and identifies a parameter that mediates this effect. I then develop a practical, web-based, asynchronous "ideation game," which allows the implementation and test of various incentive schemes. Using this system, I run two experiments, which demonstrate that incentives do have the capability to improve idea generation, confirm the predictions from the theoretical analysis, and provide additional insight on the mechanisms of ideation.

Keywords: Idea Generation, New Product Research, Product Development, Marketing Research, Agency Theory, Experimental Economics, Game Theory.

1. Introduction

Idea generation (ideation) is critical to the design and marketing of new products, to marketing strategy, and to the creation of effective advertising copy. In new product development, for example, idea generation is a key component of the front end of the process, often called the "fuzzy front end" and recognized as one of the highest leverage points for a firm (Dahan and Hauser 2001).

The best known idea generation methods have evolved from "brainstorming," developed by Osborn in the 1950's (Osborn 1957). However, dozens of studies have demonstrated that groups generating ideas using traditional brainstorming are less effective than individuals working alone (see Diehl and Stroebe 1987 or Lamm and Trommsdorff 1973 for a review). Three main reasons for this robust finding have been identified as production blocking, fear of evaluation, and free riding. The first two (which will be described in more details in Section 3) have been shown to be addressed by Electronic Brainstorming (Nunamaker et al. 1987, Gallupe et al. 1991, Gallupe et al. 1992, Dennis and Valacich 1993, Valacich et al. 1994). However free riding remains an issue. In particular, participants free ride on each other's creative effort because the output of idea generation sessions is typically considered at the group level, and participants are not rewarded for their individual contributions (Williams et al. 1981, Kerr and Bruun 1983, Harkins and Petty 1982). This suggests that idea generation could be improved by providing appropriate incentives to the participants.[1] Surprisingly, agency theory appears to have paid little attention to the influence of incentives on agents' *creative* output.[2] On the other hand, classic research

---

[1] Note that free riding is likely to be magnified as firms seek input from customers (who typically do not to have a strong interest in the firm's success) at an increasingly early stage of the new product development process (von Hippel and Katz 2002).

[2] Agency theory has been widely applied to other marketing topics such as channel coordination (e.g., Raju and Zhang 2005, Desiraju 2004, Jeuland and Shugan 1988, Moorthy 1987).

in social psychology suggests that incentives might actually have a negative effect on ideation. For example, the Hull-Spence theory (Spence 1956) predicts an enhancing effect of rewards on performance in simple tasks but a detrimental effect in complex tasks. The reason is that rewards increase indiscriminately all response tendencies. If complex tasks are defined as tasks in which there is a predisposition to make more errors than correct responses, then this predisposition is amplified by rewards. Similarly, the social facilitation paradigm (Zajonc 1965) suggests that, insofar as incentives are a source of arousal, they should enhance the emission of dominant, well-learned responses, but inhibit new responses, leading people to perform well only at tasks with which they are familiar. McCullers (1978) points out that incentives enhance performance when it relies on making "simple, routine, unchanging responses," (p. 14) but that the role of incentives is far less clear in situations that depend heavily on flexibility, conceptual and perceptual openness, or creativity. McGraw (1978) identifies two conditions under which incentives will have a detrimental effect on performance: "first, when the task is interesting enough for subjects that the offer of incentives is a superfluous source of motivation; second, when the solution to the task is open-ended enough that the steps leading to a solution are not immediately obvious." (p. 34)

In this paper I examine whether carefully tailored ideation incentives can improve creative output. In Section 2, I study the influence of incentives on idea generation using a formal model of the ideation process. In Section 3, as a step towards implementing and testing the conclusions from Section 2, I present a practical, web-based, asynchronous "ideation game," which allows the implementation and test of various incentive schemes. In Section 4, I describe two experiments conducted using this system. These experiments demonstrate that incentives do have the capability to improve idea generation, confirm the predictions from the theoretical analysis, and provide additional insight on the mechanisms of ideation. Section 5 mentions some managerial applications and Section 6 concludes.

2. Theoretical Analysis

The goal of this section is to develop a theoretical framework allowing the study of idea generation incentives. One characteristic of ideas is that they are usually not independent, but rather build on each other to form streams of ideas. This is analogous to academic research, in which papers build on each other, cite each other, and constitute streams of research. Let us define the contribution of a stream of ideas as the moderator's valuation for these ideas (for simplicity I shall not distinguish between the moderator of the idea generation session and his or her client who uses and values the ideas). The expected contribution of a stream of ideas can be viewed as a function of the number of ideas in that stream. Let us assume that the expected contribution of a stream of n ideas is a non-decreasing, concave function of n, such that each new idea has a non-negative expected marginal contribution and such that there are diminishing returns to new ideas within a stream. Note that the analysis would generalize to the case of increasing, or inverted U-shaped, returns. (This might characterize situations in which the first few ideas in the stream build some foundations, allowing subsequent ideas to carry the greatest part of the contribution.)[3]

If several streams of ideas have been established in an idea generation session, a participant has a choice between contributing to one of the existing streams, and starting a new stream. If he or she decides to contribute to an existing stream, he or she might have to choose between contributing to a longer, more mature stream, and a newer, emerging one. For ease of exposition,

[3] The three judges in experiment 1 were asked to plot the evolution of the contribution of a tree of ideas as a function of the number of ideas in that tree. One judge thought that there were diminishing returns, while the other two indicated diminishing returns only after the first couple of ideas.

let us assume that each stream of ideas reflects a different approach to the topic of the session.[4] Then the problem faced by the participant can be viewed equivalently as that of deciding which approach to follow in his or her search for new ideas.

I assume that some approaches can be more fruitful than others, and that the participant does not know a priori the value of each approach but rather learns it through experience. He or she then has to fulfill two potentially contradictory goals: increasing short-term contribution by following an approach that has already proven fruitful (exploitation), versus improving long-term contribution by investigating an approach on which little experience has been accumulated so far (exploration). This problem is typical in decision making and is well captured by a representation known as the multi-armed bandit model (Bellman 1961)[5]. For simplicity, let us first restrict ourselves to a two-armed bandit, representing a choice between two approaches *A* and *B*. Let us assume that by spending one unit of time thinking about the topic using approach *A* (respectively *B*), a participant generates a relevant idea on the topic with unknown probability p_A (respectively p_B). Participants hold some beliefs on these probabilities, which are updated after each trial.[6]

For example, let us consider a participant generating ideas on "How to improve the impact of the UN Security Council?" (which was the topic used in Experiment 1). Let us assume that the participant contemplates two approaches to the problem: a legal approach which he or she believes to be somewhat promising, and a monetary approach (devising an incentive system that forces countries to respect the decision of the council), which appears more uncertain but potentially more rewarding. The participant has to decide between an approach with a moder-

---

[4] This assumption is not crucial. We just need to assume that each stream of ideas has an intrinsic characteristic that influences the likelihood that the participant will be able to contribute to it, and that the participant learns about this characteristic by trying to contribute to the stream.
[5] Bandit is a reference to gambling machines (slot machines) that are often called one-armed bandits. In a multi-armed bandit, the machine has multiple lever (arms) and the gambler can choose to pull any one of them at a time.
[6] Starting a new stream of ideas can be viewed as a special case in which a participant tries an approach that has not been successfully used yet.

ately high expected short-term return and a non risky long term potential, and an approach with a lower expected short term return that is potentially more promising in the long term.

The traditional tension between exploration and exploitation is completed further by the assumption that ideas in a stream have a diminishing marginal contribution. Let us complement the classical multi-armed bandit model by assuming that when N trials of an approach have been made, a proportion $\hat{p}$ of which were successful, then the next idea found using this approach has an expected marginal contribution of $(\hat{p}.N+1)^{\alpha}-(\hat{p}.N)^{\alpha}$, with $0 \leq \alpha \leq 1$.

Some topics are more narrowly defined than others. Mathematical problems or puzzles are an extreme example in which the definition of the problem is such that once a solution has been proposed, there is usually little to build upon. Such topics would be represented by a low value of α. On the other hand, some problems considered in idea generation are more "open-ended" (trying to improve a product or service) and are such that consecutive ideas refine and improve each other. Such situations correspond to higher values of α.[7]

To analyze the tension in ideation, I first formally characterize different strategies in idea generation. Next I consider the simple incentive scheme consisting of rewarding participants based on their individual contributions, and show that although it leads to an optimal balance between the different strategies when α is small enough, it does not when α is large. For large α, this suggests the need for incentives that interest participants in the future of the group's output. Although such incentives allow aligning the objectives of the participants with that of the moderator of the session, they might lead to free riding. Avoiding free riding requires further tuning

---

[7] The parameter α could also depend on the preferences of the user(s) of the ideas. For example, the organizer of the session might value many unrelated, rough ideas (low α), or, alternatively, value ideas that lead to a deeper understanding of certain concepts (high α).

of the rewards and can be addressed by rewarding participants more precisely for the impact of their contributions.

