Mining Hidden Profiles in the Collaborative Evaluation of Raw Ideas

Graham Horton, Jana Goers
Computer Science Department, Magdeburg University, Magdeburg, Germany.
Zephram GbR, Magdeburg, Germany.

Corresponding author email: graham.horton@ovgu.de

Abstract

We consider the task of evaluating raw ideas by a team of experts where typically a simple GO/NO-GO vote is taken. However, since both the ideas and the evaluation criterion can be ambiguous, the experts will in general form different mental models of them, which then become the basis for their individual evaluation judgements. This effect casts doubts on the meaning and reliability of the evaluation result.

We propose a model for raw ideas and a facilitation algorithm for their evaluation in a group. The algorithm is designed to uncover hidden profiles in the raw idea and in the evaluation criteria and to treat these profiles separately. Our goal is to generate better ideas and a more precise interpretation of the evaluation criterion. An additional feature is increased transparency of the evaluation, which improves the group's acceptance of the result.

Two small examples illustrate the behaviour of the algorithm.

1. Introduction

In the first phase of an innovation process, a (possibly large) number of ideas must be evaluated. This evaluation is performed by a group of experts, usually representing different points of view within the organization. The goal of this evaluation is to determine which of the ideas are promising and merit closer attention and which should be dropped.

Typically, the ideas to be evaluated are very sketchy, consisting of no more than a single phrase or a short sentence. For this reason they are commonly referred to as *raw ideas*. Furthermore, the evaluation criteria at this stage are usually deliberately left vague. Thus, both the raw ideas and the evaluation criteria can be ambiguous and are open to varying interpretations.

Owing to time constraints, the evaluation may be very simple: ideas receive a GO or NO-GO vote from each group member either after a very short and unstructured discussion, or even without any discussion at all.

As a result of this approach, individual group members may have different mental models of the raw ideas and the evaluation criteria, and information which is relevant to the evaluation may also only be known to a subset of the members of the group. Furthermore, these locally held mental models and items of information are not made available to the rest of the group before voting takes place.

This can lead to several significant problems. If, for example, a group of six experts has voted 4:2 in favour of an idea, did all six members actually have the same idea in mind? Which criterion was used to arrive at each individual judgement? What should the process owner do next? In the worst case, the evaluation will be incorrect, or the process owner will undertake inappropriate or irrelevant next steps. In addition, the group's acceptance of the evaluation result may be compromised, if some members feel that the evaluation is inaccurate.

In our experience from a large number of innovation consulting projects, we have observed these phenomena many times. Clearly, problems such these can have a significant negative business impact in the form of wasted effort, investment in bad ideas and lost opportunities. We have also occasionally been able to resurrect raw ideas that had been initially rejected by the group of experts by ensuring that certain interpretations held by individuals were made available to the group. In Section 5 we show an example of this. These experiences formed the motivation for the work presented in this paper.

Our contribution is motivated by previous work on hidden profiles which is summarized in the next Section. However, the task of evaluating raw ideas is sufficiently distinct from the group tasks that are typically considered in the literature to necessitate a modified approach.

In this paper, we present a model for raw ideas and for a facilitation algorithm based on this model. The algorithm elicits hidden profiles and generates a more comprehensive evaluation result that distinguishes between the various mental models held by the individual group members.

The immediate goals of this work are …

- to gain insight into the variety of the mental models that can arise in the collaborative evaluation of raw ideas,
- to improve the quality of ideas and the shared understanding of possible evaluation criteria,
- to obtain useful hints for post-evaluation management activities,
- to gain pointers for improving our facilitation of evaluation workshops.

Our long-term goal is to improve the overall efficiency and effectiveness of the early phases of the innovation process, in particular with new and better algorithms for information technology support.

In Section 2, we create a context for our work. In Section 3, we present our data model. Section 4 contains the algorithm for processing raw ideas in order to obtain a differentiated evaluation result. In Section 5, we present two small examples that illustrate the algorithm and its output. The paper ends with some implications for managers and an outlook.

2. Background

2.1. Raw ideas and the front end of innovation

The traditional (and almost ubiquitous) model for the innovation process in organizations is the so-called stage gate process [3], which is illustrated in Figure 1. The stages represent phases in which ideas are processed (for example checking their technical feasibility or conducting preliminary market research), and the gates represent evaluation and selection events. At each gate, a GO/NO-GO decision for each idea is made; NO-GO means that the idea will be dropped, GO allows the idea to proceed to the next stage. At Gate 0, raw ideas are examined, which contain very little information, and the evaluation criterion is usually relatively general. (By contrast, at later gates, the ideas will be very comprehensively described, and a substantial set of precisely defined criteria will be applied. The so-called *Front End of Innovation (FEoI)* represents the activities from raw idea generation up to the decision to initiate a development project.

