How Do You Search for the Best Alternative?
Experimental Evidence on Search Strategies to Solve Complex Problems.

Abstract

Through a controlled two-stage experiment, we explore the performance of solution search strategies to resolve problems of varying complexity. We validate theoretical results that collaborative group structures may search more effectively in problems of low complexity, but are outperformed by nominal structures at higher complexity levels. We call into question the dominance of the nominal group technique. Further close examination of search strategies reveals important insights: the number of generated solutions, a typical proxy for good problem-solving performance, does not consistently drive performance benefits across different levels of problem complexity. The average distance of search steps, and the problem space coverage play also critical roles. Moreover, their effect is contingent on complexity: a wider variety of solutions is helpful only in complex problems. Overall, we caution management about the limitations of generic, albeit common rules-of-thumb such as "generate as many ideas as possible".

Keywords: complexity, problem solving, experiment, ideation, innovation

1. Introduction

Problem solving lies at the heart of innovation tasks. For that reason, the identification of a good solution to a problem has long been a subject of study in various fields, including operations research, social psychology, and innovation management. Interestingly, all fields have adopted a common conceptual lens to understand the challenges and effectiveness of problem solving: they view problem solving as a search for solution ideas. Yet, these disciplines differ as to how they operationalized the intended search to happen.

For example, the field of operations research (OR) focuses largely on the development of efficient computational methods to search for a good solution in a complex solution landscape. Metaheuristics such as simulated annealing, tabu search and GRASP allow for an extensive search of a large and complex landscape to identify a good point in the solution landscape, rather than a local optimum (see for example Blum and Roli 2003). However, OR methods assume that the performance function – albeit complex – can be specified up front.

The management literature "models" search as a boundedly rational approach to identify new alternatives within a complex solution landscape; solvers come up with new solution ideas by recombining known decision variables (e.g., Levinthal 1997, Rivkin 2000). Unlike the aforementioned metaheuristics, which carry a sense of constrained optimization, the management literature incorporates the bounded rationality of individuals and questions their ability to optimize. It typically assumes that the performance of identified solutions can be assessed, and that search steps usually make a change to just one, *randomly* chosen decision variable (local search), with larger changes (distant search) taking place only if local search does not yield the desired results. Such search behavior is empirically documented by Billinger et al. (2014), who observed individuals switching from local search to far jumps based on feedback provided.

Studies in the innovation and product development literature also consider search as a representation of the innovation process. They model search as draws from a solution space, but explore dynamic optimization decisions about which part of the solution space to search (Weitzman 1979, Loch et al. 2001, Erat and Kavadias 2008, and Erat and Krishnan 2012), as opposed to boundedly rational changes considered in the management literature discussed above.

Finally, in the social psychology literature Osborn's early work on brainstorming (1953) has triggered a lot of attention to search for good solutions in unstructured and open-ended problem statements,[1] which exhibit ambiguity as to the performance of the potential solutions, or to the possible decision variables. The social psychology literature adds an interesting angle to the overarching discussion on search, in that it explores the benefits of searching for solutions through *different types of groups*. It compares two archetypical group structures: groups that work collaboratively towards the solution to a problem, versus groups where the same number of individuals work in isolation (nominal group structure). One of the commonly cited search benefits of collaborative work in these activities is their ability to integrate a range of perspectives into the search process. Different perspectives allow a problem to be viewed through a variety of vantage points, and allow a variety of methods to be considered and combined in ways that would not be otherwise possible (Singh and Fleming 2010; Chan et al. 2017). In that sense, collaborative structures enable search outcomes that could not have been possible through individual perspectives alone. The presence of multiple individuals enhances critical examinations of potential solutions, as well as re-examinations of the underlying assumptions made regarding the problem context (solution landscape). This could prove particularly valuable in ambiguous problem contexts, where not only the magnitude of cause and effect relationships is uncertain, but also the underlying nature of cause and effect connections is unknown (Camerer and Weber 1992; Pich et al. 2002).

However, contemporary work questions the superior ability of collaborative structures, relative to nominal ones, to reliably and consistently find the best solutions (Stroebe and Diehl 1994; Paulus et al. 1996; Girotra et al. 2010). Not all the dynamics that emerge from group settings are positive. Indeed, in many contexts, the group dynamics create inefficiencies (e.g. production blocking), that limit the number of searches and diminish or even outstrip the potential benefits of complementary knowledge and perspectives. In settings where ambiguous problems are paired with increasing levels of problem

[1] For example: "suggest solutions to improve the relationship between the German population and the (foreign) guest workers" (Diehl and Stroebe 1987) or "generate ideas about the practical benefits or difficulties that would arise if everyone had an extra thumb on each hand after next year" (Gallupe et al. 1991).

complexity,[2] the collaborative structures can be particularly hindered by these negative dynamics. Kavadias and Sommer (2009) develop a theoretical model for brainstorming based on the conceptual foundations of search to find that the effectiveness of collaborative structures in solving highly complex problems may be markedly inferior to that of the same number of individuals working independently towards a solution ("nominal" structures). They show that the role of complexity is so fundamental, that in its absence, groups who can build on the best solution available, perform better than individuals.

In this paper, we test whether the theoretical predictions of the effects of problem complexity on the relative search performance of groups can be empirically validated in a controlled setting. Our experimental setting is adapted from an earlier innovation study by Ederer and Manso (2013), to accommodate the extant literature on search in solution landscapes. Such a setting allows us to compare the search performance of different experimental subjects for good solutions, without necessarily having these participants know the solution performance, and while they operate on solution landscapes of varying complexity. While the solution dimensions and the overall problem objective (profit maximization) are known, the link between the exact solution parameters (i.e. actual values assigned to different solution dimensions) and the exact solution performance is unknown to the experiment participants. The full solution landscape is only known to the researchers, and hence cannot perfectly inform participants regarding the direction of search for a good solution. Although the problem is less ambiguous than the experimental settings sometimes assumed in the broader brainstorming literature, it is ambiguous enough in the eyes of the participants to embody challenges faced during problem solving in corporations. In that regard, our study contributes to the search literature that has emerged in innovation management, by an explicit measurement of search effectiveness in a realistic setting. At the same time, it bears managerial implications to the corporate situations that engage brainstorming approaches to address company challenges.

We divide our theory development, and the subsequent analysis, into sections. In the first two sections, we consider search performance differences that various levels of problem complexity have within alternate group settings. Our results on the performance of nominal versus collaborative structures validate normative predictions (Kavadias and Sommer 2009), and echo general observations of the broader social psychology literature about the limitations of groups (e.g., Stroebe and Diehl 1994; Paulus et al. 1996). Following this, we consider the use of explicitly sequenced collaborative approaches, i.e. the well-established "nominal group technique" (NGT) and its converse. Our findings are in contrast to a strongly held belief regarding the way group collaboration can be leveraged in complex problem solving in firms (Sutton and Hargadon 1996). In particular, under high complexity we observe relatively low performance levels from the NGT recommended in organizations' literature (Robbins and Judge 2007).

---

[2] Complexity is conceptualized as the amount of interactions between the factors that determine the quality (performance) of a solution to the problem (see Simon 1969, p. 195).

Accordingly, we attempt to drill deeper into possible explanations for this deviation. We consider potential search strategies (patterns of search decisions) used by individuals in nominal or collaborative structures, with or without prior experience. We analyze these patterns as possible antecedents to the performance differences observed in collaborative and nominal settings, while we manipulate the levels of complexity. We find that the *number of solutions*, the *solution space coverage*, i.e., the breadth of the solution space examined during a search (Kornish and Ulrich 2011), and the *search step size*, i.e., the extent to which search attempts differ from one another, are informative predictors of performance. However, the magnitude, and sign of their impact depends critically on the level of problem complexity. Through these additional analyses, we are able to contribute beyond the theoretical developments of Kavadias and Sommer (2009). We highlight the instrumental role that the problem complexity carries in determining the effectiveness of problem solving search strategies. At the same token, we open up new research avenues as to the underlying mechanisms that tie distinct search strategies to problem solving effectiveness (Girotra et al. 2010), and encourage future research efforts to delve into these. We conclude with specific prescriptions to groups faced with increasingly complex and ambiguous management tasks.

2. Theory and Hypotheses

2.1 Impact of Problem Complexity on Group Search

A broad spectrum of literature identifies complexity as an important factor that constrains any organization's ability to search for and to identify the best solution to a problem (Siggelkow and Rivkin 2005). Managers can assess relative levels of complexity, and therefore a framework that allows them to choose the proper organizational approach, including group structure and sequencing, for the right type of problem task becomes of direct value for the firm.

As the complexity of a problem increases, so does the number of locally optimal solutions in the respective solution space (Kauffman 1993, Rivkin 2000). Kornish and Ulrich (2011) confirmed empirically that for more complex problems, there exists a substantial number of relatively good solutions in the set of all feasible solutions. Yet, with increasing complexity finding one locally optimal solution becomes less valuable, since the number of low performing solutions increases as well; in fact, a normative model of complexity establishes that the average solution performance of local peaks might actually decrease (Rivkin 2000). At the same time, higher complexity makes it increasingly difficult for individuals to effectively enrich their mental mappings of the solution landscape through past solutions; complexity makes causal inferences less robust (Pich et al. 2002). Hence, complexity has a negative impact on the individuals' ability to identify high performing solutions. We anticipate observing the following relationship empirically:

H1: Problem complexity decreases the value of the achieved search outcome in any group structure.

