From Problems to Solutions in Strategic Decision-making: The Effects of Generative AI on Problem Formulation[1]

Nety Wu
nety.wu@insead.edu

Hyunjin Kim
hyunjin.kim@insead.edu

Chengyi Lin
chengyi.lin@insead.edu

INSEAD

15 August 2025

Abstract

Recent studies show that large language models (LLMs) can augment the generation and evaluation of ideas, but their impact on problem formulation remains underexplored. Through a randomized controlled trial with 305 MBA students, we investigate how LLM assistance at different stages of the decision process affects strategic outcomes. Consistent with prior findings, we find that LLM assistance increases the number of alternatives generated. However, a surprising pattern also emerged: using LLMs in both problem formulation and ideation decreases strategic focus, an effect not observed when LLMs assist only in ideation. Using an abductive approach, we propose that when introduced during problem formulation, LLMs appear to shape how individuals mentally construct the strategic problem and perceive LLM's role, creating cognitive anchors that bind subsequent search. Our study advances the understanding of human-AI collaboration in strategic contexts, highlighting the importance of when and how LLM is integrated in decision-making.

Keywords: Strategic Decision-Making, Large Language Models, Human-AI Interaction, Problem Formulation, Managerial Cognition

[1] We are grateful to Nghi Truong and Simonne Pinto for their valuable support to this study. The experiments were approved by the INSEAD IRB and pre-registered on the AEA RCT registry.

INTRODUCTION

Strategic decisions are consequential, ambiguous, and cognitively demanding (Cyert & March, 1963; Leiblein et al., 2018). They typically unfold through phases of agenda setting, problem formulation, alternative generation, and evaluation, often under conditions of limited attention and bounded rationality (Simon, 1947). The recent emergence of Large Language Models (LLMs) has begun to fundamentally alter the process of strategic decision-making (Csaszar et al., 2024). As tools that extend human information processing capabilities, LLMs have demonstrated their ability to support strategic decision-making by assisting with idea generation (Boussioux et al., 2024) and evaluation (Doshi et al., 2025).

However, most of the research on LLMs and strategic decision-making has treated the problem as given, letting go of the upstream stages and focusing on solution generation and evaluation within a predetermined space. Typical tasks include "generate different future scenarios using idea A" or "propose an idea in industry B according to the evaluation criteria." In doing so, they bypass the foundational step where decision-makers must determine what the core strategic challenge actually is: whether it is a competitive positioning problem, a resource allocation decision, a market timing issue, or something else entirely. By treating the problem as given, these studies neglect problem formulation as a distinct cognitive process that precedes and shapes subsequent ideation.

Problem formulation involves constructing a mental representation of the strategic challenge, deciding which dimensions matter and how they relate (Baer et al., 2013), while ideation focuses on generating alternatives within that frame. This distinction is theoretically consequential: how a problem is framed can shape downstream reasoning and choices (Baer et al., 2013; Park & Baer, 2022). As such, understanding how LLMs intervene in this early cognitive phase is critical for evaluating their impact on strategic decision-making.

Building on these insights, we adopted an abductive approach (Heckman & Singer, 2017; King et al., 2021) to examine how LLM-assisted decision-making unfolds across problem formulation and ideation stages. We conducted a randomized controlled experiment with 305 MBA students at a leading business school. Participants were asked to formulate the key problems faced by a firm described in a case, generate potential strategic options, and select the one they believed was best. The experimental design included three conditions: a control group without LLM assistance, a partial treatment group with GPT-based LLM support for ideation only, and a full treatment group with GPT-based LLM support for both problem formulation and ideation.

The experiment revealed an interesting puzzle. Consistent with prior literature, LLM assistance increases the number of options generated. However, when participants used LLMs for both problem formulation and ideation, they were less likely to produce or select strategic options: the proportion of strategic options dropped by 7 percentage points, and the likelihood of choosing a strategic option fell by 15 percentage points. By contrast, using LLMs for alternative generation alone does not result in a change in strategic focus or option selection.

To further explore this puzzle, we followed an abductive approach (Heckman & Singer, 2017; King et al., 2021), iterating between our data and relevant theory. We began by identifying three potential cognitive mechanisms from the literature: cognitive offloading, where participants delegate strategic thinking to the LLM; attention dilution, where an LLM-generated problem framing spreads attention across multiple dimensions; and cognitive drain, where intensive early-stage engagement with LLMs depletes mental resources for subsequent tasks. For each mechanism, we derived observable implications in our setting and compared these expectations to the patterns in our data.

While no single mechanism fully accounted for the observed patterns, this iterative process pointed to an emerging synthesis: when LLMs are introduced during problem formulation, participants often anchor on the seemingly complete LLM-generated frame. Yet, since they had little insight into how the frame was constructed, they tend to exert their efforts in generating options within the frame, rather than considering whether the problem itself needs to be changed, which might limit the generation of strategic options during ideation. Participants continue to engage in active thinking but redirect their cognitive efforts in ways that constrain strategic reasoning. In contrast, when participants first formulated the problem themselves, they retained a clear sense of the assumptions underlying the problem, and were more likely to treat the LLM as an input source rather than a defining frame.

Our study makes several contributions. First, it extends our understanding of LLM assistance in strategic decision-making (Boussioux et al., 2024; Csaszar et al., 2024; Doshi et al., 2025; Otis et al., 2024). By distinguishing between AI's role in problem formulation versus alternative generation, we illuminate how the cognitive effects of AI assistance vary across decision stages, contributing to the literature on the micro foundations of strategic decision-making (Baer et al., 2013; Laureiro-Martinez et al., 2023; Newell & Simon, 1972; Simon, 1947). Second, we propose that when introduced during problem formulation, LLMs can reduce strategic focus, by possibly providing a default reference frame for subsequent thinking. This builds on work on task allocation considerations (Choudhary et al., 2025; Puranam, 2021) and extends it beyond questions of division of labor to address the cognitive integration challenges involved in strategic contexts (Gavetti & Levinthal, 2000; Laureiro-Martínez & Brusoni, 2018). Rather than simply asking "who does what," our study examines how human-AI interaction affects the cognitive processes underlying strategic decision-making. Finally, our findings open a new research agenda around the cognitive interplay between humans and AI in shaping strategic representations, with implications for strategic decision-making and the design of AI-augmented strategy processes.

THEORETICAL MOTIVATION

Strategic decision-making is a cognitively demanding process that requires managers to navigate ambiguity, surface causal structures, and formulate courses of action with far-reaching consequences (Leiblein et al., 2018). As these cognitive demands intensify in complex and uncertain environments, the question of how external tools might support strategic thinking becomes increasingly salient. Among these tools, large language models (LLMs) stand out for their ability to generate, organize, and reframe information in ways that appear well-suited to the demands of strategic decision-making.

Existing studies have demonstrated much interest in the potential of LLMs for strategic decision-making, especially for ideation and evaluation. Recent empirical studies suggest LLMs can produce strategic ideas comparable in quality to those generated by human experts (Csaszar et al., 2024) and, in some cases, even enhance decision quality through human-AI collaboration (Boussioux et al., 2024). In consumer-facing contexts, LLM-generated ideas outperform human ideas in average purchase intent, though often with lower novelty (Meincke et al., 2024). In evaluative tasks, LLM assessments have been found to align with expert judgments (Doshi et al., 2025). These findings point to the growing potential of LLMs in strategic decision-making.