In organizations that have such an innovation process, the process owner will be a person who is responsible for new products, services, or business models. Depending on the size and type of organization, this may be the CEO or a head of division, or a product, business development or innovation manager. In the following, we will assume that the process owner is an innovation manager. The function of each gate for the innovation manager is twofold: to learn which ideas have been accepted and rejected by the expert group, and to obtain suggestions for developing the ideas in the next stage. Examples of such suggestions are to activate promoters within the organization, determine technical feasibility or obtain clearance from management.

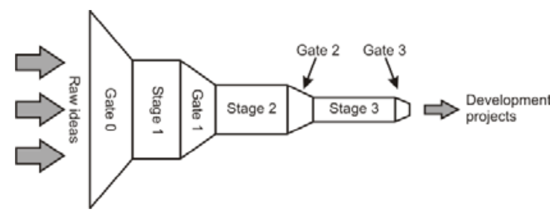

Figure 1. FEoI stage-gate process

At the front end of the innovation process it is important to avoid rejection errors, i.e. a NO-GO evaluation of an idea which would have been successful if it had been implemented. An idea that experiences a rejection error is lost forever. On the other hand, an acceptance error (a GO evaluation of an idea that later fails) is less serious, since it will probably be filtered out later when more information is available. A discussion of acceptance and rejection errors and their causes can be found in [18].

Girotra et al make the important observation that in business, it is the quality of the best idea, rather than the mean quality of a set of ideas – the metric that is usually used in academic studies – that counts [5]. The reasons for this are intuitively clear: no organization has the resources to implement a large set of ideas, and in a free market, uniqueness is a key success factor, since it makes new products distinctive and thus competitive. This observation is confirmed by our consulting practice: although hundreds of ideas may be produced, one single "hit" idea is itself enough to justify the investment in the idea generation and evaluation. The "open day" example in Section 5 illustrates how our algorithm yields one very good idea from an idea that initially had received a majority NO-GO evaluation.

2.2. Hidden profiles and mental models

Mental models are defined by Rouse and Morris as *internal representations of (aspects of) the environment that provide a conceptual framework for describing, explaining and predicting future system states* [13]. Mental models are initially unique to every individual. In the context of collaboration, the portion of each mental model of the task that is held by all members of the group is known as the *shared mental model* and is defined by Van den Bossche et al as *the overlapping mental representation of knowledge by members of a team* [19].

It seems intuitively clear, and it has been shown in many studies [19], that it is important to build a shared mental model of the task in order for a group to function efficiently. The shared mental model is built via the two processes of (co-)creation of meaning, i.e. the establishment of common interpretations, and constructive conflict, i.e. the competition between differing points of view leading to the best choice for the task [19]. Our algorithm promotes both of these processes.

When items of information affecting a group decision are held by only subsets of the members before discussion begins, i.e. they are not part of the shared mental model, we have what is referred to as a *hidden profile*. Figure 2 shows a traditional hidden profile in which three persons A, B and C have to choose between alternatives X and Y. Each of them uniquely knows a different item of information in favour of X, and all of them know two items of information in favour of Y. They thus have a shared mental model for Y, but X has a hidden profile.

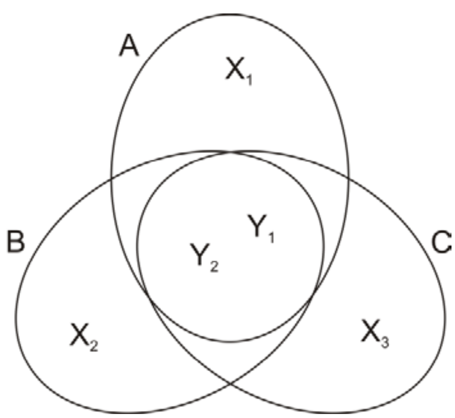

Figure 2. Classical hidden profile

Hidden profiles have been intensively studied [16], almost always in the context of group decision-making. When the task is to select the better of two alternatives, the assumption is made that the information favouring the superior alternative is unique, while the information favouring the inferior alternative is shared. Studies have shown that groups tend to converge on the shared information very quickly, and that the hidden profile is consequently never discovered. This is due to social effects such as groupthink, dominance of individuals, evaluation apprehension and free riding [15].

Milliken et al discuss the benefits of cognitive diversity, i.e. the differences in terms of what group members know or how they think about problems, in collaboration [8]. Clearly, cognitive diversity has significant advantages in creative tasks, including coming up with more innovative and higher quality ideas thanks to the wider range of perspectives that are available. However, the consideration of multiple alternatives depends on the group's willingness and ability to share unique information (i.e. their ability to solve the hidden profile problem). Evaluation apprehension and the desire for consensus can lead to premature convergence to the wrong conclusion. The solution proposed by Milliken et al is authentic dissent, which corresponds to the constructive conflict of Van den Bossche et al [19].