2.2 Group Structure and Problem Complexity

Given the deteriorating effect of complexity on search performance, we explore the search performance effects that stem from the interaction between the group structure choice and problem complexity. We focus on two distinct, yet commonly assumed group structures: a *collaborative* and a *nominal* structure. In contrast to groups working in a collaborative structure (where individuals interact directly), a nominal group structure consists of individuals working independently towards solutions to the same problem. The best solution developed among the members in a nominal structure becomes the solution implemented. The performance of that solution serves as a proxy for the overall performance of the entire group of individuals working in this nominal structure.[3]

In this instance, we focus on general task contexts where the search for solutions to problems takes place, and groups cannot perfectly describe the performance function in advance. In other words, individuals cannot "learn" a mathematical or logical tool to better or quicken their search process for solutions, as it is assumed in the problem-solving literature (Heller et al. 1992; Laughlin et al. 2006). At the same time, the search is also not assumed to be just a random walk over the solution landscape. Through some (possibly random) initial choices of solutions, individuals or collaborative structures build meaningful, yet imperfect, mental models that enable a gradual progress towards better solutions through more conscious choices. As a result, progress towards good solutions is made, an assumption underlying the local search (or adaptation) in the NK-model literature (Levinthal 1997), an assumption also made by Kavadias and Sommer (2009) in the context of idea generation.

The ability to quickly progress towards one good solution represents the potential advantage of collaborative structures: such groups can pool distinct perspectives, and exploit a variety of mental models to build on the 'running best' solution, while having the luxury of choice among several solutions (Kavadias and Sommer 2009; Sting et al. 2016). Indeed, building on disparate perspectives can be seen as virtuous in group settings (Osborn 1953, IDEO's deep dive). Groups that can do so, unconstrained by other dynamics, can pinpoint solution performance peaks within the solution landscape faster than individuals in nominal structures. Thus, collaborative structures might have an advantage in time-constrained settings.

However, complexity also constrains the search performance in collaborative group structures. As discussed above, complexity decreases the value of finding **one** locally optimal solution. In contrast, parallelism based search strategies become more promising as complexity increases (Sommer and Loch

[3] The ability to identify the best solution is clearly an assumption implicitly made in the social psychology literature, as the researchers focus solely on the idea generation process (see for example, Gallupe et al. 1991, Paulus and Dzindolet 1993). Follow-up work shows that idea generation and evaluation are not independent. Harvey and Kou (2013, pg. 347) look at different evaluation processes during the idea generation stage and demonstrate that "*a group's ability to select a final set of creative ideas therefore cannot be isolated from the process of forming their evaluations*". Knudsen and Levinthal (2007) show the importance of considering the joint evaluation and ideation process in a simulation model. In an evaluation process that falls into Harvey and Kou's category of "Evaluation as Convergent Decision Making", Girotra et al. (2010) find that in a comparison of collaborative structures and groups using the nominal group technique (NGT), those using NGT are better at identifying good ideas.

2004), especially if time is a limiting factor (Loch et al. 2001). Identifying several possible solutions is increasingly important, even if the identified solutions are not local optima. Nominal group structures achieve such strategies better because they are not subject to intra-group constraining factors which limit the ability and willingness to voice solution ideas in collaborative settings (Paulus and Dzindolet 1993, Diehl and Stroebe 1987).[4] In addition, individuals in nominal structures are not influenced by other group members' solutions. Such group effects can constrain the search path (e.g., collaborative topic fixation, Sawyer 2007, Kohn and Smith 2011) or can lead to adverse groupthink outcomes and search stagnancy (Bendoly 2014), reducing again the number of explored solutions.

Thus, we hypothesize the following relationship:

H2a: In problems with low complexity, individuals in collaborative group structures achieve better search outcomes than those in nominal group structures.

H2b: In problems with high complexity, individuals in collaborative group structures achieve worse search outcomes than those in nominal group structures.

2.3 Nominal Group Technique and Problem Complexity

In addition to the distinction between collaborative and nominal structures, the literature formally recommends the "*nominal group technique*" (NGT), which has been described as combining the best of both worlds (Delbecq, and Van de Ven 1971, Robbins and Judge 2007, Girotra et al. 2010). In NGT, problem-solving benefits from the following sequence of activity: initially, the facilitator provides sufficient time for individuals to think and search for solutions to the problem at hand by themselves (i.e. as if they were working in a nominal group structure). Following this initial stage, the facilitator convenes the individuals into a collaborative setting. Armed with their solutions derived thus far, s/he allows them to further combine existing solutions or perform additional search for newer ones to ameliorate the already found solutions. The theoretical rationale behind this mechanism is that the nominal stage forces independent thought to develop, which subsequently serves as a catalyst for informed and productive engagement during the collaborative phase.

This technique has been widely accepted as the preferred ideation mechanism in practice. In an experimental setting, Girotra et al. (2010) find that NGT outperforms groups who are given the same total time but collaborate throughout. Given the strong endorsements of this approach in the organizational literature, we proceed with hypotheses that aim to test whether the relative benefit from the NGT can be validated within a controlled problem-solving environment like the one we are using in our study.

[4] This experimental literature repeatedly identified 3 factors constraining collaborative brainstorming: *free riding*, *evaluation apprehension* (or being reluctant to participate out of an individual's fear of being negatively evaluated), and *production blocking* (or not being able to generate ideas while listening to others).

H3a: For low complexity, the NGT provides outcome improvements relative to searches that involve collaborative structures alone.

In order to further assess the impact of problem complexity, we capitalize on our prior arguments to explore it. Collaborative structures need to be able to create a shared mental model, to effectively transfer and build on solutions in the nominal work stage. However, as the complexity increases, individuals gain less of an understanding about the solution space (Pich et al. 2002). Indeed, due to the proliferation of performance peaks and valleys (Rivkin 2000), a high starting point proposed by an individual might not be a better starting point than a random starting point for finding the highest performing peak. Hence, heightened complexity might diminish any performance advantage of NGT.

H3b: As the complexity increases, there is diminishment in the performance advantage of the NGT over a search involving a collaborative structure alone.

The lack of mental models precludes firms from fully benefiting from the NGT in case of high complexity. The underlying reason for that is the proliferation of peaks and the fact that individuals may get stuck on local optima without much search.[5] One way firms can improve their search performance in complex solution landscapes is to base their initial nominal search on a lower dimensional representation of the overall landscape (Gavetti and Levinthal 2000), or by ignoring (temporarily) certain dimensions of the landscape (Ethiraj and Levinthal 2009, Csaszar and Levinthal, 2016). In our context, the better understanding of the partial solution landscape could enable an additional advantage: it makes the communication between group members working in a collaborative structure more efficient. Clearly, any constrained individual search makes the formation of a shared mental model even more difficult, since the group members now experience different parts of the landscape. However, the explicit origin of their non-overlapping experiences, makes the group members more likely to defer to others; almost like a micro-culture of "listening to the expert" within the group. Such behaviors reduce the risk of group conflict and encourage more effective value-added communication. It is within high complexity settings that such ability to effectively communicate and share good solutions becomes particularly important. Moreover, it may be difficult to achieve communication effectiveness without this constrained setting, since individuals have seemingly conflicting experiences in the same action space, i.e. spatially close solutions can have widely different performance values. Therefore, we hypothesize that in case of high complexity the efficiency of NGT can be improved by effectively constraining the search of individuals over the solution landscape. Such a constrained search can be achieved, if project managers ensure that parallel searches in the nominal setting explore clearly different approaches; and it develops naturally in cross-functional teams, where each individual contributes solutions depending on their personal expertise.

[5] Indeed, depicting the search paths of individuals during our experiment provides visual evidence that individuals get stuck in local optima, albeit sometimes only temporarily (see Appendix C).

H3c: For higher levels of complexity, the performance outcome of NGT is weakly improved, if individuals limit their search to a partial solution landscape (constrained search setting).

Finally, a full consideration of the virtues and limitations of NGT requires the consideration of sequences of engagement, such as that which is achieved when the sequencing of nominal and collaborative work is reversed ("*reverse nominal group technique*", or reverse-NGT). To our knowledge, the analysis and comparison of such a procedural structure is absent from the ideation, brainstorming and general problem-solving literatures. At the outset, during a reverse-NGT, the switch from collaborative structures to nominal structures should allow individuals to form a more coherent understanding of the solution space, through more aligned mental models, and perhaps a better starting point for their individual searches. Hence, the time allocated for nominal search should be more effective when this nominal search happens in the 2nd stage as opposed to the 1st stage of combined search efforts.

However, as we've already argued in H3b, individuals learn less about the solution landscape in cases of higher complexity, and the starting point found in the group setting, might not be closer to the overall optimum than a random starting point. Hence, the advantage of the reverse-NGT should also be less pronounced under high levels of complexity. Furthermore, due to fixation and framing effects resulting from the collaborative stage (Sawyer 2007, Kohn and Smith 2011), all individuals might search a similar part of the solution landscape, thereby reducing the overall performance of the search in the nominal structures in the 2nd stage even further.

H4a: For low complexity, the reverse- NGT provides outcome improvements relative to a search involving a nominal structure alone.

H4b: As the complexity increases, there is diminishment in the performance advantage of the reverse-NGT over a search involving a nominal structure alone.

2.4. The Impact of Search Strategies

It is obvious to state that the performance differences across different group structures can only be driven by the actual choices made during the search for solutions. This path dependent viewpoint has underpinned the theoretical search literature. Yet, the social psychology literature has exhibited a firm tendency to approximate these involved search dynamics through a mere measure of the number of solutions generated. Often, this is even proposed as a good measure of performance for brainstorming efforts, irrespective of the nature of the obtained solutions (e.g., Gallupe et al. 1991). This "limitation" is not unreasonable; it is primarily justified by the sheer difficulty of empirically recording any detailed action or decision that might be meaningful during the search process. Without any doubt, the number of solutions generated is an important metric. In statistics, the extreme value theory predicts that the higher the number of (independent) draws from a distribution the higher the expected value of the maximum draw (Dahan and Mendelson 2001, Girotra et al. 2010). In search terms, the more searches (draws) from a solution landscape (distribution) the higher the expected best solution (maximum draw).