Strategies in idea generation

Let us consider a two-period model, in which a single participant searches for ideas, and study the conditions under which it is optimal to follow the approach that has the higher versus the lower prior expected probability of success. (For simplicity, I shall refer to the former as the "better-looking approach" and to the latter as the "worse-looking approach.") I assume that the prior beliefs on p_A and p_B at the beginning of period 1 are independent and such that $p_A \sim Beta(n_{AS}, n_{AF})$ and $p_B \sim Beta(n_{BS}, n_{BF})$, with n_{AS}, n_{AF}, n_{BS}, $n_{BF} \geq 1$.[8] The corresponding expected probabilities of success are $\hat{p}_A = E(p_A) = n_{AS}/N_A$; $\hat{p}_B = E(p_B) = n_{BS}/N_B$, where $N_A = n_{AS} + n_{AF}$ and $N_B = n_{BS} + n_{BF}$. The variance of the beliefs is inversely proportional to N_A and N_B.

Two effects influence this model. First, if the uncertainty on the worse-looking approach is large enough compared to the uncertainty on the better-looking approach, it might be optimal to explore the worse-looking approach in period 1 although it implies a short-term loss. In particular, if the expected probability of success of the uncertain, worse-looking approach is close enough to that of the more certain, better-looking approach, a success with the worse-looking approach in period 1 might lead to updated beliefs that will make this approach much more appealing than the currently better-looking one.[9] Second, if the number of trials (and successes) on

---

[8] Because beta priors are conjugates for binomial likelihoods (Gelman et al. 1995), the posterior beliefs at the end of period 1 (used as priors at the beginning of period 2) are beta distributed as well.

[9] For example, let us assume that the better-looking approach at the beginning of period 1, *B*, has a very stable expected probability of success which is around 0.61. By stable we mean that it will be similar at the beginning of period 2 whether a success or failure is observed for this approach in period 1 (N_B is very large). Let us assume that the worse-looking approach, *A*, has an expected probability of success of 0.60 at the beginning of period 1. However a success with this approach would increase the posterior probability to 0.67 while a failure would decrease this posterior probability to 0.50 (These probabilities would result from N_A=5, n_{AS}=3, n_{AF}=2.) Then, playing *B* in period 1 would lead to a total expected contribution of 0.61+0.61=1.22 (ignoring discounting for simplicity). Playing *A* in

the better-looking approach gets large, this approach becomes over-exploited (due to the decreasing marginal returns) and it becomes more attractive to "diversify" and choose the other approach: although the probability of success is lower, the contribution in case of success is higher.[10]

Propositions 1a, 1b and 1c in the technical appendix formally identify the conditions under which exploration, exploitation, or diversification should be used. In particular, these propositions show that exploration is never optimal when α is small enough: at the same time as the participant would learn more about p_A, he or she would also heavily decrease the attractiveness of this approach. On the other hand, when α is large enough and if A is uncertain enough compared to B, the exploration of A might be optimal in period 1. Exploration is characterized by a short-term loss (smaller expected contribution from period 1) incurred in order to increase long-term contribution (larger expected contribution from the two periods).

Rewarding participants for their individual contributions

Consider the incentive scheme that rewards participants based on their individual contributions. In practice, this could be operationalized by having an external judge rate each idea sequentially, and rewarding each participant based on the ratings of his or her ideas. This scheme is probably among the first ones that come to mind, and addresses the free-riding issue mentioned in the introduction.

Let us introduce a second participant into the model. Assume that, at each period, both players simultaneously choose one of the two approaches. At the end of period 1, the outcomes of both players' searches are observed by both players, and players update their beliefs on p_A and

period 1 would lead to an expected total contribution of 0.60+0.60*0.67+0.40*0.61=1.25 (play A in period 2 iff it was successful in period 1), which is higher.

[10] The conflict between exploration and exploitation is typical in the multi-armed bandit literature. However diversification typically does not arise because of the assumption of constant returns.

p_B. Each player's payoff is proportional to his or her individual contribution, i.e., to the sum of the contributions of his or her ideas from the two periods. The contribution of the group is equal to the sum of the contributions of each player. Note that α applies equally for both players, such that if a player finds an idea using a certain approach, the marginal return on this approach is equally decreased for both participants.[11]

Let us consider the subgame-perfect equilibrium (SPE) of the game played by the participants, as well as the socially optimal strategy, i.e., the strategy that maximizes the total expected contribution of the group.

Low-α case (rapidly decreasing returns)

When α is low, the socially optimal strategy is always to choose the approach with the highest short-term expected contribution. Consequently, actions that are optimal for self-interested participants trying to maximize their own contributions are also optimal for the group. This is captured by the following proposition (see the technical appendix for the proofs to Propositions 2 to 4):

Proposition 2: *For all N_A, N_B, there exists α^* such that $0 < \alpha < \alpha^* \Rightarrow$ for all $\hat{p}_A$, $\hat{p}_B$ such that $\hat{p}_A < \hat{p}_B$ and $\hat{p}_A$ is close enough to $\hat{p}_B$, if participants are rewarded based on their own contributions, then a strategy (x,y) is played in period 1 in a SPE of the game iff it is socially optimal.*

[11] The assumption that players independently choose which approach to follow and then share the output of their searches is probably more descriptive of electronic ideation sessions (in which participants are seated at different terminals from which they type in their new ideas and read the other participants' ideas) than it is of traditional face-to-face sessions. As will be seen later, electronic sessions have been shown to be more effective than face-to-face sessions (Gallupe et al. 1991, Nunamaker et al. 1987).

High-α case (slowly decreasing returns)

When α is close to 1, there are situations in which exploration is optimal for the group. However, exploration results in a short-term loss suffered only by the explorer, and a potential long-term benefit enjoyed by the whole group. There are some cases in which it is socially optimal that at least one player explores in period 1, but no player will do so in any SPE of the game. This is illustrated by the following proposition when α is equal to 1:

<u>Proposition 3</u>: *If $\alpha = 1$, then for all $\hat{p}_B$, for all $\hat{p}_A < \hat{p}_B$ close enough to $\hat{p}_B$, there exists $N_{Alow} < N_{Ahigh}$ ($V_{low} < V_{high}$) such that if N_B is large enough ($Var(p_B)$ low enough), then*

- *$N_A < N_{Ahigh}$ ($Var(p_A) > V_{low}$) implies that it is not socially optimal for both players to choose B in period 1 (at least one player should explore)*
- *$N_{Alow} < N_A < N_{Ahigh}$ ($V_{low} < Var(p_A) < V_{high}$) implies that both players choose B in period 1 in any SPE of the game (no player actually explores)*

Interesting participants in the future of the group's output by rewarding them for their impact

Proposition 3 demonstrates a potential misalignment problem when building on ideas matters, that is, when the marginal returns to subsequent ideas are high. In order to overcome this problem, participants should be forced to internalize the effect of their present actions on the future of the other participants' contribution. To study the dynamic nature of the incentives implied by this observation, let us consider a more general framework, with an infinite horizon, N participants, and a per-period discount factor δ. Assume that participant i's output from period t, y_{it}, has a probability density function $f(y_{it} \mid x_{1t},\ldots,x_{Nt},(y_{j\tau})_{j=1\ldots N,\tau=1\ldots t-1})$, where $x_{j\tau}$ is participant j's action in period τ, and that participant i's contribution from period t is $C_{it}=C((y_{j\tau})_{j=1\ldots N,\tau=1\ldots t})$. The two-armed bandit model considered earlier is a special case of this general model.

Interesting participants in the group's future output allows aligning the incentives, but it might also lead to free-riding (Holmstrom 1982, Wu et al. 2004). The technical appendix illustrates this fact by considering the incentive scheme that consists of rewarding a participant for his or her action in period *t* according to a weighted average between his or her contribution in period *t* and the other participants' contribution in period *t+1*, and shows that it is in fact equivalent to a proportional sharing rule. Although the objective function of the participants is proportional to that of the moderator under such a scheme, it is possible for a participant to be paid without participating in the search. This imposes a cost on the moderator, since the rewards given to participants have to be high enough so that each participant is willing to search for ideas in each period.[12]

In order to address free riding while still aligning the objectives, one option is to reward each participant more precisely for that part of the group's future contribution that depends directly on his or her actions, i.e., for his or her impact on the group. In particular, the following proposition considers rewarding participant *i* for his or her action in period *t* based on the average between his or her contribution in period *t*, and the impact of this action on the group's future contribution. Under such a scheme, the objectives of the participants are aligned with that of the moderator, and free riding is reduced because a participant does not get rewarded for the other participants' contribution unless it is tied to his or her own actions. This reduces the share of the created value that needs to be redistributed to the participants.[13]

---

[12] I assume that the moderator's valuation for the ideas is high enough that it is possible and optimal for him or her to induce the participants to search in each period.