Stasser & Titus state [16] *the key to revealing hidden profiles seems simple: communicate unique information*. However, in our experience, this is seldom attempted at the raw idea stage, or well-known effects prevent it from functioning adequately. Thus, any algorithm that is designed to solve the hidden profile problem must promote the exchange of unique information (private mental models).

According to Tindale et al [17], a procedural approach holds the most promise for solving the hidden profile problem. In other words, a facilitator script that guides the group's discussion is most likely to be able to uncover hidden profiles. Stasser & Birchmeier also consider a procedural approach to be most effective [15]. They propose a specific Nominal Group Technique that contains both individual and group work, thus avoiding many of the pitfalls of (unfacilitated) group discussions. This is consistent with Girotra et al [5], who state, *We find strong support that the best ideas generated by a hybrid process are better than the best ideas generated by a group process.* The new algorithm presented in Section 4 is a hybrid procedural approach.

Stasser & Birchmeier claim that technology such as Group Decision Support Systems appears to alleviate many of the problems associated with hidden profiles [15]. However, a later meta-study by Lu et al [7] suggested that technology had no effect. We suspect that no specific facilitation procedures were used in the IT-based studies, which might account for the inconclusive result. However, lack of details prevents any firm conclusion being made. The same meta-study also showed that decision quality was positively correlated with the degree to which unique information was made explicit and the attention given to unique information by the group. Clearly, any hidden profile-resolving algorithm must exhibit these two properties.

To summarise: The hidden profile problem is well understood, but until now, only general suggestions for solutions have been published. Our approach is to maximize the well-known factors that are conducive to decision quality via a facilitation procedure that lends itself to computer support.

2.3. Other related work

Kempe et al have presented a computer-supported algorithm for the evaluation of raw ideas that allows the selection to be made in the presence of missing information [6]. The algorithm generates multiple evaluations based on the missing information and yields as output the tasks for the innovation manager that can provide that missing information. Similarly to the algorithm presented in Section 4, the Kempe algorithm is an example of division of labour between computers and humans in which each is assigned the tasks it can perform well.

Briggs et al have proposed six patterns of collaboration named *Generate*, *Reduce*, *Clarify*, *Organize*, *Evaluate*, and *Build Consensus* [2]. These patterns are claimed to be a canonical set of tasks that constitute any group process. They are usually considered to be distinct phases of a collaborative process (consider, for example, the widely used terms *convergent* and *divergent*, which refer to the Generation and Reduce patterns respectively). However, as Stasser and Birchmeier point out, *in practice divergent and convergent processes in collective choice are undoubtedly entwined in complex ways* [15]. In our algorithm, the Generate, Reduce, Clarify and Evaluate patterns are all interwoven and cannot be meaningfully separated.

The implications of this entwining of collaboration patterns are currently unclear to us. From the facilitator's point of view, keeping the patterns distinct makes collaboration processes easier to understand and to manage. On the other hand, in our case, it is the mixing itself that creates value, because it is the occurrence of a discrepancy in the evaluation that sets off a new round of clarification.

Finally, the embellishment and combination of ideas is important to the success of group ideation, and is one of the original brainstorming rules suggested by Osborn [10]. Our algorithm explicitly encourages these behaviours.

3. The hypothesis model

We now define our model for raw ideas, which on which the algorithm presented in Section 4 is based.

3.1. Definitions

We assume that the evaluation of raw ideas is based on three items of information in the mind of each group member: the raw idea itself, the evaluation criterion and possibly also supplementary information that affects their evaluation. These we refer to as *Proposal (p)*, *Criterion (c)* and *Support (s)*, respectively. Taken together, these three pieces of information form a group member's mental model of the evaluation task.

We refer to a vector (*p*, *c*, *s*) as a *Hypothesis*. The rationale behind this term is that it may be considered to be a tentative suggestion of how to achieve success. It is interpreted as follows: *Given that we want to achieve c and that we know s, we propose doing p*. This hypothesis is then tested in the next stage of the innovation process, which may therefore consistently be considered to be an experiment. The hypothesis metaphor is chosen in order to be consistent with the Lean Startup [1] [12] and Hypothesis-Driven Entrepreneurship [4] models for founding technological companies.

Using these terms we now illustrate our hidden profile situation. Instead of the distributed items of information being arguments for or against alternatives, they are differing interpretations of hypotheses. In the classical case, the task is to make the hidden profile explicit in order to synthesise arguments. Here, however, the unique mental models are themselves alternatives, of which some or all may be selected and treated as new individual entities.