However, statistics also predicts that additional searches will have decreasing returns. Beyond the aforementioned effect that the number of solutions has on the best outcome, the impact of problem complexity on this effect is less straightforward. On one hand, higher complexity implies more variable performance outcomes, and many local performance peaks of potentially low performance. Hence, problems that are more complex might benefit more from a larger number of solutions being explored to ensure that searches move beyond local optima. On the other hand, as argued above, complexity also results in poorer mental models, making the generation of each additional solution more random. Thus, after a reasonably good solution has been found, an additional solution search in a complex landscape is more likely to result in a poor performance outcome than such an additional search in a less complex landscape (Kaufmann et al. 2000). Therefore, we propose an open explorative statement regarding the impact of complexity on the importance of the number of solutions generated, as we cannot argue convincingly for one clear effect. Accordingly, we pose the following hypothesis:

H5a: The group's performance outcome increases in the number of solutions generated by a group, however, at a decreasing rate. This impact may increase or decrease with the complexity of the problem.

While the number of solutions generated is certainly important, the actual solutions found could be all very similar or vary widely. The coverage of the solution space has recently received a good deal of attention as an important driver of search outcomes (Kornish and Ulrich 2011, Erat 2017). Coverage is thought to be stymied by two group dynamics in particular: *collaborative fixation* (Kohn and Smith 2011) which takes place in collaborative structures and reinforces a fixation around some commonly viewed solutions, and *clustering* effects which result from nominal efforts simply concentrating on the immediately promising solutions (a 'low hanging fruit' strategy; Kornish and Ulrich 2011)[6]. These effects tend to lead to the generation of potentially many similar solutions. Erat (2017) demonstrates that the dispersion of ideas in an idea pool is an important driver of the performance of the best idea. Hence, *coverage* is a critical factor in search assessment, beyond the information that a simple count of solutions might provide. Therefore, even in the presence of high solution counts, the search coverage should have an observable impact on search performance.

Moreover, we posit that the importance of coverage might in fact be contingent on the problem complexity. When the complexity is low, groups can fairly quickly identify good regions in the solution landscape. Through such a better understanding of the solution landscape, groups should perform better if they limit their further coverage of the landscape, and perform their entire search within these good regions

[6] Erat and Krishnan (2012) provide an insightful normative model that explains how incentives push rational *individual* independent solvers (i.e., a structure similar to nominal structures) who search for the best solution to cluster around a few good solutions.

rather than "wasting" the same number of searches more widely. Therefore, we expect that for low complexity, higher coverage leads to a lower overall performance of groups in a time-constrained search. On the other hand, we have already argued that for high complexity it is difficult or borderline impossible for individuals to create a representative enough mental model of the solution landscape. Hence, to identify the best region in the landscape is not effective, and it becomes more important to cover a larger part of the solution landscape; i.e. avoid zooming into the (many) low performance peaks. We therefore suggest that for high complexity, higher coverage improves the overall performance of groups. Accordingly, we pose the following hypotheses contingent on the problem complexity:

H5b: In problems with low complexity, the group's performance outcome decreases with a greater coverage of the solution space.

H5c: In problems with high complexity, the group's performance outcome increases with a greater coverage of the solution space.

Finally, we consider the extent to which subsequent solutions differ from one another. We term this factor "*step size*". The same number of observations and the same coverage can be obtained by making small steps moving though adjacent parts of the landscape, or by jumping back and forth between far away parts of the landscape. Step size captures the dynamics (path dependency) of the search process, and as such, it enables us to understand whether certain search patterns (e.g. small ascending steps versus large exploratory jumps) achieve greater performance results. Wooten (2013) highlights the potential for large steps (leaps) to enhance the speed of solution discovery, but also suggests the caveat that such large steps may not in themselves elevate the quality of solutions found.

Once more, the knowledge of the solution landscape should matter. Initially, when the complexity of the solution space itself is still unknown, large step sizes have the potential to identify likely search area candidates as well as eliminate others more quickly. In low complexity settings, large exploratory steps provide assurances regarding the appropriateness of subsequent smaller steps in a local space. In high complexity problems, these serve a similar purpose in helping to assess appropriate areas for later focus, but it may be important to persist with larger steps for a higher number of searches (thus a higher average step size). Recent work in computationally efficient search-algorithm design has reinforced the importance of large step sizes in such terrains (Kanagaraj et al, 2014): where small average step sizes are incapable of the rapid assessment, elimination or refocus of broader search options, the inclusion of large steps (i.e. searches characterized by larger average step sizes) can do so.

As in our discussion for coverage, we propose that complexity plays a moderating role. In problems with low enough complexity, groups will quickly gain a good understanding of the search terrain, and then should use smaller steps to zoom into the actual performance peak, hence leading to a lower average step size. However, for problems with high complexity such understanding is difficult to gain. Worse, since complexity creates many local performance peaks of potentially low performance, groups using small steps might actually get stuck in a local peak. Hence, they might need to engage in some "long jumps" to get out of low, local performance peak (Levinthal 1997). This would suggest that under high complexity groups with a larger average step size should perform better.

There might be an opposing factor at play here as well. The more complex the solution landscapes, the lower the correlation between adjacent points; high performance peaks in a complex landscape have relatively small attraction basins (Rivkin 2000), and hence groups using too larger steps might simply jump over attractive regions. While the overall impact of step size in a complex solution landscape is less clear, we suggest that the risk of getting stuck in a local peak is more important. Hence, we state the following hypotheses, again contingent on the problem complexity:

H5d: In low complexity problems, group performance outcomes decrease with greater average step size.

H5e: In high complexity problems, group performance outcomes increase with greater average step size.

Figure 1 summarizes the above developed hypotheses, and show how they link to search structure, structure sequencing, and search tactics.

[Insert Figure 1 here]

3. Empirical Methods

3.1. Experimental Task

In order to provide an appropriate context for analyzing the impact of complexity on the search strategies and performance of collaborative versus nominal structures, we adapt Ederer and Manso's (2013) experimental setup. Ederer and Manso's task aims to provide a sufficiently non-trivial objective solution landscape whose global optimum (not disclosed a priori) can only be determined through iterative search attempts. As such, it offers the proper structure for researchers to study search processes.

In this setting, the decision-making task is an ostensibly straightforward one: configure a strategy for a maximally revenue generating lemonade stand. The managerial levers available for manipulation consist of three bounded continuous variables (price, % lemon content and % sugar content), as well as two discrete decisions (2 choices of color, and 3 choices of location). The advantage of this experimental context is that all participants can easily relate to the challenge problem and its parameters. This is important given that we are looking at a collaborative setting, where the participants have to discuss their choices; this is easier in a context of parameters they are familiar with e.g., the same way specialized engineers would be familiar with the parameters in the context of a larger development project.

At the globally optimal price and color selections, the objective solution landscape of Ederer and

Manso ("medium complexity") is represented visually in the second row of graphs in Figure 2. The local optima in this second row of graphs differ for each location, but a single global optimum exists (school location, high sugar, low lemon).

[Insert Figure 2 here]

As in Ederer and Manso's experiment, subjects are never explicitly shown the full landscape or set of modeling relationships between the potential decisions and the respective solution performances. Thus, the subjects face an ambiguous problem context, where the underlying structure of cause and effect connections is unknown. They are nevertheless given a starting decision point, and they are asked to improve upon it, based on very limited subjective feedback provided on each new set of decisions developed (details below).[7]

We adapt the setting, to create a low complexity task condition through a simplified objective landscape. We eliminate local optima and make the impact of location simply a step-function contribution to performance (see first row of graphs in Figure 2). For our high complexity condition, we complicate the landscape used by Ederer and Manso through additional local optimal (2 per location-specific landscape; see the 3rd row of graphs in Figure 2). In this figure, landscapes for the non-optimal (worse performing) color option are not shown, however similar albeit down-shifted forms are used in these cases as well. The global optimum, however, does not change with the change in landscape complexity.

The computer interface used in our adaption of the setting provides participants with control over the various decisions involved. The participants are allowed to make any modifications within the permitted ranges of the decision variables both in the collaborative structure and "nominal" group structure settings. Moreover, similar to the work of Shore et al. (2015), they are allowed to submit their decision set to a *market analyst* for further consideration and feedback, someone who is described as being able to make appropriate comparisons across a set of solutions considered, but cannot identify the best solution or further possible improvements in advance. He is described as a real individual, but in reality, this is a very basic artificial intelligence (AI) that evaluates the proposed solution and provides feedback. The feedback is restricted to positive comments, when the most recently considered solution results in an improvement relative to prior submitted solutions. For example, if an improvement in performance associated with a change in the lemon content is achieved, the analyst might state: "*I think there's a good chance what you're*

[7] We acknowledge that there exist more extreme settings, where decision makers face additional ambiguity regarding the factors available to manipulate in defining solutions, or even the nature of the problem objective. Clearly that is a level of ambiguity beyond the one we study. Still, even in those settings, search is the underlying process taking place. The challenge from a laboratory experiment is assessing the value of progress made in settings lacking an unambiguous objective. We deliberately eliminate this form of ambiguity for the purpose of obtaining a clear and comparable performance measure.

doing with lemon content could push us to higher profitability". The commentary list is rich enough to appear non-repetitive but fundamentally equivocal in terminology used. When small local improvements are made during the search, but a globally superior solution has previously been examined, the analyst also makes recommendations to generally look at prior solutions.

A log of solution history, collected to retain all decision sets developed, is provided for subjects to review if desired. (See Appendix A for screen capture of interface). The log of the historical solutions captures in a concise fashion the collective memory of the process, in the same way that a brainstorming session in a room would include drawings, notes and visualizations of almost all solutions proposed. When solutions do not improve performance, no feedback is given; that is, the AI does not provide any guidance about how to improve ideas, but only an evaluation of already mentioned ideas.