[13] The impact of participant *i*'s action in period *t* is defined here as: $\sum_{\tau=t+1}^{+\infty}\delta^{\tau-t}\lambda(t,\tau)\sum_{j\neq i}[C_{j\tau}-\tilde{C}_{j\tau}(i,t)]$ where $\tilde{C}_{j\tau}(i,t)$ is the expected value of $C_{j\tau}$ obtained when y_{it} is null and all players maximize the group's expected output at each period, δ is a discount factor, and $\lambda(t,\tau)=\delta^{t}.(1-\delta)/(\delta-\delta^{\tau})$ is such that $\sum_{t=1}^{\tau-1}\lambda(t,\tau)=1$ for all *t* and *τ*.

Proposition 4: Rewarding each participant for the average between his or her contribution and his or her impact aligns the objectives of the participants with that of the moderator, and is such that the total expected payoffs distributed in the session are lower than under a proportional sharing rule.

Summary of this section

This section studied incentive schemes rewarding participants for a weighted average between their individual contributions and their impact on the group. Proposition 4 showed that such a scheme can always lead to optimal search by the participants while addressing free riding. Propositions 2 and 3 studied a special case in which the weight on impact is null. A key parameter of the model is the speed with which the marginal contribution of successive ideas building on each other decreases. Unless marginal contribution decreases very quickly, rewarding participants based solely on their individual contributions leads to insufficient exploration.

At least three issues arise when trying to test these predictions experimentally. The first is a measurement issue. Impact, in particular, can be very hard to measure. The next section will address this issue by proposing an idea generation system which allows measuring (albeit not perfectly) the impact of each participant. Second, the model assumes that all participants search for ideas at each period. Hence the different incentive schemes should be compared holding participation constant. Third, the optimal relative weight on individual contribution versus impact depends on the scale of their measurements.[14] Being unable to experimentally compare the scheme rewarding individual contribution to the "optimal" scheme, I only consider rewarding participants based on their individual contributions or on their impact. The theory is silent re-

[14] For example, a theoretical prediction that the same weight should be put on impact and individual contribution does not imply that this should be the case in the ideation game proposed in Section 3, because individual contribution and impact are measured on different scales (the contribution of one idea is not assumed to be equivalent to the impact implied by one citation).

garding the relative contribution achieved by these two extreme reward schemes. However it suggests that rewarding participants based on their impact will increase the amount of exploration in the session. It also suggests that increasing the amount of exploration will be more beneficial if the marginal contribution of ideas building on each other is slowly diminishing. These predictions can be summarized by the following two hypotheses:

Hypothesis 1: *The amount of exploration in the session is higher if participants are rewarded for their impact than it is if they are rewarded for their individual contributions.*

Hypothesis 2: *Increasing the amount of exploration is more beneficial if the marginal contribution of ideas building on each other is slowly diminishing.*

3. An Ideation Game

I now describe an incentive-compatible "ideation game" that allows testing and implementing the insights derived in Section 2. In this game, participants score points for their ideas, the scoring scheme being adjustable to reward individual contribution, impact, or a weighted average of the two. The design of the game is based on the idea generation, bibliometric, and contract theory literatures. Table 1 provides a summary of the requirements imposed on the system and how they were addressed.

Addressing "fear of evaluation" and "production blocking"

Although the primary requirement imposed on this idea generation system is to allow implementing and testing the incentive schemes treated in Section 2, it should also be compatible with the existing literature on idea generation. In particular, two main issues have been identified with classical (face-to-face) idea generation sessions, in addition to free riding (Diehl and Stroebe 1987). The first one, "production blocking," happens with classical idea generation sessions when participants are unable to express themselves simultaneously. The second one, "fear of evaluation," corresponds to the fear of negative evaluation by the other participants, the moderator, or external judges.

These two issues have been shown to be reduced by electronic idea generation sessions (Nunamaker et al. 1987, Gallupe et al. 1991, Gallupe et al. 1992, Dennis and Valacich 1993, Valacich et al. 1994), in which participation is asynchronous (therefore reducing production blocking) and in which the participants are anonymous (therefore reducing fear of evaluation). More precisely, in a typical electronic brainstorming session, each participant is seated at a computer terminal connected to all other terminals, enters ideas at his or her own pace, and has access to the ideas entered by the other group members. The online system proposed here is anonymous and asynchronous as well.

Table 1
Ideation game – requirements and corresponding solutions

Requirement	Proposed solution
Address "Production Blocking"	Asynchronous
Address "Fear of Evaluation"	Anonymous Objective measures Mutual monitoring
Measure contribution and impact	Ideas structured into trees
Prevent cheating	Relational contract Mutual monitoring

Measuring the contribution and the impact of participants

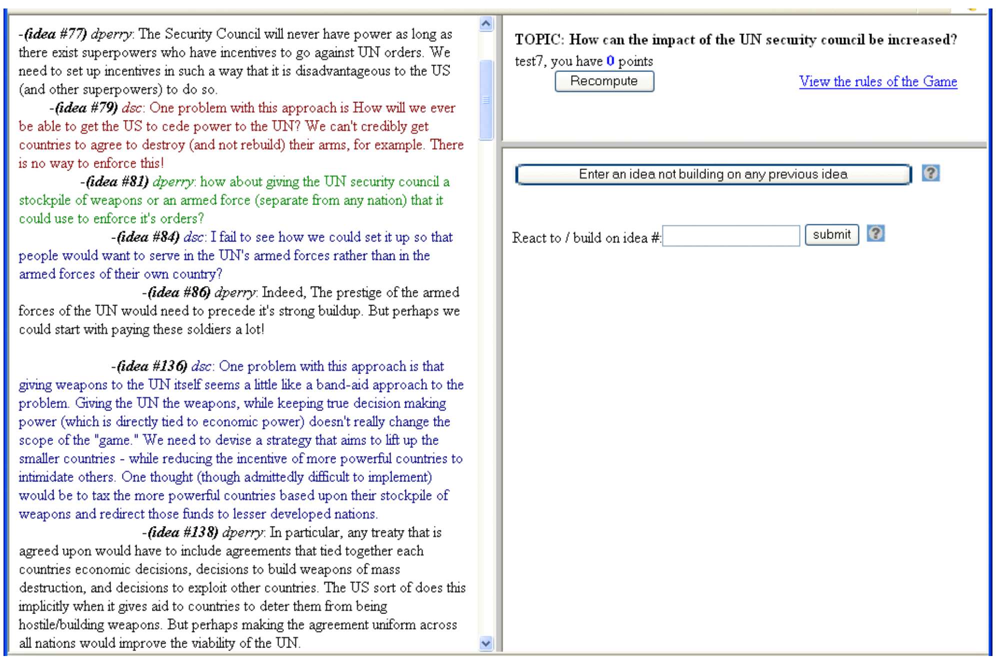
One way to measure the contribution and impact of participants would be to have them evaluated by external judges. However, this approach has limitations. First, it is likely to trigger fear of evaluation. Second, the evaluation criteria of the moderator, the judges and the participants might differ. Third, it increases the cost of the session. The proposed ideation game instead adopts a structure that provides objective measures of these quantities. These objective measures do not require additional evaluation, are computed based on well-stated rules, and offer continuous feedback on the performance of the participants.

Bibliometric research suggests that the number of ideas submitted by a participant should be a good measure of his or her contribution, and that his or her number of "citations" should be a good measure of his or her impact (King 1987). The proposed structure adapts the concept of citation count from the field of scientific publications (which are the focus of bibliometry), to idea generation. An example is provided in Figure 1. Ideas are organized into "trees," such that the "son" of an idea appears below it, in a different color and with a different indentation (one level further to the right). The relation between a "father-idea" and a "son-idea" is that the son-idea builds on its father-idea. More precisely, when a participant enters a new idea, he or she has the option to place it at the top of a new tree (by selecting "enter an idea not building on any previous idea") or to build on an existing idea (by entering the identification number of the father-idea and clicking on "build on/react to idea #..."). Note that an idea can have more than one "son," but at most one "father" (the investigation of more complex graph structures allowing multiple fathers is left for future research). The assignment of a father-idea to a new idea is left to the author of the new idea (see below how incentives are given to make the correct assignment).

If a participant decides to build on an idea, a menu of "conjunctive phrases" (e.g., "More precisely", "On the other hand", "However"…) is offered to him or her (a blank option is also available), in order to facilitate the articulation of the links between ideas. Pretests suggested that these phrases were useful to the participants.

With this structure, the impact of an idea y can be measured by considering the ideas that are lower than y in the same tree. In the experiments described in the next section, impact was simply measured by counting these "descendant" ideas. More precisely, participants rewarded for their impact scored one point for each new idea that was entered in a tree below one of their ideas.[15] For example, if idea #2 built on idea #1, and idea #3 on idea #2, then when idea #3 was posted, one point was given to the authors of both idea #1 and idea #2. Participants rewarded for their own contributions scored a fixed number of points per idea. The exploration of alternative measures is left for future research.