Yet much of this emerging research has focused on solution generation and evaluation within predetermined frameworks, treating the problem as given and ignoring the cognitively distinct stage where decision-makers define what the problem is. For example, researchers might prompt LLMs to generate ideas within pre-specified domains, elaborate strategies using predefined dimensions, or evaluate scenarios according to established criteria. While these designs offer clarity in assessing LLM capabilities for ideation and evaluation, they neglect the critical cognitive step of defining and structuring the problem before ideation begins. In doing so, existing studies have conflated two cognitively distinct stages of strategic decision-making: problem formulation, where decision-makers define what the problem is, and ideation, where they explore potential solutions. This omission is theoretically consequential because problem formulation and ideation involve fundamentally different cognitive processes. Conflating them may mask how LLMs differentially influence these distinct stages of strategic thinking.

This concern resonates with foundational insights from the Carnegie School of decision-making, which has long emphasized that strategic choices emerge not through isolated stages but through cognitively entangled processes, where problem formulation, alternative generation, and evaluation are dynamically interdependent (Cyert & March, 1963; Simon, 1947). Under bounded rationality, managers rely on simplified mental models to navigate complex environments (Gavetti, 2012; Gavetti & Levinthal, 2000). From this perspective, by providing pre-defined problems, existing studies bypass the crucial cognitive work of problem formulation, which shapes the solutions decision-makers perceive as viable (March & Simon, 1958). They may inadvertently mask how LLMs influence the transitions between framing, generating, and selecting strategic options, precisely where cognitive dynamics are most consequential. To address this gap, we propose expanding the scope of inquiry to treat problem formulation not as a background assumption, but as a distinct and theoretically meaningful phase of strategic cognition. Doing so allows for a more precise understanding of how LLMs might differentially influence the foundational task of constructing problem representations versus generating solutions within already-defined frames.

Clarifying the difference between problem formulation and ideation is therefore central to understanding how LLMs may alter strategic decision-making. Problem formulation involves the construction of mental representations that define the dimensions and boundaries of a strategic challenge (Baer et al., 2013; Nickerson & Argyres, 2018). This process establishes the "problem space"—the cognitive framework through which decision-makers subsequently search for solutions (Newell & Simon, 1972). Strategic problem formulation aims to identify nonredundant and relevant causes of complex, ill-structured problems based on observable symptoms (Baer et al., 2013). When formulating problems, managers determine which dimensions are relevant, how they interrelate, and which causal structures might explain observed symptoms (Mitroff & Featheringham, 1974). Enhancing problem formulation is crucial because it determines what problem is solved and ultimately the quality of the solution (Baer et al., 2013; Nickerson & Argyres, 2018). Effective problem formulation could help managers avoid the pitfall of solving the "wrong" problem due to an overly narrow or inappropriate formulation (Baer et al., 2013; Mitroff & Featheringham, 1974). Traditionally, problem formulation approaches have been broadly categorized as either rational (emphasizing systematic analysis) or intuitive (emphasizing the use of heuristics). Recent research suggests the effectiveness of integrating rational and intuitive thinking for novel problem formulation (Park, 2024), and moving from concrete to abstract thinking to develop comprehensive causal explanations (Park et al., 2025).

In contrast, alternative generation or ideation operates within the established problem representation to identify potential actions. This cognitive activity involves search processes that are constrained by and dependent upon the preceding problem formulation (Csaszar & Levinthal, 2016). The quality and diversity of generated alternatives thus depend substantially on the comprehensiveness and accuracy of the initial problem representation (Park & Baer, 2022). These representations can facilitate "long jumps" searches beyond local terrain (Gavetti & Levinthal, 2000) and assist managers in navigating novel situations through analogy (Gavetti et al., 2005). Empirical evidence suggests that well-defined mental representations significantly enhance search and evaluation processes, ultimately improving decision outcomes (Csaszar & Laureiro-Martínez, 2018; Heshmati & Csaszar, 2023). Thus, recognizing the distinct cognitive tasks of problem formulation and ideation highlights how effectively constructed problems could influence the subsequent generation of strategic alternatives.

Although LLMs have shown promise in ideation and evaluation, less is known about how they interact with the distinct cognitive demands of problem formulation. LLMs are designed to predict the next token using probabilistic associations learned from vast and varied textual corpora. This makes them particularly effective at tasks involving elaboration, extension, and recombination of existing inputs , which are hallmarks of ideation that involve associative thinking and pattern completion. In contrast, problem formulation involves constructing a coherent problem space from ambiguous cues, selecting relevant dimensions, and inferring potential causal relationships (Baer et al., 2013; Nickerson & Argyres, 2018), which may call on different cognitive processes than ideation. How LLMs interact with these distinct processes is not yet well understood.

Moreover, LLMs primarily interact with individual users, making individual-level cognition a natural starting point for understanding human-AI interaction in strategic contexts. While much strategic decision-making occurs in teams, individual decision-makers also play crucial roles in organizations, from identifying market opportunities and formulating responses

(Novelli & Spina, 2024) to influencing firm outcomes (Kruse et al., 2023) and shaping responses to institutional shocks (Lamberg & Peltoniemi, 2020). Recent work shows that individuals working with LLMs can match the performance of human teams (Dell'Acqua et al., 2025), highlighting the potential for AI tools to narrow the difference between solo and teamwork. Understanding how LLMs shape individual strategic cognition provides essential building blocks for future research on team-level dynamics, where multiple individuals, each potentially influenced by AI assistance, interact to produce collective strategic outcomes.

RESEARCH DESIGN

This study adopts an abductive approach that combines randomized controlled experiments with iterative theory development. The experiment systematically varies when participants receive LLM support and compares how these interventions shape their decisions. Unexpected empirical patterns from the experiment prompt the iterative cycles between data and literature, allowing us to evaluate alternative explanations and refine our theoretical account. This exploratory process ultimately converged on a synthesized explanation informed by both evidence and theory, though we caution against over-interpreting the findings.

Sample, Experimental Task, and Treatment

Our sample for the main experiment consists of 305 MBA students enrolled in a core strategy course at a leading business school. Table 1 presents the summary statistics of the sample, showing that 32% are female, and 31% were employed in management consulting before enrolling in the MBA program. Table A.1 in the Appendix shows the distribution of participants across sections and experimental groups. Table A.3 presents the pairwise comparison across experimental groups.

--INSERT TABLE 1 HERE --

The experimental task asked participants to diagnose the causes of a company's failure to achieve profitability and develop strategic options based on a pre-distributed case study of Rated (a pseudonym to maintain confidentiality), an online reviews platform (Kim, 2021)[1]. The company had established a significant user base and was leading its competitors in terms of monthly unique visitors. However, it struggled with monetization challenges after eight years of substantial growth without reaching profitability. The case provided a detailed snapshot of the company's performance, including financial data, market position, and insights from key executives. These executives identified challenges such as outdated pricing models for advertising, high turnover in the salesforce, and increased competition from newer platforms. The case was chosen because it mirrors the real-world complexity of strategic decision-making, where surface-level symptoms, such as declining profitability, may obscure deeper underlying issues. The case ended with a question of what strategic options could redefine Rated's future and improve its long-term viability and competitiveness in an evolving industry landscape. Potential strategic options typically include pivoting the business model, pursuing vertical or horizontal expansion, exploring strategic partnerships, or expanding into other geographic markets. Participants were given the case several days in advance to understand the issues at hand. Appendix B presents the full case.