This artificial feedback structure is an important element that makes the experimental search process meaningful; without it, the participants would have no means to evaluate the idea and any search effort would be totally random. Thus, the analyst substitutes for the basic contextual knowledge that participants of a product development team would have when brainstorming for a solution to a design problem, for example. It is this contextual knowledge that guides the search process in a real-world setting, but is lacking in our lab controlled experiment. Indeed, Harvey and Kou (2013) show that in reality evaluation of ideas does not - and should not - take place only after ideas are generated (as often advocated in the early literature, Osborn 1953); it should occur throughout the search process. Our analyst provides such evaluation and feedback. Since all search activities are subjected to the same analyst AI interface there is no bias introduced by the feedback provided.

3.2. Laboratory Protocol

Participants in the experiment (subjects) were scheduled to arrive at the lab at designated times in groups of three. They were contacted the day before, to ensure they would show up. They did not know who their respective group members would be prior to arrival. After signing consent forms for participation, subjects completed general questionnaires about demographic information, including background questions on their analytical coursework and deductive reasoning capacity (Clark 1998).

Immediately prior to the experimental task, all subjects were briefed on the nature of the context and how the computer interface and analyst commentary could be used to develop better solutions. Then, subjects were asked to read a two-page primer on optimization and the impact of multimodal complexity on the solution search in objective terrains (for a two-decision continuous variable setting). Following their read, they were given comprehension questions (e.g. "*based on this graphical depiction of the relationships between decisions and outcomes, what leads to a global optimum?*"). Although it was extremely rare to see incorrect answers to these questions (7 individuals in 478), in such cases the subjects were given a more thorough explanation and asked more comprehensive questions. If confusion remained, they were not included in the study (2 subjects), but compensated at a base rate for volunteering (see §3.3).

Dependent upon the treatment condition, subjects began to work towards solution development using the interface either as independent users (nominal), or as a collaborative structure. In treatments where work was independent, separate terminals with separate instances of the task were provided, and no communication between individuals was allowed; in the collaborative structure a single terminal was used and individuals coordinated the decisions with each other. Whether nominal or collaborative, the first stage of work lasted for fifteen minutes. Upon the completion of the first stage, participants either joined each other for an additional fifteen minutes of collaborative work (nominal→collaborative structure condition), or broke up to continue their work independently (collaborative→nominal structure condition). Figure 3 shows this setup visually, to provide a better understanding of the group structures and the sequencing.

[Insert Figure 3 here]

By design, approximately half of the subjects were first exposed to a nominal structure, where separate individuals searched for solutions, prior to searching for the best solution together (sequence 1, representing NGT). The remaining subjects were first exposed to a collaborative structure, prior to being broken up for independent work of the nominal structure (sequence 2, representing reverse-NGT). All groups were further subdivided in terms of the level of complexity associated with the landscape (Figure 2). Approximately one third of all subjects were exposed to a low complexity landscape setting, one third to a medium complexity landscape and the remaining to the high complexity treatment. Following the second stage of work, individuals were asked to independently complete a post-experiment questionnaire about their experience. Items on the questionnaire comprised scales commonly used in the study of group dynamics including evaluation apprehension, free-riding and blocking (items adapted from Alavi 1994, Reinig and Bongsik 2002). All survey responses and activity logs for the first and second stages of the experiment were archived for subsequent analysis [see Appendix B for questionnaire].

3.3. Recording and Further Manipulation

Given the somewhat arbitrary landscape in this experimental task, students are expected to be fairly homogeneous with respect to task specific competencies and knowledge. Such homogeneity allows us to eliminate the noise created by hard to measure knowledge differences, and also allows us to focus on other elements, such as the effect of group structure and complexity. Our experimental setup allows for the collection of search activity data and the tabulation of characteristics of distinct search strategies applied by participants (see details below). Hence, our unit of observation is each individual solution submitted by groups working collaboratively or in a nominal structure. These form the basis of all calculations aggregated at the group level, separately for both the nominal and collaborative structures.

As an additional treatment, we created a condition in a separate experimental study where we constrained the search during the nominal structure stage. These individuals all worked first in nominal group structures and then in collaborative structures. However, unlike the individuals in the "basic" setting described above, "constrained" individuals could only manipulate a subset of the problem parameters during the nominal structure stage. Of the three individuals that constitute a group, one individual could manipulate *price*, location and color, one *sugar content*, location and color, and one *lemon content*, location and color. While each individual could manipulate a different set of variables, within each complexity setting, they all faced partial landscapes of similar complexity. Hence, for the collaborative structures stage in the 2nd 15 minutes, each individual came to the group stage with know-how about the partial landscape. However, during the 2nd 15 minutes, all individuals could suggest changes in all variables; i.e., they were no longer constrained.

3.4. Subject Compensation

To elicit effort in both stages of the experimental task, a random-instance pay-for-performance scheme was applied. Subjects were told that their performance in the two problem solving trials (the collaborative structure and the isolated nominal structure) could be used to calculate additional payouts; up to $13 above the $5 base pay rate depending on their relative performance to peers in the same setting. However, they were also told that the specific trial their compensation is calculated from would be based on a coin toss. A "head" would mean that their compensation would be based only on their individual performance in the nominal structure trial. A "tail" would mean their compensation would be based on their performance in the collaborative structure setting. Hence, individuals had to reenter any solutions from the first group setting (first 15 minutes) that they believed to be particularly good in order to include them in the performance of the second group setting (second 15 minutes). In either case, the stronger they perform the higher their additional payout. The use of a random-instance payment scheme is a commonly used approach to eliciting effort in multiple trial experiments (Bearden et al. 2006, Bendoly 2011).

3.5. Measures and control variables

Dependent Variables:

Unlike most prior experiments in problem solving, ideation, and brainstorming, our setting provides us with a clear and objective measure of the performance of each solution attempted by a group. Since we are considering the search for the best solution, our dependent variable is the *best (or maximal) performance* achieved by a group. The performance in the nominal structures captures the best performance among all ideas identified by any individual in the nominal structure; the performance of collaborative structures is the performance of the best solution found by the group in the collaborative structure. As in any realistic ideation session, neither the nominal nor the collaborative structures actually have to evaluate their ideas and pick the best idea during this session. Rather we focus solely on the idea generation part, and analyze which search processes and approaches generate the objectively best ideas. We make no statement about the ability to select the best ideas. To allow for easier interpretability of the performance levels found across complexity levels and group settings, we look at the performance as a percentage of the maximal performance that could theoretically be achieved (the same value in all complexity landscapes).

Independent variables:

The *number of solutions* measures the total number of solutions found either in collaborative or nominal structures. For nominal structures, it translates into the sum of ideas found by the individuals. The sum could under some settings lead to double counting, i.e., individuals could come up with the same solutions. However, such occurrences cannot happen in our setting due to the variable continuity.

To capture the *coverage* of the solution landscape, we partition the landscape into a grid-system, counting the number of partitions visited (without replication) by a group in the collaborative structure or nominal structure in aggregate. For robustness purposes, we have conducted analysis using a 5-partition, 10-partition and 20-partition scheme for the three continuous variables (2 and 3 partitions for the two nominal variables respectively). As expected, significance of coverage degrades at higher partition levels due to multi-collinearity: With an increasing number of partitions, the coverage variable starts to resemble the "number of solutions". We hence provide the results of the 5-partition scheme.

Step size is calculated as the absolute multi-dimensional Euclidean distance between the decision variables in two consecutively explored solutions. We take the average across both all solution steps of an individual and all individuals in a group. Our step-size measure follows closely the long tradition of the search distance metrics found in the literature (Chao and Kavadias 2008, Chao and Loutskina 2012).

We argue that shared mental models enable individuals to collaborate. To verify this underlying mechanism and capture the similarity in the formed mental models, we calculated the overlap between the areas of the landscape searched by individuals of a group during the first 15 minutes (nominal structure stage). In particular, we used the following metric: *Overlap* = {Total number of areas searched by individuals in the nominal structure – Number of non-overlapping areas searched by the individuals in the nominal structure}/ Number of non-overlapping areas searched individuals in the nominal structure.

Control variables:

Finally, we control for a number of factors that differed across groups. Since the number of team members could impact a group's performance, we control in the generalist setting for *group size*. We also control for two demographic variables frequently considered in group research: the *average age* of the three members of the nominal or collaborative structures as well as *gender-mix*, which we capture with a zero-one variable, with 1 representing mixed gender groups. Finally, we also control for differences in quantitative skills, based on the average across a set of background questions relating to history of analytical coursework and deductive reasoning capacity (see Appendix B).

4. Analysis and Results

4.1 Descriptive Statistics

The subject population consisted of students from graduate university programs. In total, we had 308 individuals in generalist setting, with 58 2-person groups and 64 3-person groups. All groups were scheduled as 3-person groups, but despite every attempt to confirm appointments, several students did not show up for the actual experiment. This resulted in a number of 2-person groups. In the constrained search setting, we had 168 participants, in 56 3-person groups[8]. We conducted a 2x3 design of experiments to explore the performance of different group structures across different problem complexities, the distribution of groups across this design in the generalist setting is described in Table 1.

Table 1. Summary Table of Design of Experiments

	Complexity Treatment		
	Low	*Medium*	*High*
Group Treatment			
Individuals starting in *Nominal* Structures	20	20	20
Individuals starting in *Collaborative* Structures	22	20	20

Table 2 summarizes the key descriptive statistics for the independent and dependent variables.