Figure 1
Example of an ideation game



Preventing cheating

Although this structure provides potential measures of the contribution and impact of the different participants, if the scoring system were entirely automatic and if no subjective judgment were ever made, participants could "cheat" by simply posting irrelevant ideas in order to score unbounded numbers of points. Consequently, some level of subjective judgment seems inevita-

[15] In order to facilitate the comparison across conditions, points were scored only for the first five "descendant" ideas in the second experiment. See Section 4 for details.

ble. These judgments, however, should keep fear of evaluation to a minimum, and limit costly interventions by external judges. This is addressed by carefully defining the powers of the moderator of the session. First, the moderator does not have the right to decide which ideas deserve points. However, he or she has the power to expel participants who are found to cheat repeatedly, i.e., who submit ideas that clearly do not address the topic of the session or who wrongly assign new ideas to father-ideas. This type of relation between the participants and the moderator is called a "relational contract" in agency theory. In this relationship the source of motivation for the agent is the fear of termination of his or her relation with the principal (Baker at al. 1998, Levin 2003). In the present case, if the moderator checks the session often enough, it is optimal for participants to only submit ideas which they believe are relevant.

In order to reduce the frequency with which the moderator needs to check the session, this relational contract is complemented by a mutual monitoring mechanism (Knez and Simester 2001, Varian 1990). In particular, if participant *i* submits an idea which is found to be fraudulent by participant *j*, then *j* can "challenge" this idea.[16] This freezes the corresponding stream of ideas (no participant can build on a challenged idea), until the moderator visits the session and determines who, between *i* and *j*, was right.[17] The participant who was judged to be wrong then pays a fee to the other participant.[18] This mechanism is economical because it reduces the frequency with which the moderator needs to visit the session, without increasing its cost, since challenges result in a transfer of points between participants (no extra points need to be distributed). Another argument in favor of having participants monitor each other is the finding by Collaros and

[16] In the implementation used so far, participants cannot challenge an idea if it has already been built upon.
[17] The idea is deleted if *j* was right.
[18] The amount of this fee was fine-tuned based on pretests and was set to 5 points in experiment 1 and 25 points in Experiment 2 (the average number of points per idea was higher in experiment 2).

Anderson (1969) that fear of evaluation is greater when the evaluation is done by an expert judge rather than by fellow participants.

Relation with existing idea generation tools

Most existing idea generation tools can be thought of as reflecting one of two views of creativity. The first view is that participants should be induced to think in a random fashion. This widely spread belief, based on the assumption that anarchy of thought increases the probability of creative ideas, has led to the development of tools such as brainstorming (Osborn 1957), synectics (Prince 1970), or lateral thinking (De Bono 1970). In contrast, recent papers suggest that structure, and not randomness, is the key to creativity (Goldenberg et al. 1999a). This structured view, according to which creativity can be achieved through the detailed application of well-defined operations, has led to systematic approaches to idea generation. Two important examples are inventive templates (Goldenberg et al. 1999b, Goldenberg and Mazursky 2002), and TRIZ (Altshuller 2000). The ideation game proposed in this paper is probably closer to the latter view. By organizing ideas into trees and by offering the participants a menu of conjunctive phrases that highlight and articulate the relation between ideas, this structure requires the participants to identify the components of the previous ideas on which their new ideas build more heavily. A possible extension of this game would be to replace ideas with product configurations and to replace conjunctive phrases with the five inventive templates proposed by Goldenberg and Mazursky (2002). Instead of building on each other's ideas, participants would apply inventive templates to each other's proposed product configurations.

Practical implementation

This ideation game was programmed in php, using a MySQL database. Its structure and the instructions were fine-tuned based on two pretests. The pretests used participants from a professional marketing research/strategy firm. One pretest dealt with improving the participants' workspace and a second dealt with re-entering a market.

In the implementation, the moderator also has the ability to enter comments, either on specific ideas or on the overall session. Participants also have the ability to enter comments on specific ideas. This allows them, for example, to ask for clarification or to make observations that do not directly address the topic of the session. The difference between comments and ideas is that comments are not required to contribute positively to the session, and are not rewarded.

4. Experiments

I now report the results of two experiments run using the ideation game described above. Experiment 1 tested whether incentives have the potential to improve idea generation and allowed an initial investigation of the difference between rewarding participants for their individual contributions versus their impact. Being performed online over a few days, this experiment allowed studying the effect of incentives on the level of participation and on the dynamics of the sessions. Experiment 2 was performed in a lab, and allowed a more direct test of the hypotheses proposed in Section 2.

In both experiments, the only difference between conditions was the incentive system, that is, the manner in which points were scored in the game. Each participant was randomly assigned to one idea generation session, and each idea generation session was assigned to one incentive condition (all participants in a given session were given the same incentive scheme). The points scored during the session were later translated into cash rewards.

Experiment 1

Experimental design

Experiment 1 had three conditions. A total of 78 participants took part in the experiment over a period of 10 days, signing up for the sessions at their convenience. Three sets of three parallel sessions were run (one session per condition in each set), defining three "waves" of participants.[19] Each session lasted up to five days, and within each wave, the termination time of the three sessions was the same.

Recall that the analysis in Section 2 relies on the assumption of an equal level of participation across participants. This was addressed by calibrating the value of points in the different conditions based on a pretest, such that participants in all conditions could expect similar payoffs.[20] More precisely, the first condition (the "Flat" condition) was such that participants received a flat reward of $10 for participation. In this condition, no points were scored (points were not even mentioned). In the second condition (the "Own" condition), participants were rewarded based exclusively on their own contributions, and scored one point per idea submitted (each point was worth $3). In the third condition (the "Impact" condition), each participant was rewarded based exclusively on the impact of his or her ideas. Participants scored one point each time an idea was submitted that built on one of their ideas (see previous section for the details of the scoring scheme). Each point was worth $2 for the first two waves of participants; the value of points was decreased by half to $1 for the last wave. The results for the third wave were similar

---

[19] The first wave consisted of the first 36 participants who signed up and had 12 participants per condition, the second wave consisted of the following 20 participants, with 7 participants in the "Flat" and "Own" conditions and 6 in the "Impact" condition, and the last wave consisted of the last 22 participants, with 8 in the "Flat" condition, and 7 in the "Own" and "Impact" conditions. Participants were cyclically assigned to the "flat", "own", or "impact" session of the corresponding wave.

[20] In this pretest, a group of approximately 15 employees from a market research firm generated ideas on "How can we better utilize our office space?".

to those for the first two waves; hence the same analysis was applied to the data from the three waves.

In order to provide a strong test of the hypothesis that incentives can improve idea generation, an engaging topic, as well as some motivated participants, were selected. In particular, the topic was: "How can the impact of the UN Security Council be increased?" (the experiment was run in March 2003, during the early days of the US-led war in Iraq), and the participants were recruited at an anti-war walkout in a major metropolitan area on the east coast, as well as on the campus of an east coast university. Three graduate students in political science in the same university (naïve to the hypotheses and paid $50 each for their work) were later hired as expert judges.

Results

Both the quantitative and qualitative results are summarized in Table 2. Compared to the "Own" condition, participants in the "Impact" condition submitted significantly more unique ideas on average (p-value < 0.05),[21] where the number of unique ideas is defined as the number of ideas minus the number of redundant ideas (an idea is classified as redundant if it was judged as redundant by any of the three judges). They were also significantly more likely to submit at least one unique idea (p-value < 0.04), submitted significantly more unique ideas conditioning on submitting at least one (p-value < 0.02), and submitted ideas that were on average 76% longer.

[21] The numbers of ideas submitted by different participants are not totally independent because participants are influenced by the ideas submitted by the other members of their group. An endogeneity issue arises, because the number of ideas submitted by participant i depends on the number of ideas submitted by participant j, and vice versa. In order to limit this issue, a regression was run with the number of ideas submitted by participant i as the dependent variable, and three dummies (one per condition) as well as the number of ideas that were submitted by the other members of participant i's group *before* he or she signed up as independent variables. This last independent variable is still endogenous, but it is pre-determined, leading to consistent estimates of the parameters (Greene 2000). A simple contrast analysis yielded similar results as this regression.