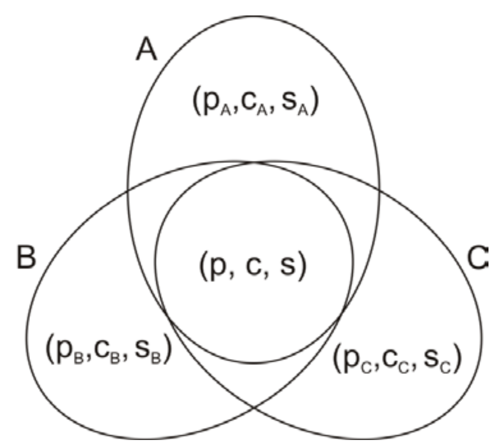

Figure 3. Hidden hypothesis profile

Figure 3 shows an example situation for a group of three individuals A, B, and C. The group has been presented with a hypothesis (p, c, s) for evaluation. In general, each group member will have a different model of the hypothesis, denoted by (p_A, c_A, s_A), (p_B, c_B, s_B) and (p_C, c_C, s_C) respectively. If no measures are taken, the group members will make their GO/NO-GO judgements about (p, c, s) based solely on their local information; if the models differ significantly, the voting outcome will be practically meaningless. The examples in Section 5 illustrate that this can indeed occur in practice.

For example, if the group is processing ideas for improving the customer-friendliness of a supermarket, they might be asked to evaluate the Proposal p = *Provide seating for our customers* with the generic Criterion c = *Is a good idea*. Group member A may form the mental model p_A = *Comfortable sofas with popular magazines*, c_A = *Attractive for senior citizens*, while group member B may form the mental model p_B = *Lounge chairs with a WiFi access point*, c_B = *Attractive for the younger generation*. (Supports s, s_A and s_B are assumed to be empty.) In this case, both A and B may well vote GO. However, they are voting for different proposals according to different criteria, which under normal circumstances is not apparent. Note that if we switch the criteria, so that A evaluates (p_A, c_B) and B evaluates (p_B, c_A) we may well obtain the opposite result, i.e. two NO-GO votes. Furthermore, even if both group members make their judgements based on (p_A, c_A), but only one of them possesses the support *Our company has a policy of not providing seating facilities for customers*, then the result will be one GO vote and one NO-GO vote.

We now consider the situation in which (at least) two group members evaluate a hypothesis differently; we call this result a *Discrepancy*. We distinguish between two different types of Discrepancy. In the first case, the group members have different mental models of the hypothesis; we refer to this case as a *Model Discrepancy*. In the second case, the subjects have reached different conclusions from the same mental models; we call this result an *Axiomatic Discrepancy*, because the difference in opinion stems from a difference in beliefs, rather than information.

3.2. Consequences and expectations

Our model recognizes two kinds of discrepancies that account for differing judgements by the group.

Model Discrepancies result from different mental models and can be resolved simply by eliciting all mental models and treating each of them as a new hypothesis to be evaluated.

Axiomatic Discrepancies cannot be resolved within the evaluation session, since they are caused by differences in beliefs. In this case, the best we can achieve is to make the Discrepancy visible. When Axiomatic Discrepancies result from dogmatically held beliefs, the innovation manager must simply accept that no unanimity can be reached. One the other hand, when the beliefs simply result from a temporary non-availability of information (for example A believes that the boss will approve of the proposal while B believes he/she will not), the innovation manager is given a clear indication of how to resolve the issue.

The primary consequence of our model is therefore that ultimately there are only two possible evaluation results for a hypothesis: either a unanimous judgement or a (non-unanimous) Axiomatic Discrepancy.

We expect that unanimous evaluations will achieve a very high acceptance within the group, since each group member will observe that all members have the same mental model of the hypothesis and have made the same judgement. We assume that the acceptance of Axiomatic Discrepancies is at least no worse than that of an undifferentiated non-unanimous evaluation.

4. The hypothesis processing algorithm

We now present our algorithm for group evaluation of raw ideas. It is based on the definitions and assumptions outlined in the previous Section.

4.1. Goals

The functional goals of the algorithm are:

- to identify the different mental models inspired by the original raw idea and to obtain separate evaluations for each of these,
- to identify any Axiomatic Discrepancies,
- to provide feedback to the innovation manager about next steps,
- to increase the scope of the evaluation criteria,
- to provide an opportunity for idea refinement.

The process goals of the algorithm are:

- to create transparency in the voting reasoning of the group members,
- to increase acceptance of the evaluation result,
- to increase unanimity in the evaluation,
- to benefit from diversity in the group without suffering from the problems arising from false dissent.