Table 2. Descriptive statistics of the sample: means, standard deviations and correlations

Variable	Mean	Std. Dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
collaborative performance (1)	0.677	0.280											
nominal performance (2)	0.790	0.218	0.466										
collaborative # of solutions (3)	34.753	12.465	0.140	0.202									
collaborative coverage (4)	8.646	3.531	0.113	0.276	0.229								
collaborative step size (5)	0.255	0.107	-0.062	0.035	-0.524	0.407							
nominal # of solutions (6)	116.292	34.285	-0.253	0.036	0.184	0.149	-0.063						
nominal coverage (7)	15.899	4.232	-0.085	0.067	-0.052	0.371	0.287	0.399					
nominal step size (8)	0.262	0.094	0.238	-0.008	-0.239	-0.089	0.205	-0.321	-0.028				
group size (9)	2.669	0.472	-0.027	0.059	0.087	0.038	-0.102	0.544	0.170	0.226			
average age (10)	22.641	2.349	0.019	0.100	0.149	-0.041	-0.146	-0.227	-0.220	-0.112	-0.166		
avg. quantitative skills (11)	4.191	0.554	0.220	0.195	0.158	0.182	-0.046	0.028	-0.044	-0.285	-0.219	0.007	
gender mix (12)	0.640		-0.230	0.013	0.048	0.094	0.027	0.401	0.459	0.082	0.418	-0.064	-0.210

Amongst our main independent variables, collaborative structures step size is highly negatively correlated with the number of solutions and highly positively correlated with coverage. The former suggests either a quick incremental search or more thought out larger step sizes. The latter hints that when collaborative structures search in larger steps, they really search a larger solution landscape – instead of jumping back and forth between few solution areas. The existence of such correlation is desirable, showing

[8] In the constraint search setting, if a student did not show up, the remaining 2 students were asked to sign up again for a different time and participated in a different experiment instead, or if no different experiment was available for them, they were dismissed with a basic rate for showing up.

reasonable search strategies rather than random search behaviors. For nominal structures, the number of solutions and the coverage are positively correlated, as one might expect. No other search variables show high levels of univariate correlation. From our control variables, age and quantitative skills were not very strongly correlated with other factors. Group size has as expected a high correlation with the number of solutions voiced in the nominal setting, providing additional reasons to control for it. The likelihood of mixed gender groups was obviously also higher in larger groups.

4.2 Hypothesis Tests

4.2.1. Group Structure and Problem Complexity Effects

In order to test Hypotheses 1 through 3, we compare the relative search performance across different complexity levels and group structures. The graph on the left side of Figure 4 plots the performance of nominal and collaborative structures during the initial task exposure (the first 15 minutes) for the three levels of complexity. The black line shows the average of the best performances of nominal structures, i.e., the average calculated across the best performing solutions for each set of individuals in the nominal structure. The grey line plots the average of best solution performances for the collaborative structures. The right side of Figure 4 shows the average of the best performances derived for the exact same individual participants (subjects) reintroduced to the very same task during the second 15-minute stage, where their group structure has been flipped. Hence, the performance of the collaborative structure in the left graph (grey line) and the performance of the nominal structure in the right graph (grey line) is based on the *same* groups of individuals (and similarly for the black lines).

[Insert Figure 4 here]

Table 3. Impact of complexity on performance

	Performance				Performance			
compared complexity levels	Nominal (1^{st} 15 min)		rev. NGT (2^{nd} 15 min)		Collaborative (1st 15 min)		NGT (2^{nd} 15 min)	
	change	p-value	change	p-value	change	p-value	change	p-value
low to medium	0.055	0.498	-0.161	0.004	-0.495	0.000	-0.223	0.003
medium to high	-0.193	0.004	-0.112	0.074	-0.046	0.256	-0.309	0.001
low to high	-0.139	0.095	-0.273	0.000	-0.541	0.000	-0.533	0.000

Apparent from these graphs is the general decline of performance as a function of complexity, regardless of the group structure and prior problem exposure (i.e., whether they are in the basic nominal or collaborative structures, or in the NGT or reverse NGT setting). T-tests with unequal variances reveal that the performance decrease is in most settings statistically significant (see Table 3). This analysis provides support for Hypothesis 1; complexity negatively impacts the performance of both group structures, with or without prior exposure.

4.2.2 Group Structure and Problem Complexity

The hypotheses H2a and H2b collectively state that collaborative structures outperform nominal structures under low levels of complexity, but that this advantage reverses as complexity increases. To test these hypotheses, we compare the relative performance of nominal and collaborative structures during the first 15 minutes. The left-hand side of Figure 4 suggests that during the first 15 minutes collaborative structures perform better than nominal structures at a low level of complexity, but they lose this advantage and perform worse than nominal structures at higher levels of complexity. (We explore the second period performance and the nature of dynamics between the two periods in Hypothesis 3 below.)

T-tests with unequal variances confirm that the performance differences are significant at all three levels of complexity for the first 15 minutes. Furthermore, an ANOVA analysis with interaction terms confirms that indeed the type of group (nominal vs. collaborative structures), the level of complexity and the interaction between these two factors have a significant impact on the groups' performances for the first 15 minutes (significant at 2% level). Hence, complexity moderates the effectiveness of different group settings, providing support for Hypothesis 2a and 2b. Under low complexity collaborative structures obtain a better solution by building on the overall best idea if the search is time constrained; when complexity increases this advantage is out-weighed by the ability of individuals in nominal structures to pursue different search paths and search a larger number of solutions. Table 4 summarizes these results.

Table 4. Differences between performances of nominal and collaborative structures

	Complexity level		
	Low	Medium	High
Nominal Performance in 1st 15 minutes *minus* Collaborative Performance in 1st 15 minutes	-0.129	0.420	0.273
p-value	0.035	0.000	0.000

4.2.3 NGT and Problem Complexity

Next, we test whether the NGT (i.e., preceding a collaborative search structure stage with an initial individual search stage) improves the performance relative to a collaborative structure alone (H3a). Comparing the collaborative structure's performance in the first 15 minutes (grey line on the left side of Figure 4) to the outcome of NGT (black line on the right side of Figure 4), we find that NGT is useful for the *low and medium complexity setting*, increasing the performance significantly. In case of low and medium complexity, the performance of NGT (collaborative structures during the second 15 minutes) is significantly better than the performance of groups using a collaborative structure without prior individual reflection about the problem (collaborative structures during the first 15 minutes) at the 1% level (see also left side and center of Table 5). Hence, we find general support for hypothesis H3a.

Table 5. Performance differences between nominal group technique (NGT) and collaborative structures alone

	Complexity level		
	Low	Medium	High
NGT overall Performance *minus* Collaborative Performance in 1st 15 minutes	0.053	0.324	0.061
p-value	0.002	0.000	0.116

A couple of points, however, deserve additional attention. First of all, allowing for prior exposure to the problem – and hence for additional time to reflect about the problem – does not improve the performance in the case of high complexity (see also right side of Table 5). Further, groups using the NGT perform similarly to collaborative structures without prior problem exposure. In the collaborative stage of NGT, they actually remain below the performance of the nominal structures used in the first 15 minutes, despite the additional time enjoyed.[9] The fact that nominal structures appear to find a significantly better solution during the first 15 minutes than what the group proposes after a subsequent collaborative stage (Figure 4) suggests that not all members contribute their solutions during the collection of idea or not all members are actually listened to. We theorized that this could be due to limited understanding of solution options and their relative benefits (possibly resulting in different mental maps). To test whether constraining the search of individuals during the nominal structure stage (the first 15 minutes) can improve the performance of the subsequent collaboration stage (collaborative search in the second 15 minutes; H3c), we compare the performance of the NGT in the basic (unconstrained) setting to the performance of the NGT in the constrained search setting (see Figure 5 and Table 6).

Figure 5 shows that in low and medium complexity settings the performance of collaborative structures after an initial constrained search is not significantly different than the performance of collaborative structures in the base case; however, in higher complexity settings, collaborative structures after an initially constrained search perform significantly better, providing support for Hypothesis H3c.

Finally, we compare the performance of reverse-NGT (grey line on right side of Figure 4) to nominal search alone (black line on left side of Figure 4). The comparison suggests that preceding the nominal structure stage with a collaborative structure stage (reverse-NGT) is again beneficial relative to nominal work alone in *low* complexity settings (see also Table 7), but not under medium or high complexity (despite the additional time spent). Hence, we find support for H4a and H4b[10]. Interestingly, our finding

[9] We are thankful to one reviewer and the Associate Editor for pushing us to clarify the importance of additional time as an explanation for our findings related to hypothesis 3a.

[10] Echoing our previous discussion, we need to keep in mind that while the additional time certainly plays a role for our findings in the low complexity setting, the change into a different group setting also encourages individual to search again. We frequently observed teams that stopped searching within the 15 minutes time frame. A change in the setting, together with new incentives (for individual performance) encouraged the individuals nevertheless to search again.

related to H4a echoes a similar (in spirit) finding that Chan et al. (2017) document in their analysis of success in filing design patents. The authors observe that successfully filed design patents by individual designers greatly benefit from these designers having collaborated in other design patents in the past.

[Insert Figure 5 here]

Table 6: Performance comparison of the basic nominal group technique (NGT) and of the constrained nominal group technique

	Complexity level		
	Low	Medium	High
Constrained NGT Performance *minus* Basic NGT Performance	-0.007	-0.084	0.172
p-value	0.653	0.844	0.008

Taken together, the results in this section suggest that under lower levels of complexity, collaboration can indeed pay off. However, under higher levels of complexity, isolated and focused nominal work continues to perform well as argued in the broader ideation and brainstorming literature. In fact, one could argue that the extremely ambiguous challenges set forward in the studies performed in that literature (e.g. see footnote 1 in this paper) might be the reason why these studies have so unanimously promoted the dominance of the nominal structures.