Participants in the "Own" condition, in turn, performed better than participants in the "Flat" condition, although not significantly so on all metrics.[22]

Table 2
Results

	Impact	Own	Flat
Quantitative results			
Number of unique ideas per participant	4.6*	2.3	0.8
Proportion of participants who posted at least one unique idea	68%*	54%	48%
Number of unique ideas given that at least one	6.8*	4.3†	1.6
Number of words per idea	79.3*	44.9	45.6
Qualitative ratings			
Total contribution	5.8	4.7	3.6
Number of star ideas	6.2	3.3	1.7
Proportion of ideas identified as stars by at least one judge	8.0%	8.2%	10.0%
Breadth	6.1	5.3	3.2
Depth	6.8	4.6	3.6
Novelty	5.6	5.2	4.4
Thought-provoking	5.8	5.2	4.0
Interactivity	7.3	4.7	3.2

* = Impact significantly larger than Own at the 0.05 level. † = Own significantly larger than Flat at the 0.05 level.

The qualitative results reported in Table 2 were obtained by averaging the ratings of the three judges. The ratings were found to have a reliability of 0.75 (Rust and Cooil 1994), which is above the 0.7 benchmark often used in market research (Boulding et al. 1993).[23] (Given the low number of judges, significance tests for the qualitative ratings are not available.) The sessions in the "Impact" condition were judged to have a larger overall contribution, more "star" (i.e., very good) ideas (although the *proportion* of "star" ideas was non-significantly different across condi-

[22] *p*-values of 0.18, 0.17, and 0.04 respectively for the number of unique ideas, for the proportion of participants who submitted at least one unique idea, and the number of unique ideas given that at least one was submitted.
[23] All ratings, except for the number of star ideas, are on a 10-point scale. For each measure and for each judge, the ratings were normalized to range between 0 and 1. Then ratings were classified into three intervals: [0;1/3[, [1/3;2/3[, and [2/3;1]. The same procedure was used in experiment 2.

tions), more breadth, depth, and to have ideas that on average were more novel, thought-provoking and interactive (interactivity is defined here as the degree to which a "son-idea" builds on its "father-idea" and improves upon it).

The effect of incentives on participation

Incentives seem to have had an effect on the effort spent by participants. Indeed, 100% of the participants who submitted at least one unique idea in the "Flat" condition did so in the first 30 minutes after signing up, versus 65% for the "Impact" condition and 64% for the "Own" condition. In particular, 18% (respectively 21%) of the participants who submitted at least one idea in the "Impact" condition (respectively the "Own" condition) did so more than three hours after signing up. This is consistent with the hypothesis that all three conditions faced the same distribution of participants, but that when incentives were present, participants tried harder to generate ideas and did not give up as easily.

For participants who submitted at least one idea, incentives also influenced the frequency of visit to the session as well as the number of ideas per visit. Let us define "pockets" of ideas such that two successive ideas by the same participant are in the same pocket if they were submitted less than 20 minutes apart (the time at which each idea was submitted was recorded).[24] Table 3 reports the average number of ideas per pocket for the participants who posted at least one idea, as well as the average number of pockets. Participants in the "Own" and "Impact" conditions submitted significantly more ideas per visit compared to participants in the "Flat" condition (p-values < 0.04). However, participants in the "Impact" condition did not submit significantly more ideas per visit compared to the participants in the "Own" condition (p-value = 0.35). The results so far have suggested a similar effect of incentives on participation in the "Own" and

[24] Similar results were obtained when defining pockets using 10 or 30 minutes.

"Impact" conditions. However, participants in the "Impact" condition visited the session marginally significantly more often than participants in the "Own" condition (p-value = 0.09). Participants in the "Own" condition did not visit the session significantly more often than participants in the "Flat" condition (p-value = 0.19).

Table 3
Dynamics

	Impact	Own	Flat
Number of ideas per pocket	2.9	2.3	1.2
Number of pockets per participant	2.6	1.6	1.2

The effect of incentives on the level of interactions

It is interesting to study how incentives influenced the overall number of ideas submitted in the session as a function of time. Figure 2 (corresponding to the first wave of participants - the graphs for the other two waves having similar characteristics) shows that the process in the "Impact" condition is S-shaped, whereas the process in the "Flat" condition tends to be smoother. [25] The process in the "Own" condition is somewhat intermediary between these two extremes.

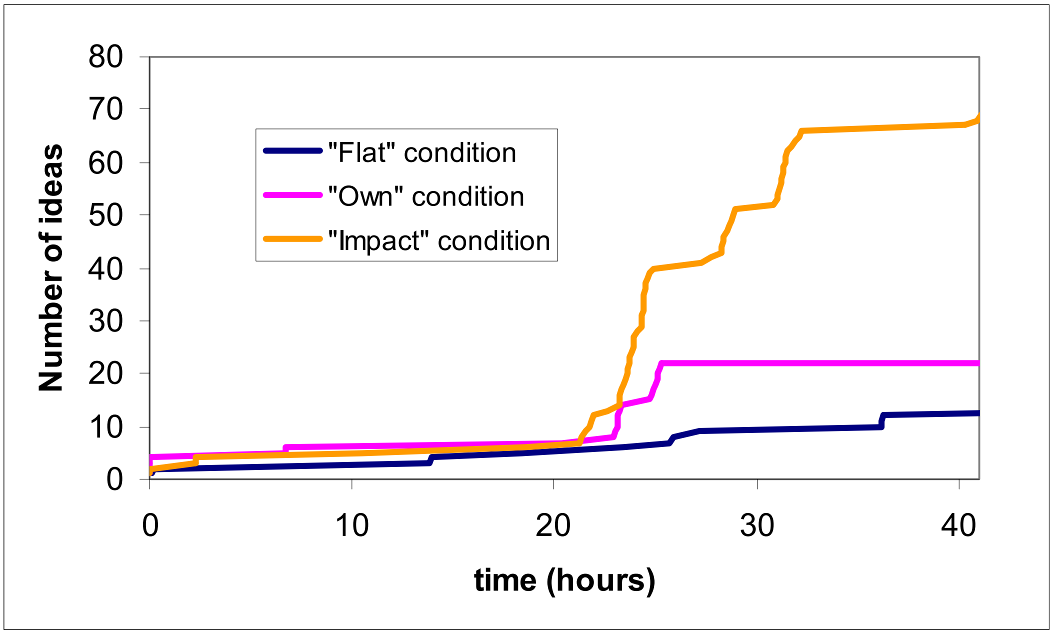
In order to better understand the shapes of these graphs, it is useful to plot similar graphs at the individual level. Figure A1 in the appendix gives an example for a particular session, with one graph per participant.[26] These graphs indicate that the shape of the aggregate process in the "Impact" condition comes from the fact that at the individual level, participation in this condition is also characterized by a period of inactivity followed by a period of high activity during which

[25] Time on the x axis is the total cumulative time spent by participants in the session when the idea was submitted, divided by the number of participants. For example, if an idea was posted at 5 pm, and at this time two participants had signed up, one at 2 pm and one at 3 pm, then the reported time is [(5-2)+(5-3)]/2=2.5 hours. A simpler measure of time, the total clock time elapsed since the first participant signed up, gives similar results.

[26] Clock time is used in the x axis for the individual-level graphs.

participation is roughly linear in time. Moreover, the periods of high activity start at very similar times for the different participants within a session, suggesting a high level of interaction between participants. On the other hand, Figure A2 shows that sessions in the "Flat" condition displayed much less structure, which explains the smoother aspect of the aggregate graphs (the aggregate graphs in this condition are the sum of independent graphs).

Figure 2
Number of ideas as a function of time



Another approach to studying interactions among participants is to investigate the influence of incentives on the relative proportion of the following three categories of ideas: "new ideas" that do not build on any previous ideas; "self-citations" that build directly on an idea from the same participant; and "non self-citations" that build on another participant's idea. A higher proportion of "non-self citations" may be interpreted as reflecting a higher level of interactions between participants. As indicated by Table 4, the proportion of "non self-citations" was highest in the "Impact" condition, followed by the "Own" condition and the "Flat" conditions (all numbers in the last row are significant at the p<0.05 level)

Table 4
Proportions of “new ideas”, “self citations” and “non-self citations”
Experiment 1

	Impact	Own	Flat
“New ideas”	13.5%	30.8%	53.3%
“Self citations”	4.5%	9.0%	10.0%
“Non self citations”	82.1%	60.3%	36.7%

Reconciling the results with the social psychology literature

The results of this experiment suggest that incentives have the potential to improve idea generation. However, as was mentioned in the introduction, the social psychology literature seems to predict that incentives are more likely to *hurt* idea generation (Spence 1956, Zajonc 1965, McCullers 1978, McGraw 1978). This prediction relies in great part on the observation that idea generation is not a task in which there exist easy algorithmic solutions, achievable by the straightforward application of certain operations. McGraw (1978) contrasts tasks such that the path to a solution is “well mapped and straightforward” with tasks such that this path is more complicated and obscure. He argues that incentives are likely not to help with this second type of tasks, but notes that “it is nonetheless possible for reward to facilitate performance on such problems in the case where the algorithm (leading to a solution) is made obvious” (p. 54). Such facilitation might have happened in the present experiment. In particular, the structure of the ideation game might have made the mental steps leading to new ideas more transparent, allowing the participants to approach the task in a more “algorithmic” manner.