4.2. Notation

The algorithm presented in Section 4.3 makes use of the following variables:

- H: a list of hypotheses yet to be evaluated,
- p: the initial proposal,
- c: the initial criterion,
- s: the initial support (which may be empty),
- J: the set of judgements made by the members of the group,
- A: the set of axioms held by the group in the case of an Axiomatic Discrepancy,
- h: the hypothesis that is currently being evaluated,
- M: the set of mental models held by the group as a result of being exposed to a hypothesis h.

4.3. Algorithm

The group hypothesis processing algorithm is given in pseudo-code notation in Figure 4.

```
 1 H = {(p, c, s)}
 2 repeat
 3   h <- GetHypothesis(H)
 4   J <- ElicitJudgements(h)
 5   M <- ElicitAndSelectModels(h)
 6   if (SizeOf(M) = 1) then
 7     if (SizeOf(J) = 1) then
 8       Output(Unanimous(J), h)
 9     else
10       A <- GetAxioms(M)
11       Output(AxiomaticDiscrepancy, h, A)
12   else
13     UpdateHypotheses(M)
14 until Empty(H)
```

Figure 4. Algorithm in Pseudo-Code

The input to the algorithm is the hypothesis to be evaluated (the initial hypothesis). The output of the algorithm consists of the hypotheses generated and the type of voting result (unanimous or axiomatic discrepancy) for each.

There now follows a line-by-line explanation of the algorithm:

- Line 1: The set of hypotheses to be processed H is initialized with the initial hypothesis.
- Lines 2 & 14: Keep going until there are no more hypotheses left to process.
- Line 3: Select the next hypothesis from the list for processing.
- Line 4: Elicit the judgements.
- Line 5: Elicit the mental models.
- Line 6: Test if all models are the same.
- Line 7: Test if all judgements are the same.
- Line 8: Output the unanimous result.
- Line 9: Enter this branch if judgements are not unanimous.
- Line 10: Elicit conflicting axioms.
- Line 11: Output the result Axiomatic Discrepancy.
- Line 12: Enter this branch if there is more than one model in play.
- Line 13: Enter the new hypotheses into the list based on the mental models of Line 5.
- Line 14: Test if the set of hypotheses is empty.

Figure 5 shows the same algorithm in terms of possible facilitator instructions. The line numbers correspond to those in Figure 4.

```
 1  <empty>
 2  <begin algorithm>
 3  "I am now removing the next hypothesis
from the list for you to work on. <Read out
the hypothesis.>"
 4  "Now evaluate this hypothesis with 'GO'
or 'NO-GO'"
 5a "Now write down your interpretations of
the proposal, the evaluation criteria you
used, and any supporting knowledge you might
have used to make your judgement."
 5b "Now each of you in turn read out your
interpretations."
 5c "Now please select from the
interpretations you have just heard those
that you would like to be treated as new
hypotheses."
 6  <If all models are essentially
equivalent then carry out 7-11.>
 7  <If all judgements are identical then
carry out 8, otherwise carry out 10 and 11.>
 8  "I am now adding your unanimous
interpretations and judgements to the
output." <Continue at 13.>
 9  <empty>
10  "Although you all have the same mental
model of the hypothesis, you have still
given differing judgements. Please write
down what belief or opinion led to your
judgement."
11  "I am now adding your results to the
output as an Axiomatic Discrepancy."
12  <empty>
13  "I am now adding the new hypotheses you
selected in step 5c to the list."
14  <If there is at least one hypothesis
left in the list return to 2.>
```

Figure 5. Algorithm as facilitated procedure

4.4. Discussion

The algorithm terminates when there are no more hypotheses left to be processed. In principle, the algorithm could never terminate, if the group members continually developed more and more refined models. In practice, however, we have never observed more than three levels of refinement.

As the depth increases, the Proposals tend to become less and less abstract, and the (shared) number of criteria and the volume of support grows.

In practice, one or more concrete evaluation criteria may be supplied. These can be input into the algorithm (as the value of c in Line 1). The algorithm itself is insensitive to the choice of starting criterion.

The algorithm supports the successive refinement and expansion of the initial evaluation criterion. Each mental model of the evaluation criterion that is discovered expands the set of criteria shared by the group. This is advantageous, because the goal of the first evaluation stage is to identify all raw ideas that are worthy of further consideration, and a raw idea may qualify because of any of a number of different criteria. For example, an idea for a new product may be worthy of further investigation because it adds customer value, complements an existing product, fills a gap in the portfolio, or has been requested by an important customer.

In the extreme case, the evaluation could be started with a generic criterion such as *is a good idea for us*, or even with none at all, and the group will generate a set of relevant criteria from this on its own. This extreme case may be permissible if the group is diverse and contains members from all relevant fields (such as Management, Marketing, Engineering and Sales for a new product idea). The examples in the next Section illustrate the variety of criteria that a group can generate autonomously from a generic initial criterion.