The task structure followed Simon's decision-making model (Simon, 1947), excluding the agenda-setting stage[2]. Participants first identified potential factors causing the company's low profitability, then generated as many strategic options as possible and selected the best one. We

[1] Holding the strategic context constant in a single-case experiment increases internal validity by allowing us to isolate cognitive mechanisms that might be obscured by contextual variation. Single-case designs are common in strategy research for tracing decision processes (e.g., Gavetti & Rivkin, 2007; Tripsas & Gavetti, 2000). To begin exploring generalizability beyond this single case, we conducted a follow-up study with executive teams working on diverse strategic challenges from their own organizations (described in Appendix G).

[2] Agenda setting is often externally imposed in real-world organizational contexts—such as when strategic challenges are handed down by superiors or shaped by environmental constraints—and is therefore less amenable to experimental manipulation.

adopted this sequential design to isolate how LLM support introduced at different stages impacts decision outcomes. While problem formulation and ideation can be iterative in practice, there are also many real-world settings where these phases are deliberately staged or decoupled. For example, strategy consultants often present diagnostic analyses before co-developing solutions with clients. Similarly, design thinking and stage-gate innovation models encourage separating diagnosis from solution generation. These organizational practices highlight the value of studying the phases sequentially. Appendix A presents the full survey flow.

Participants were randomly assigned to three experimental groups, stratified by their section[3], gender, and a binary indicator of whether their last job was in consulting. The control group completed the task without LLM assistance (103 participants). The full treatment group received access to a GPT-4-based LLM from the problem formulation stage (102 participants). The partial treatment group received access to a GPT-4-based LLM from the ideation stage, after problem formulation (100 participants). Figure A.2 in the Appendix shows the assignment of GPT access across experimental groups. We provided suggestive prompts (see Appendix Table A.2 for details)[4] to all participants to ensure that any effects could be attributed to the presence of LLM rather than the framing provided by the prompts. Participants were allowed to modify the prompts as needed.

A potential fourth condition, where LLM-assisted problem formulation followed by human-only ideation, was deliberately excluded from the experimental design for two reasons. First, this study aims to examine how using LLMs during problem formulation sets a trajectory

[3] Students are pre-assigned to different sections by the MBA program administration office prior to and independent of the experiment.
[4] The prompts were carefully crafted through an iterative process to support the diagnostic task in problem formulation, which aimed at surfacing possible root causes, and the generative task in ideation, which aimed at producing possible solutions (Newell & Simon, 1972).

that might carry through to ideation, rather than isolating the effects of LLMs at each stage. Adding a fourth condition would distract from the goal and dilute statistical power, reducing the precision of inferences for our core research question. Second, such a condition would introduce interpretive ambiguity: any observed effects could stem from the influence of LLMs during problem formulation, from the removal of LLM support during ideation, or from their interaction. Enforcing a strict separation between stages would also be impractical, as participants who used LLMs in problem formulation might continue using them during ideation, which might confound treatment effects.

The experiments were run across four sections. In the first two sections, participants randomly assigned to treatment were instructed to use the standard ChatGPT web interface. For the latter two sections, we developed a custom wrapper built on OpenAI's API (gpt-4) and pre-loaded with the case information. For the pre-prompted version, we recorded all interaction logs between participants and the LLM (88 participants in total, with 47 from the full treatment group and 41 from the partial treatment group). Additional survey questions were administered to assess participants' perceptions of the task, familiarity with LLMs, and cognitive load (using a math question). The experiment took place within the classroom. Participants worked individually and were provided with unique links to an online survey interface to receive task instructions and submit their responses. Task instructions were provided sequentially, with participants receiving guidance for the subsequent task only after completing the previous stage. Participants completed the experimental task in an average of 44 minutes (M = 44.39, SD = 15.29). The experiment was pre-registered at the AEA RCT Registry and conducted in compliance with IRB approval from the authors' institution.

Outcome Measures

To characterize the nature of the alternatives participants generated, we distinguished between two types of managerial decisions. Drawing on Porter's (1996) distinction between strategy and operational effectiveness, we coded each option as strategic or operational. Strategic options involve long-term, high-level plans that are costly or difficult to reverse, creating path dependencies, and reflect choices about a firm's direction, scope, or core activities (Csaszar, 2018; Leiblein et al., 2018; Van den Steen, 2017). Examples from our data include "become a food delivery service," "acquire a company in reservation booking," and "expand to other regions."

Operational options, in contrast, focus on improving existing processes, capabilities, or efficiency to better utilize inputs (Porter, 1996). Examples include "reduce expenses related to sales and marketing," "increase customer retention by incrementing UX of the site," and "implement targeted sales training." All options are coded using a fine-tuned GPT-3.5 Turbo model trained on a dataset of human-coded options (see Appendix A for more details).

We use the term "strategic focus" to denote the proportion of strategic options participants generated. This measure reflects the degree to which a participant's ideation emphasizes options that involve long-term, high-impact considerations in competitive positioning, as opposed to incremental or operational refinements. This measure is not a normative judgment about the superiority of strategic over operational options, but rather serves as an assessment of how LLM assistance shifts participants' orientation in the decision-making process. Prior work shows that managers under pressure often prioritize immediate fixes (Aminov et al., 2019; Berwick, 2025), remain locked in incremental approaches (West et al., 2024), and default to established routines and operational actions (Mintzberg, 1994). The appropriate balance between strategic and incremental options may vary across organizational contexts, depending on factors such as environmental turbulence, resource constraints, or time horizons.

We aggregated the option-level coding to the participant level to calculate: (1) the total number of options, (2) the number of strategic options, (3) the number of operational options, (4) the proportion of strategic options, (5) the proportion of operational options. We also constructed a binary indicator of whether the participant's best option was strategic. Appendix D presents additional results based on whether an option represents a continuation of the current strategy.

Empirical Specification

To explore the impact of the treatment on decision-making, we employed the following specification:

$$y_{is} = \beta_0 + \beta_1 T_i^F + \beta_2 T_i^P + \gamma_s + \epsilon_i \quad (1)$$

where T_i^F is an indicator variable that equals 1 if participant i was assigned to the full treatment group, T_i^P is an indicator variable that equals 1 if participant i was assigned to the partial treatment group. γ_s controls for strata fixed effects, where each randomization stratum was defined as a combination of section, gender, and whether their last job was in consulting. This improves the precision of results by accounting for any section-specific factors or variation in treatment effects by gender or prior working experience (Cox & Reid, 2000). ϵ_i is an idiosyncratic error term. Models are estimated with robust standard errors at the participant level, which is the level of randomization (Abadie et al., 2023). The coefficients of interest are β_1, which captures the difference in the average outcome between the full treatment and control groups, and β_2, which captures the difference in the average outcome between the partial treatment and control groups. We also test whether the two coefficients are significantly different from each other by conducting a Wald test of the null hypothesis that $\beta_1 = \beta_2$. This test yields an F-statistic, which allows us to assess whether the partial and full treatments have statistically different impacts on the outcome of interest.