Discussion

Experiment 1 suggests that incentives do have the potential to improve idea generation. It also suggests that rewarding individual contribution and rewarding impact lead to very different behaviors. However it does not allow a direct test of Hypotheses 1 and 2. In particular, incentives seem to have had an influence on the participants' level of involvement in the sessions, while the hypotheses assume that participation in the "Own" and "Impact" conditions is similar. Although the monetary value of points was calibrated to equate the expected payoff per idea across conditions, the calibration was based on a pretest using a different topic, and did not take participants' subjective beliefs into account. Hence the expected payoff per idea might have been perceived as higher in the "Impact" condition. Furthermore, experiment 1 does not allow comparing topics with different speeds of diminishing marginal contribution, and does not allow studying the level of exploration as a function of the incentives.

These limitations were addressed in three ways in experiment 2. First, the amount of time spent by participants on the task was controlled by running the sessions in a laboratory, which also allows a more direct investigation of exploration. Second, the expected payoff per idea in the "Impact" condition was bounded such that ideas scored points for their first 5 citations only. Third, two different topics were tested, with different speeds of diminishing marginal contribution.

Experiment 2

Experimental design

110 students from a large east coast university took part in 45 minute idea generation sessions in groups of up to 4 participants. Each group was randomly assigned to one of four conditions in a 2 (type of incentives = "Own" versus "Impact") x 2 (topic: slowly diminishing marginal contribution versus quickly diminishing marginal contribution) design. Participants rewarded for their number of ideas scored 2 points per idea, and participants rewarded for their number of citations scored 1 point per citation, with a maximum of 5 points per idea. Each point

was worth 20 cents, and participants received an additional $3 show up fee. The average number of points per idea in the "Impact" conditions was 1.43, making the expected payoff per idea comparable under both incentive schemes (the number of points per idea in the "Own" conditions was calibrated based on a pretest). The topic of each session was either "How could [*our University*] improve the education that it provides to its students?" or "How to make an umbrella that would appeal to women?". The first topic was chosen because it appeared to be one of the broadest and richest possible topics in a New Product Development context, asking the consumers of an organization to provide ideas on how to improve its product. Hence it was expected to be characterized by a slowly diminishing marginal contribution. The second topic was chosen for its apparently narrower scope, at least to the participant population. This topic was expected to be characterized by a faster diminishing marginal contribution. I will later attempt to validate this assumption of different speeds of diminishing marginal contribution.

Results – slowly diminishing marginal contribution

Let us first consider the sessions on "How could [*our University*] improve the education that it provides to its students?". Hypothesis 1 predicts that the amount of exploration will be higher when participants are rewarded for their impact. Exploration is characterized by a lower short-term expected contribution but a higher potential long-term impact. This suggests that a participant who engages in exploration will produce fewer ideas per unit of time, but will produce ideas that are more impactful.[27]

Let us use the difference between the submission time of an idea and the submission time of the previous idea by the same participant as a proxy for the amount of time spent generating

[27] In the context of the model from Section 2, exploration consists in following an approach with an a priori lower probability of success. However, if exploration is successful and an idea is generated, the expected long-term contribution from this approach is higher.

the idea, and adopt the same classification of ideas as in the previous experiment. Table 5 indicates that participants in the "Impact" condition spent significantly more time generating "new ideas" compared to participants in the "Own" condition (p-value < 0.01). Moreover, the correlation between the time spent generating a "new idea" and the total number of ideas submitted by other participants lower in the same tree was significantly positive ($\rho = 0.17$, p-value < 0.01).[28] This is consistent with exploration: participants are slower at generating new ideas, but the ideas are more impactful.[29]

Table 5
Average time to generate ideas (in seconds)
Slowly diminishing marginal contribution

	Impact	Own
"New ideas"	115.7	77.7
"Self citations"	82.5	67.5
"Non self citations"	120.1	113.6

Table 6
Proportions of "new ideas", "self citations" and "non-self citations"
Slowly diminishing marginal contribution

	Impact	Own
"New ideas"	26.8%	45.4%
"Self citations"	21.5%	21.6%
"Non self citations"	51.7%	33.0%

[28] The same result holds if the analysis is done on all ideas (p-value for time < 0.01 and p-value for correlation < 0.02). However focusing on "new ideas" allows a more direct investigation of exploration, because time is driven by more factors (e.g., how inspiring the previous ideas are) when the idea is not "new".
[29] This is unlikely to be due to differences in the lengths of the ideas, since the average number of characters per "new idea" is actually lower in the "Impact" condition (p-value = 0.06). Moreover, the correlation between the number of characters in a "new idea" and the total number of ideas submitted by other participants in the same tree is non-significant ($\rho = -0.02$, p-value = 0.64)

Moreover, as indicated in Table 6, the proportion of "non self citations" was much greater in the "Impact" condition compared to the "Own" condition (p-value < 0.01), suggesting a greater level of interactions.

With higher levels of exploration and interactions, one might expect the number of ideas to be higher in the "Impact" condition. However, the average number of ideas per participant was non-significantly *lower* in the "Impact" condition compared to the "Own" condition (23.3 ideas on average in the "Impact" condition versus 28.9 in the "Own" condition, p-value= 0.326[30]). (See Tables A1 and A2 in the appendix for a summary of the quantitative and qualitative ratings for experiment 2.) Table 5 sheds some light on this apparently contradicting result. First, participants in the "Impact" condition were slower to generate "new ideas". Consequently, their ideas were more impactful, which increased the number of "non-self citations" in the session. This further reduced the speed at which ideas were produced, since Table 5 indicates that building on other participants' ideas is more time consuming than generating one's own ideas ("non self citations" are the most time consuming ideas to generate).[31] In other words, both the higher level of exploration and the higher level of interactions seem to contribute to a *decrease* in the number of ideas in limited time sessions.

The transcripts were rated by one Ph.D student in marketing and three graduate students in education (blind to the hypotheses and paid $50 each for their work). The judges gave an overall contribution score between 0 and 10 to each session, and identified "star" ideas. The overall reliability of the four judges was 0.67. Removing one of the judges increased reliability

[30] In this experiment the p-values on the number of ideas were obtained from a two-way ANOVA of the number of ideas on the incentive scheme and the number of participants in the group.
[31] Understanding the source of this difference is beyond the scope of this paper and is left for future research. The effect remains significant after controlling for differences in the length of the ideas, as well as taking into account the time required to read other participants' ideas. Details are available from the author.

to 0.72. The ratings as well as the proportion of "star" ideas were slightly higher for the "Impact" sessions (average evaluations: 5.58 vs. 5.36, proportion of ideas identified as stars by at least one judge: 12.0% vs. 10.5% - *p*-value=0.35). Note that the correlation between a "star" dummy variable (equal to 1 if the idea was judged as a star by at least one judge) and the number of ideas submitted by other participants lower in the same tree was significantly positive (ρ = 0.105, *p*-value < 0.01), suggesting that star ideas tend to have been more impactful.

Results – quickly diminishing marginal contribution

Hypothesis 2 suggests that exploration is less beneficial with topics characterized by a quickly diminishing marginal contribution. Indeed, the correlation between the time spent generating a "new idea" and the total number of ideas submitted by other participants lower in the same tree was no longer significant when the topic was "How to make an umbrella that would appeal to women?" (ρ = -0.005, *p*-value = 0.90). However Table 7 indicates that participants still spent significantly more time on average generating "new ideas" in the "Impact" condition (*p*-value < 0.01), suggesting that the level of exploration was still higher although exploration was less fruitful.

Table 7
Average time to generate ideas (in seconds)
Quickly diminishing marginal contribution

	Impact	Own
"New ideas"	103.7	79.1
"Self citations"	55.8	88.1
"Non self citations"	111.0	113.1

Table 8 indicates that the level of interactions was no longer higher in the "Impact" condition. On the contrary, the proportion of "self citations" was significantly higher in the "Impact" condition (*p*-value<0.01). This suggests that at least some participants perceived that they were unlikely to generate ideas that would inspire other participants, and instead cited themselves in order to increase their payments.