On a final note, the reverse-NGT might be worthwhile to consider for high complexity settings, in which a collaboration is desirable for reasons beyond the initial idea generation, e.g., to create commitment to the chosen idea, or to instigate alignment between the team members in terms of the dimensions of the solution space. While the performance is no better than the performance of nominal structures, the final set of suggestions does not suffer from the performance advantage limitations and risks the NGT (and hence 2nd stage collaboration dynamics) appears to suffer from in high complexity settings. If it was not clear whether the level of problem complexity were Low or High, it does appear that reverse-NGT might provide some relatively safe guarantees of performance.

Table 7: Performance comparison of the reverse NGT to nominal structures alone

	Complexity level		
	Low	Medium	High
Reverse NGT Performance *minus* Nominal Performance in 1st 15 minutes	0.180	-0.036	0.046
p-value	0.007	0.707	0.219

4.2.4. The Impact of Search Strategies

In order to gain greater insight into the sources of performance differences between the nominal and collaborative group structures, we now consider the influence of search "strategies" in these settings. We focus on the observable search behavior (rather than on psychological antecedents that drive search behaviors). Figure 6 summarizes objective observations in search behaviors by group structure and timing. In our sample, nominal structures both in the first 15 minutes and those in the reverse NGT explored significantly more solutions and cover significantly larger solution landscapes than collaborative structures in the first 15 minutes or those collaborating in the NGT, suggesting that this distinction in search behavior could largely account for some of the observed performance differences.

[Insert Figure 6 here]

In order to explore how salient this distinction in search behavior may be in these different group settings, as well as that of other search strategy dimensions, we proceed with a more comprehensive model estimation. Table 8 summarizes the OLS regression results different group structures.

Table 8: The role of search strategies: Regression model estimates

	Performance				Performance					
	Nominal (1st 15 min)		Reverse NGT (2nd 15 min)		Collaborative (1st 15 min)		NGT (2nd 15 min)		NGT (2nd 15 min)	
Independent Variables	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**	**Coef.**	**p-value**
Constant	0.564	0.336	**1.189**	0.001	-0.163	0.682	**0.825**	0.017	**0.832**	0.007
Medium complexity	0.867	0.118	-0.180	0.528	**-1.255**	0.000	**-1.499**	0.000	**-1.511**	0.000
High complexity	-0.572	0.227	-0.390	0.204	**-0.381**	0.074	**-1.694**	0.000	**-1.453**	0.000
Number of solutions	-0.009	0.162	**0.007**	0.030	**0.040**	0.001	0.019	0.018	**0.018**	0.012
(Number of solutions)^2/100	**0.006**	0.056	-0.001	0.604	**-0.048**	0.002	**-0.021**	0.020	**-0.018**	0.018
Coverage	**-0.072**	0.002	**-0.025**	0.003	0.005	0.672	-0.014	0.165	**-0.018**	0.052
Avg. step size	-0.723	0.539	**0.932**	0.022	0.027	0.932	**-1.585**	0.001	**-1.421**	0.001
Number x medium complexity	**-0.017**	0.001	0.003	0.131	0.004	0.544	**0.022**	0.001	**0.019**	0.001
Number x high complexity	**-0.011**	0.033	**-0.004**	0.004	**-0.007**	0.031	0.005	0.249	0.001	0.788
Coverage x medium complexity	0.034	0.235	-0.028	0.162	**0.093**	0.001	-0.009	0.717	0.008	0.736
Coverage x high complexity	0.052	0.103	**0.072**	0.000	0.007	0.604	**0.030**	0.012	**0.042**	0.000
Avg. step size x medium complex.	1.460	0.512	0.299	0.764	**-1.321**	0.079	1.252	0.141	1.139	0.148
Avg. step size x high complexity	**3.078**	0.023	**-2.515**	0.005	0.208	0.633	**3.461**	0.000	**2.504**	0.001
overlap (nominal stage)									**0.245**	0.025
Group size	**0.644**	0.000	**-0.281**	0.000	0.011	0.737	**0.164**	0.002	0.036	0.574
Avg. age	-0.004	0.800	**-0.018**	0.061	0.007	0.338	0.005	0.511	0.010	0.176
Avg. quantitative skills	0.056	0.264	**0.069**	0.076	-0.011	0.759	-0.018	0.681	-0.003	0.944
Gender mix	**0.179**	0.037	-0.056	0.320	0.022	0.509	**-0.278**	0.000	**-0.212**	0.000
Number of observations	60		62		62		60		60	
Adj. R^2	74.2%		81.1%		92.8%		93.1%		95.0%	

The first search strategy effect we wanted to examine is the impact of the number of solutions on the group performance. We had hypothesized a concave increasing relationship (H5a). Looking at the low complexity case (direct effect), the number of solutions has a significant positive impact in most settings, while the negative sign on the squared term confirms the concavity of this relationship for the collaborative structures (first 15 minutes and under the NGT). However, for individuals working in nominal structures during the first 15 minutes, neither factor is significant. A possible reason is the higher correlation between coverage and the number of solutions in this setting (we observe a correlation of ~0.5 between coverage and the number of solutions for nominal structures during the first 15 minutes). However, if we ignore the extremes (i.e. five groups with the lowest performance, or one group with the highest overall performance), the coefficients for the number of solutions searched and its square are again of the anticipated sign (positive and negative respectively) and significant at 6% and 1% level respectively. Still, for consistency reasons we have chosen to report regression results without dropping outlier values.

In the case of high complexity (i.e., looking at the interaction term "Number x high complexity"), the impact of the number of solutions on performance is significantly lower for the nominal and collaborative structures in the first 15 minutes, and the reverse NGT (insignificant in the NGT). This resolves our competing hypotheses part towards the viewpoint that additional searches are less effective in high complexity settings. Once a reasonably good solution is found, every additional search step has a high chance to land in a lower performing area (Kauffmann et al. 2000), given that the understanding of the landscape is less clear in the higher complexity settings. We should make an important distinction here: this result does not mean that the number of solutions has a negative impact on performance; the interaction term only captures the difference to the low complexity case. Looking at the overall effect, the sum of the two coefficients remains positive for the high complexity setting. The medium complexity case is less clear cut. A larger number of searches improves the performance of the NGT, but decreases that of nominal structures in the first 15 minutes (insignificant otherwise). A simple explanation for this difference is that nominal structures search significantly more than the NGT, see Figure 6), and hence the nominal structures might pick up again the concave effect of this relationship. Thus, overall, we find partial support for H5a.

Next, we turn to the *coverage* dimension. For low complexity (the direct effect) coverage has a negative effect on performance in all settings except the collaborative one during the first 15 minutes; (the latter is insignificant). Thus, we again find support for H5b. For higher levels of complexity, we need to look again at the interaction terms (third block of independent variables). We hypothesised (H5c) that under high complexity the performance increases in coverage. While this variable is not always significant in all complexity levels and group settings, we find evidence that coverage is more beneficial (significant and positive coefficients for interaction terms) in case of higher levels of complexity. Unlike the impact of the number of solutions, the sign of the coverage effect actually reverses, becoming positive in those settings(comparing the magnitude of the coefficients of coverage alone and the interaction with higher levels of complexity). Thus, we find again partial support for H5c.

Finally, we look at the impact of the *step size*. For low complexity (direct effect of step size), we hypothesised the step size to have a negative effect (H5d), once teams gained an initial understanding of the landscape. We find support for this hypothesis in collaborative structures: without prior exposure to the problem, step size is not significant, but it is negative and significant for the NGT, when individuals had prior exposure to the problem. However, for the reverse NGT the sign is positive and significant. Possibly, individuals searching alone need to use large enough steps to approach a good peak; indeed, their average step size is small, and having larger than average steps might be better given the time constraint.

For high complexity, we suggested that larger steps will increase the performance, since they will prevent teams from getting stuck in low local performance peaks (H5e). Looking again at the interaction effects and the magnitude of the coefficients, we find some support for nominal structures for the 1st 15 minutes, and for NGT (for the collaborative structures during the first 15 minutes, the interaction term is positive but not significant). But for the reverse NGT, the effect of step size is negative at high levels of complexity. We can again only speculate as to the cause. Possibly getting stuck in local peaks might matter less for nominal structures. Since each nominal structure consists of several individuals conducting their own searches, it might indeed be best if they all use small steps to climb up one performance peak, albeit a different one.[11] If they each get stuck in a different peak, the overall performance is still the best of these parallel trials. However, for individuals searching together (collaborative structures or NGT), the situation is different. They cannot afford climbing just one peak and getting stuck in one local peak; especially in the high complexity setting; sufficiently large steps on average ensure that these groups can get out of the basin of attraction of a low peak. Taken together, we find support that the step size matters for performance, but the impact of step size is not always what we hypothesised.

In summary, across the H5 hypotheses we can state that search strategies indeed drive the performance of and the performance differences of the different group structures. As a final post-hoc analysis to supplement these investigations, we conduct a test of whether the difference in mental models might indeed lower the performance of groups using the nominal group technique (NGT). We therefore include *overlap* between the areas of the landscape searched by individuals of a group during the first 15 minutes (during the nominal structure stage), and test its impact on the group performance during the second 15 minutes (during the collaborative stage, i.e., the NGT). Overlap is positive and significant, suggesting that teams that (by chance?!) searched similar areas in the landscape and hence had more similar mental models indeed performed better than those searching very different areas. Anecdotal evidence from quotes

[11] Under high complexity, the magnitude of the step size is comparable to the collaborative structures (not shown).

in our open-ended survey questions also provide evidence that a lack of shared mental models can result in bickering, and can hamper progress, unless one person takes control (see Appendix D).