Supplementary Qualitative Data

To generate qualitative insight beyond the setting in the main experiment, we conducted a supplementary team-level study with 98 high-level executives enrolled in an executive education program. Participants were randomly assigned to 24 teams of four to five members. We randomly varied when the custom web interface (based on gpt-4o) was available at the team level: half of the teams received access for both the problem formulation and ideation stages (full treatment), while the other half gained access at the ideation stage (partial treatment). Unlike the main experiment, which used a standardized, pre-distributed case, each team worked on a different problem drawn from their own experience (see Appendix Table G.1 for more details). The group task lasted approximately two hours. Immediately after the group task, participants completed an individual reflection survey. Appendix G provides more details of the follow-up study.

MAIN RESULTS

The experimental results reveal a surprising puzzle: in line with prior research, LLM assistance increased the number of options generated; however, participants using LLMs for both problem formulation and ideation showed a reduced proportion of strategic options and were less likely to select a strategic option, an effect not observed when LLMs were used for ideation alone. This section unpacks these patterns in greater detail.

LLM assistance significantly increased the number of options generated in both treatment conditions. As shown in Column 1 of Table 2, participants in the full treatment group (LLM assistance in both problem formulation and ideation) generated 1.55 more options than the control group (a 27% increase over the control group mean, $p < 0.001$), while the partial treatment group

(LLM assistance only in ideation) generated 2.68 more (a 47% increase, $p < 0.001$). The partial treatment group also generated significantly more options than the full treatment group (F-test $p = 0.086$).

--INSERT TABLE 2 HERE --

This increase was evident for both strategic and operational options. Specifically, the full treatment group generated 0.53 more strategic options (a 16% increase, $p = 0.094$) and 1.02 more operational options (a 42% increase, $p < 0.001$) than the control group. Similarly, the partial treatment group generated 1.55 additional strategic options (a 48% increase, $p < 0.001$) and 1.13 more operational options (a 46% increase, $p = 0.001$). Importantly, the partial treatment had a significantly larger effect on generating strategic options compared to the full treatment.

Kolmogorov-Smirnov tests show significant rightward shifts in the distribution of the number of options (Figure 1(a)), strategic options (Figures 1(b)), and operational options (Figures 1(c)) for both treatment groups compared to the control group. Direct comparisons between two treatment groups also reveal meaningful differences in the number of strategic options generated (full vs. partial treatment K-S test $p = 0.058$), but not in the total number of options, or the number of operational options.

--INSERT FIGURE 1 HERE --

Examining the proportion of strategic and operational options further reveals the asymmetry in the two treatments. Column 4 of Table 2 shows that full treatment participants (LLM assistance in both problem formulation and ideation) generated 7 percentage points lower proportion of strategic options than the control group ($p = 0.042$). Given the binary coding of options as either strategic or operational, this implies a corresponding increase in the proportion of operational options (Column 5 of Table 2). These effects were not observed in the partial treatment condition: compared to the control, there was no significant difference in the proportion of strategic options ($p = 0.876$). This suggests that the inclusion of LLMs in the problem formulation stage reduces participants' emphasis on generating strategic options.

This pattern persists when examining option selection outcomes. Column 6 shows that participants in the full treatment group were 15 percentage points less likely to select a strategic option as their final choice ($p = 0.032$). In contrast, participants in the partial treatment group showed no significant difference from the control group. These findings suggest that the influence of LLMs during problem formulation carries downstream effects that shape both the generation and selection of strategic alternatives.

Taken together, these findings reveal a surprising pattern: LLM assistance during both problem formulation and ideation reduces strategic focus—both in the options generated and the choices made—whereas LLM assistance during ideation alone does not. To better understand these dynamics, we now explore cognitive mechanisms that may explain this puzzle.

POTENTIAL MECHANISMS

To unpack this puzzle, we adopt an abductive approach (Heckman & Singer, 2017; King et al., 2021), using the empirical results as a starting point for iterative theorizing. We turn to existing theory and explore three potential theoretical lenses—cognitive offloading (whether participants delegate thinking to AI), attention dilution (whether expanded problem framing dilutes attention), and cognitive drain (whether early-stage cognitive demands deplete later performance). The aim is not to "test" these mechanisms, but to probe the dynamics underlying the observed effects. Through this iterative process, a deeper insight emerged: when introduced during problem formulation, the AI-generated frame appears to have an anchoring effect on decision-makers, reducing their likelihood of critically interrogating or revising it. In this section, we show how each theoretical lens revealed partial insights before arriving at this more integrated understanding.

Cognitive offloading

One way to interpret the decline in strategic focus is through the lens of cognitive offloading—the process of shifting mental tasks to external tools or agents (Risko & Gilbert, 2016). Faced with complex and ambiguous tasks, individuals often offload memory, analysis, or reasoning to tools or collaborators to reduce mental demands. Recent field evidence by Dell'Acqua (n.d.) shows that recruiters working with high-quality AI tend to exert less effort and become less attentive to their tasks, leading to worsened performance compared to those with lower-quality AI. This effect may be more pronounced when LLM is introduced earlier. As individuals experience success with external aids in the early phases, their metacognitive beliefs shift, increasing the likelihood of future offloading (Risko & Gilbert, 2016). This may lead to increased reliance and reduced cognitive effort during ideation, resulting in lower strategic focus.

Several patterns are consistent with cognitive offloading. First, participants in the full treatment group spent significantly less time overall (–4.71 minutes) and during ideation specifically (–2.59 minutes), as shown in Columns 1 and 2 of Table 3, indicating reduced cognitive effort. Second, those with frequent prior LLM use showed a greater decline in strategic focus under full treatment (interaction effect = –0.12, column 3 of Table 3). While somewhat noisy, the result aligns with the offloading explanation: habitual users may be more prone to defer tasks to LLM, reducing personal cognitive engagement and, in turn, strategic focus. Third, problem statements in the full treatment group were more similar to one another than in the partial treatment group (Appendix Figure E.1, Full Treatment vs. Partial Treatment K-S Test p-value: 0.063), suggesting convergence on a shared LLM-influenced frame. This pattern implies that participants may follow a common framing from LLM, supporting the idea of cognitive offloading.

--INSERT TABLE 3 HERE --

However, the LLM reliance pattern challenges this explanation. If participants were offloading cognitive work to LLM, we would expect them to rely more heavily on LLM-generated content. Instead, we observe the opposite: full treatment participants actually relied less on LLM-generated ideas. On average, the full treatment condition has 32.7% options matched to the LLM output, while the partial treatment condition has 55.6% (t-test $p = 0.0176$), where "matched" is defined as sharing at least a 15-gram sequence with the LLM outputs. This pattern contradicts the prediction of cognitive offloading that people would increasingly defer to AI-generated content.