Table 8
Proportions of "new ideas", "self citations" and "non-self citations"
Quickly diminishing marginal contribution

	Impact	Own
"New ideas"	32.9%	51.9%
"Self citations"	48.4%	18.4%
"Non self citations"	18.7%	29.7%

One alternative explanation for the higher average time spent on "new ideas" in the "Impact" condition is that participants generated the new idea as well as the self citations at once. However the correlation between the time spent generating a "new idea" and the number of "self-citations" of this idea was significantly negative (ρ = -0.17, *p*-value < 0.01). Instead, further analysis suggests that participants in the "Impact" condition can be classified into two groups: one that relied heavily on "self-citations" and spent relatively little time generating "new" ideas, and another group that did not rely as heavily on "self-citations" and spent more time on average generating "new" ideas. One might infer that the latter group unsuccessfully attempted to explore, while the former did not believe that exploration would be fruitful and adopted another strategy in order to increase their payments. Figure A3 in the Appendix illustrates these two clusters of participants.[32]

Due to the higher proportion of "self-citations", the average number of ideas per participant was (non-significantly) higher in the "Impact" condition (28.9 ideas on average in the "Impact" condition versus 25.9 in the "Own" condition, *p*-value=0.327). More interestingly, the av-

[32] Each dot in Figure A3 represents one participant.

erage number of "star" ideas, as evaluated by an Executive MBA student interested in launching a new umbrella brand, was also higher in the "Impact" condition (10.6 star ideas per session on average in the "Impact" condition versus 6.3 in the "Own" condition, p-value = 0.19).

Speed of diminishing marginal contribution

I have assumed that the topic "How could [*our University*] improve the education that it provides to its students?" is characterized by a more slowly diminishing marginal contribution than the topic "How to make an umbrella that would appeal to women?". In the "Own" conditions, the proportion of ideas that built on other ideas ("self citation" or "non self citation") was significantly higher with the former topic compared to the latter (54.6% versus 48.1%, p-value < 0.02). In other words, even when participants did not have any direct incentive to build on ideas, they did so significantly more under the first topic. This suggests that building on ideas was more rewarding under the first topic, which indicates a more slowly diminishing marginal contribution.[33]

Discussion

As predicted by Hypothesis 1, experiment 2 suggests that rewarding participants for their impact leads to more exploration. Moreover, as predicted by Hypothesis 2, this experiment also suggests an interaction between speed of diminishing marginal contribution and payoff from exploration. More precisely, the correlation between the time taken to generate a "new" idea and its impact was significant only under the topic with the slower diminishing marginal contribution. An interesting result from Experiment 2, which might be a topic for future research, is that an increase in exploration and in the level of interactions between participants leads to a decrease in

[33] In terms of the model from Section 2, building on an idea for an expected contribution of $(n+1)^{\alpha} - n^{\alpha}$ is relatively more rewarding compared to investigating a new approach for an expected contribution of 1. This suggests that α is larger.

the number of ideas when participation is held constant. This suggests that the higher number of ideas observed in the "Impact" condition of experiment 1 relied on increased participation.

5. Managerial applications

This paper introduces a new tool for idea generation. In order to maximize the contribution of the participants, the theory recommends rewarding them for a weighted average between their individual contribution and their impact on the group. The optimal weight between individual contribution and impact could be determined empirically through experimentation.

Such experimentation has begun. At least three companies have applied the ideation game. Colgate Palmolive involved 41 of its employees from different functions in sessions on the following topic: "People are living longer than ever, but not necessarily feeling "old!" How do we deliver meaningful benefits to this growing population of aging but "ageless" consumers?", resulting in 180 ideas. The company was satisfied with the pilot and plans additional experimentation. Mark Andy, a manufacturer of large-scale, industrial printing presses, used the game to generate ideas around designing a press that addresses the customer needs identified through a Voice of the Customer process. The two week sessions involved a mix of 26 customers and 12 employees located in three different continents, and gave rise to over 200 ideas. The company found the tool to be an efficient way to capture customer-generated ideas on a global scale, and felt that the output was high-quality, actionable, and rich in details. Finally, the ideation game has been tested by a leading semiconductor manufacturer.

6. Summary and Future Research

In this paper, I propose a quantitative framework for the study of a mostly qualitative topic, idea generation, with a focus on the effect of incentives. I first derive an analytical model of the idea generation process. This model provides some insights on the search behaviors induced by different incentives under different conditions. I then develop an incentive compatible "ideation game," which allows implementing the insights from the theoretical analysis and testing some of the predictions. Finally, I show experimentally that incentives have the capability to improve idea generation, in a manner consistent with the theory.

Several opportunities for future research can be identified in addition to the ones mentioned throughout the paper. For example, it would be interesting to study in more details how incentives influence the evolution with time of the participants' allocation of effort between exploitation and exploration. This allocation could be compared to a normative benchmark, obtained by solving the corresponding multi-armed bandit problem. Future research might also seek a better understanding of the cognitive skills that are crucial to idea generation. As described earlier, there exist two major and apparently contradicting views on this issue: a "random" view (suggesting that randomness of thoughts is a key driver of creativity) and a "structured" view (suggesting that structure is the key to creativity). In a different paper (*author* 2005), I attempt to identify conditions under which each one of these two views is more likely to be valid.

References

Altshuller, Genrich (2000), *The Innovation Algorithm*, Technical Innovation Center, Inc., Worcester, MA.

Author (2005), "Bounded rationality models of idea generation", working paper.

Baker, G., R. Gibbons and K. Murphy (1998), "Relational Contracts and the Theory of the Firm", mimeo, MIT.

Bellman, R. (1961), *Adaptive Control Process: A guided Tour*, Princeton University Press.

Boulding, William, Richard Staelin, Valarie Zeithaml, and Ajay Kalra (1993), "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions", *Journal of Marketing Research*, 30 (February), 7-27.

Collaros, Panayiota A., and Lynn R. Anderson (1969), "Effect of Perceived Expertness Upon Creativity Of Members of Brainstorming Groups", *Journal of Applied Psychology*, Vol. 53, No. 2, 159-163

Dahan, Ely, and John R. Hauser (2001), "Product Development – Managing a Dispersed Process", in the *Handbook of Marketing*, Barton Weitz and Robin Wensley, Editors.

De Bono, Edward (1970), *Lateral thinking: a textbook of creativity*, Ward Lock Educational, London.

Dennis, Alan R., and Joseph S. Valacich (1993), "Computer Brainstorms: More Heads Are Better Than One", Journal of Applied Psychology, vol. 78, No. 4, 531-537

Desiraju, Ramarao (2004), "Costs and Benefits of Inducing Intrabrand Competition: The Role of Limited Liability", Marketing Science, VOl. 23, No. 3 (Summer), 429-450.

Diehl, Michael, and Wolfgang Stroebe (1987), "Productivity Loss in Brainstorming Groups: Toward the Solution of a Riddle", *Journal of Personality and Applied Psychology*, Vol. 53, No.3, 497-509

Gallupe, Brent R., Lana M. Bastianutti, and William H. Cooper (1991), "Unblocking Brainstorms", *Journal of Applied Psychology*, vol. 76, No. 1, 137-142

_______________, Alan R. Dennis, William H. Cooper, Joseph S. Valacich, Lana M. Bastianutti, and Jay F. Nunamaker (1992), "Electronic Brainstorming and Group Size", Academy of Management Journal, vol. 35, No. 2, 350-369

Gelman, Andrew B., John S. Carlin, Hal S. Stern, and Donald B. Rubin (1995), *Bayesian Data Analysis*, Chapman & Hall/CRC

Goldenberg, J., D. Mazursky, and S.Solomon (1999a), "Creative Sparks", *Science*, 285 1495-1496.

_______, _______, _______ (1999b), "Toward identifying the inventive templates of new products: A channeled ideation approach, *Journal of Marketing Research*, 36 (May) 200-210.

_______, _______ (2002), *Creativity in product innovation*, Cambridge University Press.

Greene, William H. (2000), *Econometric Analysis*, Prenctice Hall International, Inc.

Harkins, Stephen G., and Richard E. Petty (1982), "Effects of Task Difficulty and Task Uniqueness on Social Loafing", *Journal of Personality and Applied Psychology*, Vol. 43, No.6, 1214-1229

Holmstrom, Bengt (1982), "Moral hazard in teams", *Bell Journal of Economics*, 13(2): 324-340

Jeuland, Abel P., Steven M. Shugan (1983), "Managing Channel Profits", *Marketing Science*, Vol. 2, No.3, 239-272.

__________ , ________________(1988), "Managing Channel Profits: Reply" *Marketing Science*, Vol. 7, No.1, 103-106.

Kerr, Norbert L., and Steven E. Bruun (1983), "Dispensability of Member Effort and Group Motivation Losses: Free-Rider Effecs", *Journal of Personality and Applied Psychology*, Vol. 44, No.1, 78-94

King, Jean (1987), "A Review of Bibliometric and Other Science Indicators and Their Role in Research Evaluation", *Journal of Information Science*, 13, 261-276

Knez, Marc, and Duncan Simester (2001), "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines", *Journal of Labor Economics*, vol. 19, no. 4, 743-772

Lamm, Helmut and Gisela Trommsdorff (1973), "Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review", *European Journal of Social Psychology*, 3 (4), 361-388

Levin, J. (2003), "Relational Incentive Contracts", *American Economic Review*, vol. 93, No. 3 (June)

McCullers, J.C. (1978), "Issues in learning and motivation", in M.R. Lepper & D. Greene (Eds.), *The hidden costs of reward* (pp. 5-18), Hillsdale, NJ: Erlbaum

McGraw, K.O. (1978), "The detrimental effects of reward on performance: A literature review and a prediction model", in M.R. Lepper & D. Greene (Eds.), *The hidden costs of reward* (pp. 33-60), Hillsdale, NJ: Erlbaum

Moorthy, K. Sridhar (1987), "Managing Channel Profits: Comment", *Marketing Science*, Vol. 6, No.4, 375-379.