5. Discussion and Conclusion

The effective engagement of groups during problem solving efforts for ambiguous and complex problems remains a management challenge. Different literatures have tried to address the challenge, by adopting a common conceptualization: problem solving as iterative search efforts within a solution landscape (Kornish and Ulrich 2011). Recent theory has provided suggestions as to the kind of group structures that suit the various levels of problem complexity (Kavadias and Sommer 2009, Girotra et al. 2010). Additional anecdotal evidence offers recommendations for the use and sequencing of group structures. In this paper, we present the first empirical examination of the interplay between problem complexity and the benefit of collaborative and nominal structure search sequences. We further consider how performance differences between group structures may be driven by their differences in the way they search for the possible solutions, i.e. their search strategies.

We use a controlled 2-stage experimental design, with exogenously varying group structure and problem complexity. We collect data on search decisions and patterns, to analyze their performance effects. In line with theoretical predictions, we find support for the performance dominance of collaborative structures in low(er) complexity tasks, and of nominal structures in high(er) ones. We document an additional improvement in collaborative performance made possible by preceding such collaborative work with nominal structures activity (nominal group technique, or NGT). Interestingly however, in high problem complexity settings, we also observe the performance of NGT to be dominated by nominal work alone, as well as by the reverse order of NGT. The finding may in some form provide analogy to the top-down, bottom-up process discussions of authors such as Hutchison-Krupat and Kavadias (2015); emphasizing that individual thought may have a much more instrumental role in complex settings, than simply as a context priming mechanism.

These findings are also of importance to practice, as NGT has been viewed in both the academic and practitioner literature as the panacea of effective problem solving. Our results suggest that management should be careful when employing NGT, since it may not be the most effective approach in all settings. To this point, we also find evidence that using constrained search improves the performance over NGT within high complexity settings. This suggests that management can improve the performance of NGT by (a) inducing individuals to share all individual ideas in writing prior to the collaborative stage, and (b) inducing initial individual exploration of identifiably *different* solution approaches or by involving different specialists, who would naturally focus on different parts of the solution landscape.

An examination of search strategies employed within the different group structures reveals notable distinctions: Collaborative structures generally explore a smaller number of potential solutions, exhibit a lower search coverage, and employ a larger search step size. The subsequent analysis of the impact of these search strategies on performance reveals some important insights: The total number of solutions, often used as a proxy for the group problem solving performance, is not the only important driver of performance; rather two additional factors play a role: the average step size and coverage. While these two factors are correlated, our regressions demonstrate that these are separate factors, and that their importance is contingent on the level of complexity. Low coverage works well for low levels of complexity where groups can quickly zoom into a good region, and therefore the better groups do not waste time in unnecessary search. In high complexity problems, good performance requires higher levels of coverage. Hence, any encouragement or incentive to cover a wide variety of solution approaches are particularly helpful for complex problems. In addition, our results suggest that the number of solutions, in itself, can be a poor proxy for group performance; the total number of solutions tends to be characterized by a concave relationship with performance. Hence, management should be cautious about the limitations of generic rules like "generate as many ideas as possible", a suggestion made often in the brainstorming literature. Interestingly, our analysis shows that as complexity increases the marginal value of an additional solution decreases, so the number of ideas generated should be even higher to achieve good search performance.

Table 9 summarizes once more the hypotheses, the theoretical foundations and the level of support we found for these hypotheses.

Table 9. Summary of Theorized Hypotheses and Results

H	Shorthand Description	Theoretical foundation	Support
H1	More Complexity leads to Loss in Performance	Mental model completeness (Pich et al. 2002)	**Strong**
H2a	Under Low Complexity: Performance of Collaborative > Nominal Structure	Convergence by pooling (Kavadias and Sommer 2009)	**Strong**
H2b	Under High Complexity: Performance of Collaborative < Nominal Structure	Parallelism (Sommer and Loch 2004) and groupthink (Bendoly 2014)	**Strong**
H3a	Under Low Complexity: Performance of NGT >> Collaborative alone	NGT (Delbecq and VandeVen 1971) and idea generation theory (Girotra et al. 2010)	**Strong**
H3b	More Complexity reduces Performance benefits of NGT vs. Collaborative alone	Mental model completeness (Pich et al. 2002) and local search traps (Rivkin 2000)	**Strong**
H3c	Under High Complexity: Diverse specialization constraints to Nominal Structure lead to higher Performance of NGT	Partitioned expertise and recombining (Gavetti and Levinthal 2000, Csaszar and Levinthal 2016)	**Strong**
H4a	Under Low Complexity: Performance of Reverse NGT >> Nominal alone	Coordination and shared mental models (Bendoly 2014)	**Strong**
H4b	More Complexity reduces Performance benefits of Reverse NGT vs. Nominal alone	Mental model completeness (Pich et al. 2002) and fixation (Kohn and Smith 2011)	**Strong**
H5a	Each additional solution provides a smaller positive Gain in Performance	Extreme values (Dahan and Mendelson 2001, Kaufmann et al. 2000)	*Partial*
H5b	Under Low Complexity: More search area leads to Loss in Performance	Refinement and idea generation theory (Girotra et al. 2010)	**Strong**
H5c	Under High Complexity: More search area leads to Gains in Performance	Clustering (Kornish and Ulrich 2011) and dispersion (Erat 2017)	*Partial*
H5d	Under Low Complexity: Larger search steps lead to Loss in Performance	Computational efficiency (Kanagaraj et al. 2104, Rivkin 2000)	*Partial*
H5e	Under High Complexity: Larger search steps lead to Gains in Performance	Computational efficiency (Kanagaraj et al. 2104, Levinthal 1997)	*Partial*

Our study is not without limitations. For one, our experimental setting did not allow subjects to designate rationale or specifically which factors they felt were critical in each steps of their search. This was a critical aspect of our controlled experiment, in that requiring individuals to document such rationale prior to each solution proposed would break the flow of the search efforts in ways that would be unrepresentative of most real-world settings. Similarly, we do not formally control past experience as a factor influencing subsequent search efforts, and hence can only rely on what we observe as solutions in the first stage searches as possible drivers of second stage dynamics. Future studies might consider additional treatments that start some groups at a higher level of performance (less room to improve, and hence less positive feedback) than others. If coupled with some light documentation for each submission in certain treatments (e.g. "Specify which factor you believe is most critical to improve your solution"), a more detailed story of learning effects in simple and complex problem landscapes might emerge.

With respect to the field of practice in innovation, we only consider the idea or concept generation stage. We also compare the performance across group settings purely based on their ability to recognize, but not their ability to select ideas, which is certainly a possible direction for further research. Furthermore, our experimental task does not allow subjects to freely come up with factors available for modification. Rather, in small group structures, we provide a fixed set of factors and fixed ranges of manipulation up front. This allowed us to obtain a clear and comparable performance measure to explore the impact of complexity, but at the same time reduces the comparability to prior experimental work on brainstorming, which used more open ended, ideation tasks. Because of these limitations, we encourage future research extensions to this study that examine the robustness of these findings to alternate settings, representative of other forms of complex group work. Finally, our comparison of the NGT to nominal structures and collaborative structures but also our analysis of the reverse of NGT (a sequence that somehow has been neglected in the extant literature) could benefit from additional future research which would ensure that the additional time available to NGT in our setting is not shaping the results recorded.

Lastly it goes without saying that group work involves social elements, and that the communication of information and the adoption of knowledge in these settings is highly dependent on individual psychology. The work of Ederer and Manso (2013) for example, highlights the importance of informational spillovers, and the impact of perceived responsibilities and guilt for failing to fulfill these. In our original design, we in fact inquired into perceived social dynamics in the settings examined. Post-task assessments of evaluation apprehension, perceived production blocking and free-riding were each collected, yet failed to provide consistent predictive strength with regards to performance. This may simply be an artifact of statistical power, or it may be that these particular subjective scales captured post-task do not sufficiently capture the social challenges individuals from different backgrounds encounter. The literature would benefit from future work focusing on connections between these social phenomena and performance in distinct group structure and sequencing approaches.

References

Alavi, M. (1994) Computer-mediated collaborative learning: an empirical evaluation. *MIS Quart.* 18(2): 159-174.

Bearden, J.N., Rapoport, A., Murphy, R.O. (2006) Sequential Observation and Selection with Rank-Dependent Payoffs: An Experimental Study, *Management Sci.* 52(9): 1437-1449.

Bendoly, E. (2011) Linking task conditions to physiology and judgment errors in RM systems. *Production Oper. Management* 20 (6): 860–876.

Bendoly, E. (2014). Systems dynamics understanding in project execution: information sharing quality and psychological safety, *Production Oper. Management* 23(8): 1352–1369.

Billinger, S., Stielglitz, N. and Schumacher, T.R. (2014) Search on rugged landscapes: an experimental study. *Organ. Sci.* 25(1): 93-108.

Blum, C.; Roli, A. (2003) Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys 35(3)*: 268–308.

Camerer, C., Weber, M. (1992) Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk Uncertainty* 5: 325-370.

Chan, T.H., J. Mihm, M. Sosa. (2017) Revisiting the Role of Collaboration in Creating Breakthrough Inventions. *working paper*.

Chao, R., S.Kavadias (2008) A Theoretical Framework for Managing the NPD Portfolio: When and How to Use Strategic Buckets. *Management Sci*. 54(5): 907-921.

Chao, R., Lipson, M. and E. Loutskina (2012) How Organizational Complexity Impacts R&D Portfolio Decisions and Technological Search Distance, Darden School of Business *working paper*.

Clark, J. (1998) Fairness preferences and optimization skills: are they substitutes? an experimental investigation. *Journal of Economic Behavior and Organization* 34: 541-557.

Csaszar, F.A., D.A. Levinthal (2016) Mental representation and the discovery of new strategies. *Strat. Mgmt. J.* 37: 2031–2049.

Dahan, E., Mendelson H. (2001) An extreme value model of concept testing. *Management Sci*. 47(1): 102–116.

Delbecq, A. L., Van de Ven, A. H (1971) A Group Process Model for Problem Identification and Program Planning. *Journal of Applied Behavioral Science* 7: 466–91.