In order to accelerate execution, the algorithm can easily be reorganised so that line 5 is only executed if judgements differ. This saves a considerable amount of time at the price of not eliciting hidden profiles when judgements are unanimous.

5. Examples

In this Section, we show two small examples to illustrate the behavior and output of the algorithm.

5.1. Case 1: Innovation lab project

The first example is taken from an innovation lab which is run jointly by a university and a consulting company. Past activities in the lab include innovation consulting and startup coaching. The ideation task was *What new projects should our lab carry out?* and the initial hypothesis to be evaluated was (p = *We should participate in government-sponsored projects*, c = *is good for us*, *s*= *<empty>*). The participants in the evaluation were the CEO of the consultancy, the Chief Scientist of the lab, two employees and two student interns.

The six mental models of the initial proposal were (explanatory comments are given in square brackets):

1. *Provide support for medical technology startups coming out of the university.* [Medical technology is one of the university's key research areas.]
2. *Do core competence training for teachers.* [The lab has carried out several such trainings in the past, albeit for other target groups.]
3. *Participation in government projects.*
4. *Compete for calls for project proposals.*
5. *(Re-)position the lab as a partner for government projects.*
6. *Participate in DFG research projects.* [DFG (Deutsche Forschungsgemeinschaft) is the German funding agency for basic research, similar to the National Science Foundation in the USA.]

Models 1, 2 and 6 are concrete suggestions that are also very different, model 5 is more of a strategic proposal and models 3 and 4 are hardly more specific than the original proposal.

A total of eleven mental models of the initial evaluation criterion were generated:

1. *Improve the reputation of the lab*.
2. *Create value for the local economy*.
3. *Strengthen the relationship between the company and the university*.
4. *Stabilise the lab's income*.
5. *Increase the visibility of the lab*.
6. *Keep the risk of project failure low*.
7. *Do cool projects*.
8. *Autonomy*.
9. *Projects must be compatible with our image*.
10. *Do not expend an inappropriate amount of effort*.
11. *Projects must be compatible with our core competencies*.

We find it encouraging that even after just one evaluation step the group had already generated such a varied and useful set of evaluation criteria.

Proposal #1 was immediately voted 6:0 GO. Proposal #3 was initially voted 1:5 GO:NO-GO, it then split into two variants, one of which was voted 0:6 NO-GO, the other yielded an Axiomatic

Discrepancy centred around the question whether this type of project could be carried out profitably or not.

The maximum level of refinement that the group needed was three.

The algorithm also yielded actionable results, including *Make contact with the speaker of the Medical Technology project at the university* and *Determine the fees that can be charged for training seminars given to the government*.

5.2. Case 2: Open day

The second example arose from the annual "Night of Science", an open day in which members of the public can visit the colleges and research centres in the city from 6pm until midnight. A group of six students was given the task of generating ideas for the University Computer Science department that would attract visitors into their building. The proposal to be evaluated was *We should project a large image onto the exterior wall of the Computer Science building*, and the initial criterion was simply *Is a good idea*. Again, the criterion was deliberately vague, in order to be able to observe the mental models that the group members generated of their own accord.

Figure 6 illustrates the development of the hypotheses generated by the six members of the group. A check mark denote a GO vote and an X a NO-GO vote. The initial hypothesis 1 generated a model discrepancy with five distinct mental models of the proposal (proposals 1.2 and 1.5 were deemed to be essentially identical.) Hypothesis 1.1 immediately received a unanimous GO vote using participant #1's criterion, while hypothesis 1.4 was immediately and unanimously rejected. Hypotheses 1.2, 1.3 and 1.6 generated model discrepancies, which at the next level of refinement resulted in two axiomatic discrepancies (hypotheses 1.2.1 and 1.6.1) and a unanimous GO vote (hypothesis 1.3.1).

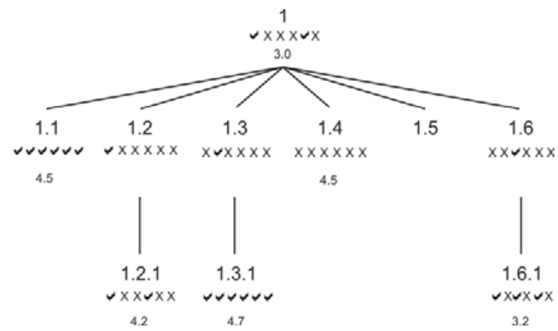

Figure 6. Development of hypotheses

The small numbers below the evaluation protocol in Figure 6 show the mean acceptance of the evaluation based on the Likert question *To what extent do you accept the group's evaluation of this hypothesis?* Possible answers ranged from 1 *Not at all* to 5 *Completely*. The mean values for the six group members are shown. The original hypothesis, which received a 4:2 vote, had a mean level of acceptance of 3.0. It can be seen that unanimous results receive a high level of acceptance (4.5, 4.7 and 4.5), whereas the two axiomatic discrepancies generated had a mixed improvement in the level of acceptance (4.2 and 3.2). Of course, these results are insufficient to draw any firm conclusions, but they do suggest that the algorithm is able to raise the level of acceptance by the group.