Additional behavioural patterns suggest that full treatment participants may have engaged with LLM in a different way. Specifically, the full treatment group initiated significantly more interactions with the LLM after formulating the problem (3.6 vs. 2.3 queries on average), suggesting deeper integration of the LLM into their decision process. Interestingly, despite more overall queries, the full treatment group asked fewer questions about options (1.1 vs. 1.9 queries related to options, Appendix Table E.1), but more about reflections and contextual insights. 21% of participants in the full treatment group asked the LLM to reflect on their behalf or validate their reasoning process for generating the options, compared to only 5% in the partial treatment group. These behaviours suggest a shift in how participants used the LLM: those in the full treatment group integrated it into their problem-solving workflow, whereas those in the partial treatment group primarily used it to generate content.

Taken together, these findings suggest something more profound than cognitive offloading. Although full treatment participants spent less time ideating and produced more similar problem formulations, they still generated original options. Rather than outsourcing cognitive work entirely, participants changed how they engaged with both the AI and the problem. This pattern prompts us to consider how LLM may also influence users' perceptions of what is relevant, thereby directing their attention in different ways.

Attention dilution

The attention dilution lens builds on the attention-based view of the firm (Cyert & March, 1963; Ocasio, 1997), which treats managerial attention as a scarce resource. Because of bounded rationality, decision makers cannot attend to all potentially relevant information simultaneously and must allocate attention across competing demands (Newell & Simon, 1972; Simon, 1947). If LLMs expand the problem space beyond what individuals would naturally generate, this added complexity may overwhelm their cognitive capacity and impair their judgment, leading to a decrease in strategic focus.

To test this explanation, we coded the problem formulation and options into different categories according to the keywords developed in Table E.2. The data provide mixed support for this account. As predicted, participants in the full treatment condition generated a significantly broader set of problem framings, as measured by the number of unique problem categories coded in their problem statements (Column 1, Table 4, a 13.8% increase over the partial group mean). However, this expanded problem space did not translate into more diffuse ideation. Participants in the full treatment group did not cover more categories in their ideas (Column 2, Table 4), nor did they generate fewer ideas per category (Column 3, Table 4). Both coefficients were small and statistically insignificant. Appendix E.2 provides more details.

We further tested whether the overlap between problem and option moderated the effect of LLM. If attention dilution were driving the decline in strategic focus, we would expect that participants with higher overlap—those who attempted to address more of the categories identified in their problem formulation—would show a greater decline in strategic focus under full LLM treatment. The idea here is that if LLMs introduce a wide range of categories that participants then attempt to follow up on, participants may attempt to address each one, leading to stretched cognitive bandwidth and reduced strategic focus. We defined overlap as the proportion of problem categories that participants subsequently addressed in their generated options (see Appendix E.2 for more details). However, the interaction terms between the full treatment binary variable and overlap were not statistically significant (Columns 4 and 5, Table 4), providing no empirical support for this attention dilution explanation.

--INSERT TABLE 4 HERE --

Overall, the results offer limited support for the attention dilution explanation. LLM-assisted problem formulation did expand the problem space. However, the predicted downstream effects on idea dispersion were not observed. This points to potential cognitive costs that might arise because of the expanded problem space, raising the question of whether early AI engagement creates other forms of cognitive costs.

Cognitive drain

The cognitive drain lens hypothesizes the depletion of mental resources caused by intensive early-stage cognitive activity, which leaves fewer resources available for subsequent stages of decision-making (Baumeister et al., 1998; Vohs et al., 2008). When early-stage tasks require substantial cognitive effort, downstream performance in reasoning, creativity, or evaluation can suffer (Pignatiello et al., 2020). For instance, engaging in complex choices or sustained attention tasks has been shown to impair self-regulation and higher-order cognition in subsequent tasks (Baumeister et al., 1998; Vohs et al., 2008). From this perspective, if LLMs introduced during problem formulation increase cognitive demands, which might happen due to processing LLM-generated framings, integrating LLM-generated content with existing knowledge, or managing the cognitive overhead of human-AI interaction, this heightened early-stage effort may have depleted participants' cognitive resources, leaving few available for later tasks. By the time they reached the ideation stage, they may have effectively "run out of steam," thus defaulting to less cognitively demanding operational options instead of generating more strategic ones, which typically require greater abstraction and creative synthesis.

However, the empirical evidence offers limited support for this explanation. First, participants in the full treatment condition did not spend significantly more time on problem formulation compared to the partial treatment group (Column 1 in Table 5), contradicting the prediction of increased early-stage cognitive demands. Second, while participants in the full treatment condition produced shorter responses during ideation as measured by the average word count per idea (Column 2 in Table 5), this might also reflect efficiency rather than resource depletion. Third, and most importantly, participants in the full treatment condition completed the test more quickly (Column 3 in Table 5) without sacrificing accuracy (Column 4 in Table 5). This pattern contradicts the cognitive drain explanation, which would predict slower completion times or reduced accuracy on cognitively demanding tasks due to depleted mental resources.

--INSERT TABLE 5 HERE --

If cognitive drain were strongly at play, we would also expect a progressive decline in strategic quality over time. Yet, we observed no significant deterioration in the strategic-orientation of options as the ideation task progressed (Column 7 in Table 5). Comparing the strategic-orientation of the first versus the last option also revealed no significant decline in the full treatment group (Column 6 in Table 5). These findings suggest that while participants may be producing shorter options overall, the decline does not necessarily come with a decline in strategic orientation as the task progresses. Appendix E.3 provides more details

Finally, the cognitive drain explanation would predict that participants with more domain expertise (e.g., consulting experience) should be less susceptible to this drain, as they can process problem framings more efficiently, reducing the cognitive cost of early engagement. Moderator analyses provided limited support for the role of expertise in buffering cognitive drain. The interaction effect is noisy (Column 5 in Table 5).

In summary, the evidence supports a limited cognitive drain effect. While participants in the full treatment group produced shorter options, there was no additional evidence suggesting that cognitive resources are depleted.

How LLM Shapes Strategic Cognition

After cycling through different theoretical lenses, the indirect evidence from multiple measures suggests that the observed effects are not fully explained by simple offloading, diluted attention, or cognitive drain. Cognitive offloading aligns with the reduced time spent and convergent framings in the full treatment group, but it cannot explain the higher proportion of independently generated ideas. Attention dilution is consistent with the broadened problem space but not with the absence of dispersed ideation. Cognitive drain might explain the shorter responses in the full treatment, but it fails to show other consistent evidence.

Synthesizing the empirical patterns points to an alternative explanation: LLMs provide comprehensive problem formulations that lead participants to anchor on the LLM-generated problem frame, which is consistent with the expanded yet convergent problem formulation. In addition, full treatment participants were more likely to delegate reflective reasoning to the LLM (21% vs 5% reflection questions in Appendix Table E.1) and return to LLM-generated framings throughout the task repeatedly. As one full treatment participant observed, "it [LLM took] me back to what it thought was the right answer," while another noted: "It helped us in coming back to the core problems again and again." Similar patterns were also seen in teams: “It helped us in coming back to the core problems again and again as answers by AI is concise” (Group 1, FT).