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

Ederer, F., Manso, G. (2013) Is pay-for-performance detrimental to innovation? *Management Sci*. 59(7): 1496-1513.

Erat, S., Kavadias, S. (2008) Sequential Testing of Product Designs: Implications from Learning. *Management Sci.* 54 (5): 956-968.

Erat, S., Krishnan, V. (2012) Managing delegated search over design spaces. *Management Sci*. 58(3): 606-623.

Erat, S. (2017) Making the best even better: The Role of Idea Pool Structure. *Production Oper. Management* 26 (10): 1946-1959.

Ethiraj SK, Levinthal DA (2009) Hoping for A to Z while rewarding only A: complex organizations and multiple goals. *Organ. Sci*. 20(1): 4–21.

Gallupe, R.B., Bastianutti, L.M., Cooper, W.H. (1991) Unlocking brainstorms. *Journal of Applied Psychology* 76: 137-142.

Gavetti G., D. Levinthal (2000) Looking Forward and Looking Backward: Cognitive and Experiential Search. *Admin. Sci. Quart* 45: 113-137.

Girotra, K., Terwiesch, C., Ulrich, K. (2010) Idea generation and the quality of the best idea. *Management Sci*. 56(4): 591-605.

Harvey, S., C. Kao. (2013) Collective Engagement in Creative Tasks: The Role of Evaluation in the Creative Process in Groups. *Admin. Sci. Quart* 58 (3): 346–386.

Heller, P., R.Keith, and S. Anderson. (1992) Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *Am.J.Phys*. 60(7): 627-636.

Hutchinson-Krupat, J., Kavadias, S. (2015) Strategic resource allocation: top-down, bottom-up, and the value of strategic buckets. *Management Sci.* 61(2): 391-412.

Kanagaraj, G., Ponnambalam, S.G., Jawahar, N., Mukund, J. (2014) An effective hybrid cuckoo search and genetic algorithm for constrained engineering design optimization, *Eng. Optimiz.* 46(10): 1331-1351.

Kauffman, S. A. (1993) The Origins of Order: Self-Organization and Selection in Evolution, New York: Oxford University Press.

Kaufmann S. A., Lobo and McReady (2000) Optimal Search on a Technology Landscape, *Journal of Economic Behavior and Organization*, 43 (2): 141-166.

Kavadias, S., Sommer, S. (2009) The effects of problem structure and team diversity on brainstorming effectiveness. *Management Sci*. 55(12): 1899-1913.

Kohn, N.W., Smith, S.M. 2011. Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology* 25(3): 359-371.

Kornish, L.J., Ulrich, K.T. (2011) Opportunity spaces in innovation: empirical analysis of large samples of ideas. *Management Sci*. 57(1): 107-128.

Knudsen, T., D. A. Levinthal (2007) Two Faces of Search: Alternative Generation and Alternative Evaluation. *Org. Sci.* 18(1): 39-54.

Laughlin, P.R., E.C. Hatch, J.S. Silver, L.Boh. (2006) Groups perform better than the best individuals on Letter-to-Numbers Problems: Effects of Group Size. *J Pers Soc Psychol* . 90(4) 644-651.

Lazer, D. Friedman, A. (2007) The Network Structure of Exploration and Exploitation. *Adm Sci Q* 52: 667–694.

Levinthal, D.A. (1997) Adaptation on Rugged Landscapes. *Management Sci.* 43(7): 934-950.

Loch, C.H., C. Terwiesch, S. Thomke. (2001) Parallel and sequential Testing of Design Alternatives. *Management Sci*. 45(5): 663-678.

Osborn, A.F. (1963) Applied imagination: Principles and procedures of creative problem solving (Third Revised Edition). New York, NY: Charles Scribner's Sons.

Paulus, P.B., Dzindolet, M.T. (1993) Social influence processes in group brainstorming. *J Pers Soc Psychol* 64: 575–586.

Paulus, P. B., V. Brown, A. H. Ortega. (1996) Group creativity. R. E. Purser A. Montuori, eds. Social Creativity in Organizations. Hampton Press, Cresskill, NJ.

Rivkin, J.W. (2000) Imitation of Complex Strategies. *Management Sci.* 46(6): 824-844.

Pich, M.T., Loch, C.H., De Meyer, A. (2002) On uncertainty, ambiguity and complexity in project management. *Management Sci.* 48(8): 1008-1023.

Reinig, B.A., Bongsik, S. (2002) The Dynamic Effects of Group Support Systems on Group Meetings. *Journal of Management Information Systems* 19(2): 303 - 325.

Robbins, S. P., T. A. Judge. (2007) Organizational Behavior. Prentice Hall, Upper Saddle River, NJ.

Sawyer, R. K. (2007) Group Genius: The Creative Power of Collaboration. Chap. 4. Basic Books, Perseus Books Group, Cambridge, MA.

Shore, J., Bernstein, E. and Lazer, D. (2015) Facts and figuring: an experimental investigation of network structure and performance in information and solution spaces. *Organ. Sci*. 26(5): 1432-1446.

Siggelkow, N., J.W. Rivkin. (2005) Speed and Search: Designing Organizations for Turbulence and Complexity. *Organ. Sci*. 16(2): 101–122.

Simon, 1969. The Science of the Artificial, 2nd ed. MIT Press, Boston, MA.

Singh, J., L. Fleming. (2010) Lone Inventors as Sources of Technological Breakthroughs: Myth or Reality? *Management Sci*. 56: 41-56.

Sommer, S., Loch, C. (2004) Selectionism and learning in project with complexity and unforeseeable uncertainty. *Management Sci*. 50(10): 1334-1347.

Sting, F.J., J. Mihm, C.H. Loch. (2016). Collaborative Search. *INSEAD Working Paper* No. 2016/60/TOM

Stroebe, W., M. Diehl. (1994) Why are group members less effective than their members: On productivity losses in idea-generating groups. *Eur. Rev. Soc. Psych*. 5: 271–303.

Sutton, R. I., A. Hargadon. (1996) Brainstorming groups in context: Effectiveness in a product design firm. *Admin. Sci. Quart*. 41: 685–718.

Wason, P.C. (1966) Reasoning, in: Foss, B.M. (Ed.), New Horizons in Psychology. Penguin Books, Harmondsworth, 135-151.

Weitzman, M. L. (1979) Optimal Search for the Best Alternative. *Econometrica*, 47(3): 641-654.

Wooten, J.O. (2013) Innovation Tournaments: Improving Ideas through Process Models. University of Pennsylvania Dissertation.

APPENDIX A - Screenshot of Computer Interface

[Insert Image A.A1 here]

APPENDIX B - Selected Questionnaire / Survey-Items used in Analysis

Post-Task Survey Quantitative Skill Assessment

{Wason Selection Test} --- Below is a picture of 4 cards spread out on a table. You only see the side facing up. The following claim has been made:

"If a card has a vowel on one side, then it has an even number on the other side."

1) Which card(s) would you need to turn over in order to know if the claim is true or false? (circle all cards you believe need to be turned over).

   [[E]] [[K]] [[4]] [[7]]

2) If the claim <u>is</u> true, a card with a letter "A" on one side cannot have a "3" on the other side (T / F)
3) If the claim <u>is</u> true, a card with a letter "8" on one side cannot have a "B" on the other side (T / F)

How would you rate your knowledge of the following college course topics?

Non-linear Algebra; Calculus; Optimization; Logical Reasoning; Project Management

{all Likert-type scales from 1-7, 1="No knowledge", 7="Very strong knowledge"}

Post-Task Retrospective Survey Items

<u>Evaluation Apprehension</u>

evi: The members of my work group viewed my ideas as . . . (not at all valuable / highly valuable)

ev2: In regard to offering contributions to the discussion, I was . . . (very apprehensive / not at all)

ev3: There were times when I refrained from participating because I felt others might not accept my ideas, (strongly disagree / strongly agree)

<u>Free Riding</u>

frl : To be honest, I just took it easy and let the other members of the group do most of the discussing, (strongly disagree / strongly agree)

fr2: How satisfied are you with your own performance as a group member in this task? (very dissatisfied / very satisfied)

fr3: How much do you feel you participated in group idea generation? (not much at all / a lot)

<u>Production Blocking</u>

pbl : In the group, when I thought of an idea I . . (could express it immediately / had to wait to express it)

pb2: In the group, did you express your ideas . . . (soon after you thought of them / after waiting awhile)

pb3: I got my ideas out to the group as soon as they occurred to me. (strongly agree / strongly disagree)

<u>Open-ended experience questions</u>

In which setting do you believe the most progress in increasing potential revenue was attained: Independent-work or Group-collaboration?

Please describe the strategy used (if any) in the group setting? Specifically, how did your group decide which issues to focus on for modification? What kinds of things triggered changes in the general direction the group followed?

Please describe the strategy you used (if any) when working independently? If it differed from the group strategy, make sure to explain how. Consider your response to the previous question in outlining any differences.

APPENDIX C - Evidence of groups getting stuck in local optima

[Insert Image A.C1 here]

[Insert Image A.C2 here]

[Insert Image A.C3 here]

[Insert Image A.C4 here]

APPENDIX D: Quotes from open-ended questions suggesting difficulties to integrate differing mental models in the collaborative stage:

Group with Low Final Performance:

"…We couldn't agree on whether to raise or lower the lemon content. One of the group members said they could get better results if it was lowered. We went back and forth and spent most time on it, but maybe should have thought more about the other issues. Still the debating probably helped use avoid bad solutions."

Group with High Final Performance:

"The mouse was mine, so probably did more than the others. They both wanted to go in two different directions, and not what I thought was best (who knows). I kind of tuned out early on and drove. I kept saying I'd "test that after this" but since we had momentum we usually didn't go back…"