Hypothesis #	Proposal	Criteria (C) or Axioms (A)	Result
1.1	A photo collage about research and life at the department	C: Illustrates what goes on in the department	6:0 GO
1.2.1	An image film about the department	A: Image films are always boring; A: If edited well, an image film can be interesting	2:4 AD
1.3.1	A giant department logo projected into the night sky	C: Cool C: Visible from far away C: My first thought: Batman!	6:0 GO
1.4	A large abstract image or entertaining YouTube videos	C: Has nothing to do with Science. (Anyone could do this, you don't have to be a university.)	0:6 NO-GO
1.6.1	A 3D projection	A: Is technically feasible A: Is not technically feasible	3:3 AD

AD = Axiomatic Discrepancy

Figure 7. Output of algorithm (Case 2: Open day)

Figure 7 shows the output of the algorithm. The numbers of the hypotheses correspond to those in Figure 6. Hypothesis 1.5 is omitted because it was essentially identical to hypothesis 1.2. In order to keep the table legible, we have omitted the support generated by the algorithm. Some individual examples of support are given at the end of the Section.

Hypothesis 1.2.1 reached an axiomatic discrepancy: two participants believed that no image film could be interesting, while others believed that an appropriately edited one could be. This difference could not be resolved. If the process owner decided to pursue this proposal, he/she would know what question needs to be answered by further investigation.

Hypothesis 1.3.1 is the most interesting case. Participant #2 had the model (*Project a still image of the department logo onto the wall of the building*, *Is boring*, .) and had voted NO-GO. In the second round, another participant modified the hypothesis to (*Project the logo into the sky*, *Is cool*, .), which all others accepted in round 3 and voted unanimously GO. Ultimately, this proposal turned out to be the overall winner.

Hypothesis 1.4 was originally voted NO-GO by participant #4, no other participant offered an alternative model, and the hypothesis was rejected unanimously.

Hypothesis 1.6.1 resulted in an axiomatic discrepancy, as participants had different (and irreconcilable) opinions as to the technical feasibility of creating and projecting a 3D film. Again, the process owner is given a clear indication of what needs to be researched.

Some examples of support that the algorithm produced are:

- *X is already planning to do that.*
- *I have seen something similar, and I was very impressed.*
- *We have the necessary equipment and expertise to build that.*
- *Our lecture theatre has a capacity of 120 (which is enough).*
- *X tried something similar last year, and it worked very well.*
- *It doesn't get dark until about 10pm at that time of year.*

Making items of information such as these part of the shared mental model of the group can often lead directly to unanimity.

6. Conclusion

6.1. Summary

We have proposed a data model and an algorithm for the evaluation of raw ideas by a group. The most important implication of the model is that there are only two possible causes for disagreements in the evaluation judgements, one of which can always be resolved to a unanimous vote.

The evaluation algorithm proceeds by eliciting the mental models of each group member and then processing each of these as a new hypothesis to be evaluated. In this manner, variants and improvements in the original raw idea can be found and the evaluation criterion refined.

Two very small examples are used to illustrate the algorithm, and the results obtained suggest that the acceptance of the result by the group is improved. We speculate that this is due to the increased transparency and the greater degree of differentiation of the evaluation process.

6.2. Implications for managers

The hidden profiles of hypotheses mask the variety of mental models and criteria that are generated by an evaluation team. The investment in uncovering these models can be worthwhile, especially considering that the quality of the best idea is the correct performance indicator.

In the evaluation of raw ideas, the criterion and any supplementary information are just as important as the proposal itself.

Given the opportunity, the evaluation team will come up with an appropriate set of evaluation criteria on its own.

We recommend distinguishing between model discrepancies and axiomatic discrepancies and making these explicit to the evaluation team. This can increase the acceptance of the evaluation result.

Axiomatic discrepancies made explicit can be indicators to next steps in the innovation process.

The algorithm proposed can be used in both a stage-gate-style innovation process and also in a hypothesis-driven process such as Lean Startup. In the latter case, the algorithm should re-framed from a decision gate to an experimental hypothesis.

6.3. Outlook

The examples given in this paper serve only to illustrate the behavior of the method. Formal studies are still needed in order to draw robust conclusions. These will include larger samples and additional test cases. Several effects need to be tested rigorously, including the creation of acceptance of the evaluation result, both in the cases of unanimity and of Axiomatic Discrepancies.