Qualitative evidence suggests that the perceived completeness of LLM-generated problem frames drives this anchoring. As participants in teams observed: “The AI often provided well-framed, familiar strategic models, which made them seem credible and easy to accept at face value” (Group 2, FT). Another team noted that they accepted LLM’s problem frame “mostly because the ideas felt well-articulated and aligned with our existing understanding of the problem” (Group 6, FT), and noted that “some of us were quick to accept ChatGPT’s suggestions because they sounded polished” (Group 6, FT).

However, when participants anchor to these frames, they tend to accept them as given and generate options within those boundaries. Even though the LLM’s framing spanned many categories, participants tended to elaborate within that predefined structure rather than introducing new dimensions. This pattern is evident in our data (Appendix Table E.6): on average, participants in the full treatment condition explored 0.48 fewer categories beyond their initial problem formulation (a 29.6% decrease over the partial group mean, $p = 0.009$). As one participant observed: “The group often agreed with AI’s suggestions without much questioning... though this occasionally limited deeper critical exploration” (Group 1, FT). As another group reflected: “During the ideation stage, what helped our group generate and focus on strategic options was the clarity we had from a well-defined problem… our options naturally aligned with those themes” (Group 6, FT). This anchoring is consequential: strategic options typically emerge from questioning the problem formulation itself (Csaszar & Levinthal, 2016; Gavetti & Rivkin, 2007).

Even when the LLM provides a wide-ranging frame, it may lock participants into a particular way of seeing the problem. As a result, decision-makers may generate more options, but fewer that challenge the fundamental premises of the situation, reducing the likelihood of true strategic alternatives.

In contrast, partial treatment participants formulated the problem themselves, retaining implicit knowledge of the assumptions and trade-offs made. For example, one partially treated participant described their problem formulation process as “Structured as 'what' (i.e. what I observed) and 'why' (i.e. hypotheses for why that could be the case, based on data),” which highlights their awareness of the underlying assumptions and causal links of the problems. Another partially treated participant noted that “To determine the strategic causes I have used all exhibits to formulate hypothesis, and then used Stakeholder’s interviews to materialize those hypothesis,” suggesting that the process of triangulating multiple sources helped maintain their sensitivity to the trade-offs involved in their chosen framing. Similarly, teams showed flexibility in reframing the problem: “The clarity in problem definition emerged through structured analysis, open dialogue and a willingness to reframe the issue” (Group 22, PT), while another noted: “I initially outlined the broad problem, holding back from leading too much so that others could shape the direction. This encouraged diverse perspectives” (Group 17, PT).

With their self-generated problem frames, these participants seemed to maintain clearer cognitive boundaries and LLM as an additional input source. For example, one participant noted, “I had 2-3 solutions at first, but ChatGPT [gave] me additional 2-3 options that I overlooked,”. Another reflected that “It listed quite a few other avenues I hadn't thought about,” positioning the LLM as expanding rather than defining the solution space. Several partial treatment participants also described using the LLM as a supplementary input. One participant noted: “I believe we were initially fixated on just one solution. But, based on the AI tool probs we questioned more and came up to discussing deeper level solutions" (Group 22, PT). Similarly, another shared: "We used it more as a checker mechanism of sorts to ensure we aren't missing anything rather than relying on it for creative solutioning" (Group 18, PT).

Taken together, these findings suggest that when LLMs provide comprehensive problem formulation, participants tend to anchor on these frames, which may constrain the likelihood of stepping outside predefined boundaries to reframe the question. Since strategic options often arise from questioning core assumptions (Csaszar & Levinthal, 2016; Gavetti & Rivkin, 2007), this might explain the observed reduction in strategic focus. While exploratory, these results offer a foundation for future research to more precisely identify and test the mechanisms by which LLM framing shapes strategic thinking.

DISCUSSION

In this paper, we documented an interesting puzzle that emerged from an experiment with 305 MBA students and 98 executives. Although LLM assistance broadens the scope of alternative generation, having access to LLM in both formulation and ideation appears to decrease strategic focus, whereas access limited to ideation does not.

To unpack this puzzle, we adopt an abductive approach (King et al., 2020), using empirics for iterative theorizing (Heckman & Singer, 2017). Our data provide initial evidence of altered behaviours when LLMs are introduced during problem formulation: expanded framings, reduced time investment, and shifted interaction patterns. While we cannot definitively identify the cognitive mechanisms underlying these changes, we propose that when introduced during problem formulation, the LLM-generated frame appears to offer a sense of completeness that anchors subsequent reasoning, shaping the options participants generate and select.

Our findings offer a cognitive lens through which to understand and build effective human-AI collaboration in strategic decision-making. While organizational design elements such as task allocation and sequencing provide valuable insights (Choudhary et al., 2023; He et al., 2023; Puranam, 2021), our study emphasizes the role of managerial cognition in shaping strategic decisions and outcomes (Gavetti & Levinthal, 2000; Laureiro-Martínez & Brusoni, 2018). Our approach also contributes to emerging work on how humans perceive and cognitively engage with AI in organizational settings (Choudhury et al., 2024; Vanneste & Puranam, 2024), extending these insights from perception and delegation to representation and reasoning.

More broadly, this study contributes to ongoing discussions around the effectiveness of LLMs in strategic decision-making (Boussioux et al., 2024; Csaszar et al., 2024; Doshi et al., 2024; Otis et al., 2024). Our findings suggest that their influence depends critically on when they are introduced and how individuals cognitively engage with them. By highlighting the representational consequences of early LLM interaction, we offer a nuanced perspective on how AI assistance shapes the mental models that guide strategic behaviours.

Limitations and Research Agenda

Using an abductive approach, this study offers a cognitive perspective on how LLMs influence strategic decision-making and lays the groundwork for a broader research agenda on AI-augmented strategic cognition. While conducted in a controlled setting and subject to several limitations, it serves as a starting point to open new lines of inquiry into how AI may reshape the fundamental processes of strategic decision-making.

From classroom to field: Validity and context

One limitation of this study is that the main experiment involved MBA students working individually in a classroom environment, which may not fully capture how experienced managers or executives engage with LLMs in high-stakes, real-world scenarios. While this raises questions about external validity, it is worth noting that MBA students in our sample were experienced professionals with an average age of 30, and 31% worked in consulting. These participants have been widely used in research on decision-making and organizational behavior, especially when the goal is to understand generalizable cognitive processes (Frechette, 2015; Levine et al., 2023; Stevenson & Josefy, 2019).

A related limitation concerns the realism and contextual scope of the task environment. Decision-making in classroom-based studies typically occurs under lower stakes and without the accountability pressures in real-world strategic contexts. These conditions may shape how participants engage with LLMs, and should be considered when interpreting the boundary conditions of our findings. Nonetheless, laboratory-style experiments are increasingly used in the study of strategic decision-making because they allow researchers to isolate causal relationships and control for confounding variables (Billinger et al., 2021; Gary et al., 2012; Laureiro-Martínez et al., 2015; Shapira & Shaver, 2014; Turner & Makhija, 2012). In our abductive approach, the controlled conditions help surface emergent patterns that might otherwise be obscured in the field by uncontrolled influences (Falk & Heckman, 2009). The main experiment relied on a single case and a sequential, non-iterative task structure, which allowed for tight control over complexity and comparability but limited the variation in contexts, problem types, and decision processes observed. We also excluded agenda-setting to ensure comparability and used selection only as a dependent measure, which leaves open how LLMs influence earlier or later stages of decision-making.