Nunamaker, Jay F., Jr., Lynda M. Applegate, and Benn R. Konsynski (1987), "Facilitating group creativity: Experience with a group decision support system", *Journal of Management Information Systems*, Vol. 3 No. 4 (Spring).

Osborn, A.F. (1957), *Applied Imagination*, (Rev. Ed.), New York: Scribner

Prince, George M. (1970), *The Practice of Creativity; a manual for dynamic group problem solving*, New York, Harper & Row.

Rust, Roland, and Bruce Cooil (1994), "Reliability Measures for Qualitative Data: Theory and Implications", *Journal of Marketing Research*, vol. XXXI (February 1994), 1-14.

Spence, K.W. (1956), *Behavior theory and conditioning*, New Haven: Yale University Press.

Valacich, Joseph S., Alan R. Dennis, and Terry Connolly (1994), "Idea Generation in Computer-Based Groups: A New Ending to an Old Story", Organizational Behavior and Human Decision Processes, 57, 448-467

Varian, Hal R. (1990), "Monitoring Agents with Other Agents", *Journal of Institutional and Theoretical Economics (JITE)*, Vol. 146, 153-174

Von Hippel, Eric, and Ralph Katz (2002), "Shifting Innovation to Users via Toolkits", *Management Science*, vol. 48, No.7, pp 821-833.

Williams, Kipling, Stephen Harkins, and Bibb Latané (1981), "Identifiability as a Deterrent to Social Loafing: Two Cheering Experiments", *Journal of Personality and Applied Psychology*, Vol. 40, No.2, 303-311

Wu, Dazhong, Gautam Ray, Xianjun Geng, and Andrew Whinston (2004), "Implications of Reduced Search Cost and Free Riding in E-Commerce", *Marketing Science*, vol. 23, No.2 (Spring), 255-262.

Zajonc, Robert B. (1965), "Social Facilitation", *Science*, Vol. 149, No. 16 (July), 269-274

Appendix

Figure A1

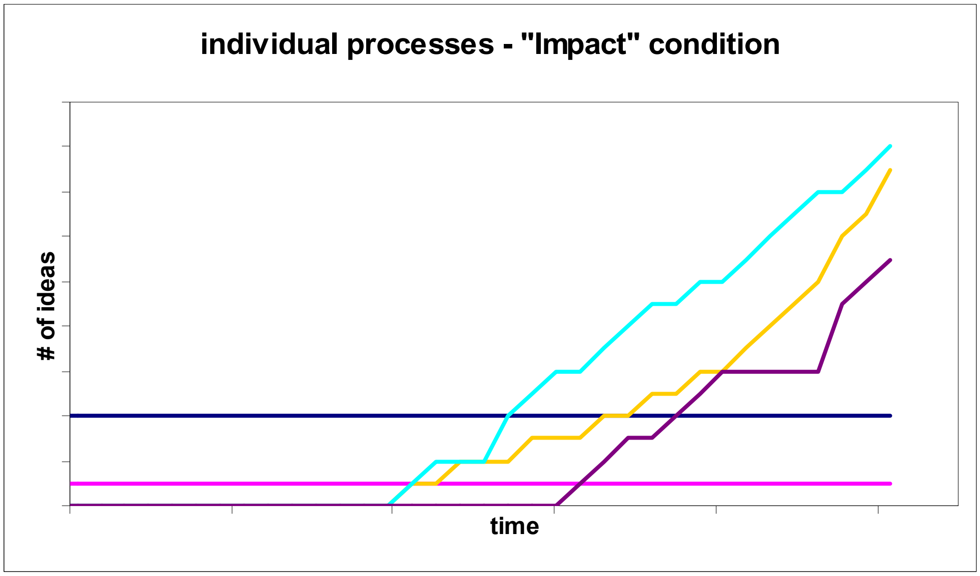


Figure A2

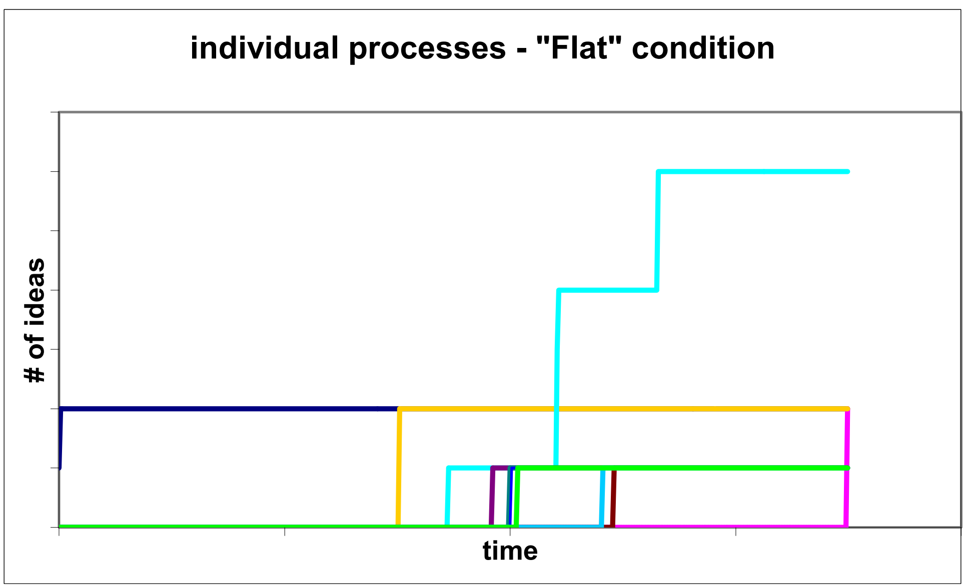


Figure A3
Exploration attempts versus self-citation

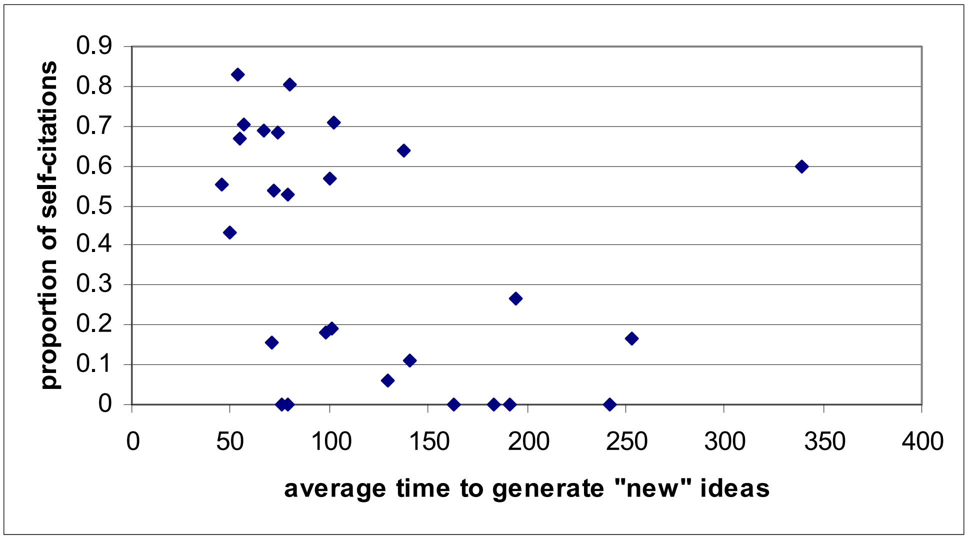


Table A1
Results – Experiment 2, slowly diminishing marginal contribution

	Impact	Own
Quantitative results		
Number of unique ideas per participant	23.3	28.9
Number of characters per idea	127.5	135.5†
Qualitative ratings		
Total contribution	5.6	5.4
Number of star ideas	2.8	3.1
Proportion of ideas identified as stars by at least one judge	12.0%	10.5%

†: Own significantly larger than Impact at the 0.05 level.

Table A2
Results – Experiment 2, Quickly diminishing marginal contribution

	Impact	Own
Quantitative results		
Number of unique ideas per participant	28.9	25.9
Number of characters per idea	80.0	89.8†
Qualitative ratings		
Total contribution	N/A	N/A
Number of star ideas	10.6	6.3
Proportion of ideas identified as stars	12.3%*	8.8%

*: Impact significantly larger than Own at the 0.05 level. †: Own significantly larger than Impact at the 0.05 level.
Note: the only output provided by the judge was the identification of "star" ideas.