We also need more practical experience with the algorithm in order to learn to what depth groups typically go and whether the cognitive load needed to switch between similar hypotheses is acceptable. Since the algorithm can be quite time-consuming, we are also looking for heuristics to accelerate and simplify its implementation.

Our next practical steps will include creating a software prototype for facilitating the evaluation algorithm and to determine to what extent it simplifies and accelerates the execution of the algorithm compared to a human-only facilitation. (Facilitating the algorithm is quite challenging.) This in turn will lead to questions regarding appropriate visualisation methods and interaction techniques.

The major practical difficulty in facilitating the algorithm is deciding which (if any) of the set of hypotheses stated by the members of the group at each stage are to be dismissed, kept or amalgamated with others (and, in the latter case, using what formulation). In the experiments described in this paper, this was an ad hoc group decision directed by the facilitator; however, we feel that a more structured approach is needed.

The popular Lean Startup approach to founding technology companies favours a rapid cycle of developing and testing ideas. We plan to extend our algorithm in order to be compatible with the wider context of a Lean Startup project. This will necessitate replacing the simple GO/NO-GO evaluation task by conditional predicates such as "GO, if ..." and "NO-GO, unless …".

The algorithm presented in this paper mixes different patterns of collaboration, whereas usually, a facilitator would prefer to keep these separate. Studying the effects of pattern mixing may yield some interesting new research questions.

Finally, of course, we will integrate the recommendations contained in Subsection 6.2 into our own consulting practice.

7. References

[1] S. Blank, 2013, Why the Lean Startup Changes Everything, Harvard Business Review, http://hbr.org/2013/05/why-the-lean-start-up-changes-everything/

[2] R. Briggs, G. Kolfschoten, G. de Vreede, and D. Dean, 2006, "Defining Key Concepts for Collaboration Engineering," in Americas Conference on Information Systems, Acapulco, Mexico.

[3] R. Cooper, S. Edgett, E. Kleinschmidt, 2001, "Portfolio Management for New Products", Perseus Books, New York, 2001.

[4] T. Eisenmann, E. Ries, S. Dillard, 2012, "Hypothesis-Driven Entrepreneurship: The Lean Startup", Harvard Business School Entrepreneurial Management Case No. 812-095, http://ssrn.com/abstract=2037237

[5] K. Girotra, C. Terwiesch, K. Ulrich, 2010, "Idea Generation and the Quality of the Best Idea", Management Science, 56:591-605.

[6] N. Kempe, R. Buchholz, J. Goers, G. Horton, 2012, "An Optimal Algorithm for Raw Idea Selection under Uncertainty", 45th Hawaii International Conference on System Sciences, Maui, HI.

[7] L. Lu, Y. Luan, P. McLeod, 2012, "Twenty-Five Years of Hidden Profiles in Group Decision Making: A Meta-Analysis", Personality and Social Psychology Review 16(1) 54-75.

[8] F. Milliken, C. Bartel, T. Kurtzberg, 2003, "Diversity and Creativity in Work Groups", in [Paulus&Nijstad]

[9] C. Nemeth, B. Nemeth-Brown, 2003, "Better than Individuals? The Potential Benefits of Dissent and Diversity for Group Creativity", in [Paulus&Nijstad]

[10] A. Osborn, 1953, "Applied Imagination", Scribner.

[11] P. Paulus and B. Nijstad (eds.), 2003, "Group Creativity: Innovation through Collaboration". Oxford University Press, New York.

[12] E. Ries, The Lean Startup, Penguin, 2011.

[13] W. Rouse, N. Morris, 1986, "On looking into the black box: Prospects and limits in the search for mental models", Psychological Bulletin, 100, 359–363.

[14] S. Schulz-Hardt, F. Brodbeck, A. Mojzisch, R. Kerschreiter, D. Frey, 2006, "Group Decision Making in Hidden Profile Situations: Dissent as a Facilitator for Decision Quality", Journal of Personality and Social Psychology, Vol 91, No. 6, 1080-1093.

[15] G. Stasser, Z. Birchmeier, "Group Creativity and Collective Choice", in [Paulus&Nijstad]

[16] G. Stasser, W. Titus, 2003, "Hidden Profiles: A Brief History", Psychological Inquiry, Vol. 14, No. 3&4, 304-313.

[17] R. Tindale, T. Kameda, V. Hinsz, 2003, "Group Decision Making" in M. A. Hogg & J. Cooper (Eds.) Sage handbook of social psychology, Sage Publications, Inc., London.

[18] J. Tropman, 1998, "The Management of Ideas in the Creating Organization", Greenwood Publishing Group.

[19] P. Van den Bossche, W. Gijselaers, M. Segers, G. Woltjer, P. Kirschner, 2011, "Team Learning: Building Shared Mental Models", Instr. Sci. 39:283-301, DOI 10.1007/s11251-010-9128-3.