Our supplementary study involving senior executives working in teams on real strategic challenges across diverse sectors (see Appendix Table E.1) helps ease these concerns by expanding the range of problems and shifting the focus from individuals to teams. However, its small sample size precludes robust statistical conclusions, and it remains within the classroom setting. Field research is needed to further test these dynamics and examine how different organizational contexts may moderate the cognitive effects of LLMs.

Cognitive mechanisms: Opening the black box

While our abductive analysis provides preliminary evidence for plausible mechanisms, it does not yield definitive answers. These mechanisms remain important avenues for investigation, as our evidence is indirect, and they may interact with each other in shaping strategic thinking. Future research should use targeted experiments to test these mechanisms, identify causal pathways, and explore boundary conditions. Additional factors such as the content of AI output, its perceived authority, and the cognitive stance users adopt toward machine assistance (e.g., trust, emotional responses) should also be systematically explored.

While our study used GPT-based models, the underlying cognitive mechanisms we identify likely persist across different architectures. Nonetheless, differences in training data, domain coverage, and interface design may shape how they manifest. This challenge is shared across much of the emerging literature on LLMs and decision-making (Boussioux et al., 2024; Csaszar et al., 2024; Dell'Acqua et al., 2023; Doshi et al., 2025; Doshi & Hauser, 2024; Otis et al., 2024). Future research should compare across models and interaction modalities to establish boundary conditions for our findings and explore whether the mechanisms operate consistently in different contexts.

Future work could also compare LLM-supported decision-making with decision-making aided by trained human collaborators under equivalent informational constraints, to isolate whether delegation to AI engages distinct cognitive pathways or simply amplifies dynamics common to human-assisted decision-making. Another promising direction is to vary the quantity and diversity of root causes identified during problem formulation—whether generated independently or with LLM assistance—to assess how such “front-loading” of the problem space shapes subsequent solution generation. In addition, extending beyond a single snapshot, longitudinal designs could track whether sustained AI use augments or substitutes for human strategic thinking, moving beyond static assessments of AI accuracy or productivity toward dynamic theories of cognitive co-evolution that account for reciprocal adaptation between human and AI.

Concluding Remark

Using an abductive approach, we examined why LLM assistance in both problem formulation and ideation reduced the likelihood of generating and selecting a strategic option compared to ideation alone. Our abductive analysis suggests that decision-makers may anchor on the LLM-generated frame and stay within its boundaries, leading to fewer strategic options considered and chosen.

LLM technologies are evolving rapidly, yet we know little about how organizations can deploy them most effectively. Our findings indicate that their effects vary not only across tasks (Dell’Acqua et al., 2023) but also with the timing and manner of their integration into workflows. Effective use of LLMs requires substantial know-how about timing, domain-specific applications, and cognitive integration. Despite the apparent ease and reduced costs of using LLM, significant frictions may hinder realizing their full potential. Moreover, our results highlight the need to understand how LLMs shape human thinking: even when human decision-makers do not lessen cognitive effort, there may be unintended consequences that shift what they attend to and how they reason.

REFERENCES

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35.

Aminov, I., De Smet, A., Jost, G., & Mendelsohn, D. (2019, April 30). *Decision making in the age of urgency*. McKinsey & Company. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/decision-making-in-the-age-of-urgency

Baer, M., Dirks, K. T., & Nickerson, J. A. (2013). Microfoundations of strategic problem formulation. *Strategic Management Journal*, *34*(2), 197–214.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–1265.

Berwick, I. (2025, August 7). *Lessons in leadership: Facing up to the 'perfect storm.'* https://www.ft.com/content/3056919b-a8b6-4f47-a2e0-913f18b63e0e

Billinger, S., Srikanth, K., Stieglitz, N., & Schumacher, T. R. (2021). Exploration and exploitation in complex search tasks: How feedback influences whether and where human agents search. *Strategic Management Journal*, *42*(2), 361–385. https://doi.org/10.1002/smj.3225

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The Crowdless Future? Generative AI and Creative Problem-Solving. *Organization Science*. https://doi.org/10.1287/orsc.2023.18430

Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. (2025). Human-AI Ensembles: When Can They Work? *Journal of Management*, *51*(2), 536–569. https://doi.org/10.1177/01492063231194968

Cox, D. R., & Reid, N. (2000). *The Theory of the Design of Experiments*. CRC Press.

Csaszar, F. A. (2018). What Makes a Decision Strategic? Strategic Representations. *Strategy Science*, *3*(4), 606–619. https://doi.org/10.1287/stsc.2018.0067

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors. *Strategy Science*, *9*(4), 322–345. https://doi.org/10.1287/stsc.2024.0190

Csaszar, F. A., & Laureiro-Martínez, D. (2018). Individual and Organizational Antecedents of Strategic Foresight: A Representational Approach. *Strategy Science*, *3*(3), 513–532. https://doi.org/10.1287/stsc.2018.0063

Csaszar, F. A., & Levinthal, D. A. (2016). Mental representation and the discovery of new strategies. *Strategic Management Journal*, *37*(10), 2031–2049. https://doi.org/10.1002/smj.2440

Cyert, R., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Prentice-Hall.

Dell'Acqua, F. (2022). Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters. *Working Paper*.

Dell'Acqua, F., Ayoubi, C., Lifshitz-Assaf, H., Sadun, R., Mollick, E. R., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., & Lakhani, K. R. (2025). *The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise* (SSRN Scholarly Paper 5188231). https://doi.org/10.2139/ssrn.5188231

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (SSRN Scholarly Paper 4573321). https://papers.ssrn.com/abstract=4573321

Doshi, A. R., Bell, J. J., Mirzayev, E., & Vanneste, B. S. (2025). Generative artificial intelligence and evaluating strategic decisions. *Strategic Management Journal*, *46*(3), 583–610.

Falk, A., & Heckman, J. J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science*, *326*(5952), 535–538. https://doi.org/10.1126/science.1168244

Frechette, G. (2015). Laboratory Experiments: Professionals versus Students. In G. Frechette & A. Schotter (Eds.), *Handbook of Experimental Economic Methodology* (pp. 360–390). Oxford University Press.

Gary, M. S., Wood, R. E., & Pillinger, T. (2012). Enhancing mental models, analogical transfer, and performance in strategic decision making. *Strategic Management Journal*, *33*(11), 1229–1246. https://doi.org/10.1002/smj.1979

Gavetti, G. (2012). PERSPECTIVE—Toward a Behavioral Theory of Strategy. *Organization Science*, *23*(1), 267–285. https://doi.org/10.1287/orsc.1110.0644

Gavetti, G., & Levinthal, D. (2000). Looking Forward and Looking Backward: Cognitive and Experiential Search. *Administrative Science Quarterly*, *45*(1), 113–137.

Gavetti, G., Levinthal, D. A., & Rivkin, J. W. (2005). Strategy making in novel and complex worlds: The power of analogy. *Strategic Management Journal*, *26*(8), 691–712.

Gavetti, G., & Rivkin, J. W. (2007). On the Origin of Strategy: Action and Cognition over Time. *Organization Science*, *18*(3), 420–439. https://doi.org/10.1287/orsc.1070.0282

Heckman, J. J., & Singer, B. (2017). Abducting Economics. *American Economic Review*, *107*(5), 298–302. https://doi.org/10.1257/aer.p20171118

Heshmati, M., & Csaszar, F. A. (2023). Learning Strategic Representations: Exploring the Effects of Taking a Strategy Course. *Organization Science*.

King, A., Goldfarb, B., & Simcoe, T. (2021). Learning from Testimony on Quantitative Research in Management. *Academy of Management Review*, *46*(3), 465–488.

Kruse, S., Bendig, D., & Brettel, M. (2023). How Does CEO Decision Style Influence Firm Performance? The Mediating Role of Speed and Innovativeness in New Product Development. *Journal of Management Studies*, *60*(5), 1205–1235.

Lamberg, J.-A., & Peltoniemi, M. (2020). The nanoeconomics of firm-level decision-making and industry evolution: Evidence from 200 years of paper and pulp making. *Strategic Management Journal*, *41*(3), 499–529. https://doi.org/10.1002/smj.3080

Laureiro-Martinez, D., Arrieta, J. P., & Brusoni, S. (2023). Microfoundations of Problem Solving: Attentional Engagement Predicts Problem-Solving Strategies. *Organization Science*, *34*(6), 2207–2230. https://doi.org/10.1287/orsc.2019.13213

Laureiro-Martínez, D., & Brusoni, S. (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal*, *39*(4), 1031–1058. https://doi.org/10.1002/smj.2774

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*(3), 319–338.

Leiblein, M. J., Reuer, J. J., & Zenger, T. (2018). What Makes a Decision Strategic? *Strategy Science*, *3*(4), 558–573. https://doi.org/10.1287/stsc.2018.0074

Levine, S. S., Schilke, O., Kacperczyk, O., & Zucker, L. G. (2023). Primer for Experimental Methods in Organization Theory. *Organization Science*.

March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.

Meincke, L., Girotra, K., Nave, G., Terwiesch, C., & Ulrich, K. T. (2024). *Using Large Language Models for Idea Generation in Innovation* (SSRN Scholarly Paper 4526071).

Mintzberg, H. (1994). The Fall and Rise of Strategic Planning. *Harvard Business Review*. https://hbr.org/1994/01/the-fall-and-rise-of-strategic-planning

Mitroff, I. I., & Featheringham, T. R. (1974). On systemic problem solving and the error of the third kind. *Behavioral Science*, *19*(6), 383–393. https://doi.org/10.1002/bs.3830190605

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Prentice-Hall.

Nickerson, J., & Argyres, N. (2018). Strategizing Before Strategic Decision Making. *Strategy Science*, *3*(4), 592–605. https://doi.org/10.1287/stsc.2018.0066

Novelli, E., & Spina, C. (2024). Making business model decisions like scientists: Strategic commitment, uncertainty, and economic performance. *Strategic Management Journal*, *45*(13). https://doi.org/10.1002/smj.3636

Ocasio, W. (1997). Towards an Attention-Based View of the Firm. *Strategic Management Journal*, *18*, 187–206.

Otis, N., Clarke, R., Delecourt, S., Holtz, D., & Koning, R. (2024). *The Uneven Impact of Generative AI on Entrepreneurial Performance* (SSRN Scholarly Paper 4671369). Social Science Research Network. https://papers.ssrn.com/abstract=4671369

Park, C. H. (2024). Finding a road less traveled: Combining analysis and intuition to develop novel problem formulations. *Strategic Management Journal*, *45*(11), 2368–2392.

Park, C. H., & Baer, M. (2022). Getting to the Root of Things: The Role of Epistemic Motivation and Construal Levels in Strategic Problem Formulation. *Strategy Science*, *7*(4), 284–299. https://doi.org/10.1287/stsc.2022.0155

Park, C. H., Baer, M., & Nickerson, J. (2025). Looking at the Trees to See the Forest: Construal Level Shift in Strategic Problem Framing and Formulation. *Organization Science*. https://doi.org/10.1287/orsc.2024.19134

Pignatiello, G. A., Martin, R. J., & Hickman, R. L. (2020). Decision Fatigue: A Conceptual Analysis. *Journal of Health Psychology*, *25*(1), 123–135.

Porter, M. (1996). *What Is Strategy?* https://hbr.org/1996/11/what-is-strategy

Puranam, P. (2021). Human–AI collaborative decision-making as an organization design problem. *Journal of Organization Design*, *10*(2), 75–80.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Shapira, Z., & Shaver, J. M. (2014). Confounding changes in averages with marginal effects: How anchoring can destroy economic value in strategic investment assessments. *Strategic Management Journal*, *35*(10), 1414–1426. https://doi.org/10.1002/smj.2165

Simon, H. (1947). *Administrative behavior; a study of decision-making processes in administrative organization*.

Stevenson, R. M., & Josefy, M. (2019). Knocking at the gate: The path to publication for entrepreneurship experiments through the lens of gatekeeping theory. *Journal of Business Venturing*, *34*(2), 242–260. https://doi.org/10.1016/j.jbusvent.2018.10.008

Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, *21*(10–11), 1147–1161.

Turner, K. L., & Makhija, M. V. (2012). The role of individuals in the information processing perspective. *Strategic Management Journal*, *33*(6), 661–680. https://doi.org/10.1002/smj.1970

Van den Steen, E. (2017). A Formal Theory of Strategy. *Management Science*, *63*(8), 2616–2636. https://doi.org/10.1287/mnsc.2016.2468

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, *94*(5), 883–898. https://doi.org/10.1037/0022-3514.94.5.883

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, *7*(9), 1526–1541. https://doi.org/10.1038/s41562-023-01659-w

West, A., Koller, T., & Bhargava, R. (2024, May 20). *Tying short-term decisions to long-term strategy*. McKinsey & Company. https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/tying-short-term-decisions-to-long-term-strategy

TABLES

Table 1: Descriptive statistics by experimental group

	(1) Full sample				(2) Full Treatment		(3) Partial Treatment		(4) Control	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD
Baseline Variables										
Female	0.32	0.47	0	1	0.37	0.49	0.30	0.46	0.28	0.45
Consulting	0.31	0.47	0	1	0.32	0.47	0.31	0.46	0.31	0.47
Age	30.07	2.47	25	37	30.15	2.57	30.22	2.36	29.84	2.48
Main Outcome Variables										
Number of Options	7.12	4.09	1	40	7.31	4.05	8.41	5.23	5.68	1.81
Number of Strategic Options	3.96	2.73	0	22	3.83	2.83	4.85	3.25	3.23	1.65
Number of Operational Options	3.16	2.33	0	18	3.48	2.24	3.56	2.90	2.45	1.53
Proportion of Strategic Options	0.56	0.22	0	1	0.51	0.23	0.58	0.20	0.57	0.24
Proportion of Operational Options	0.44	0.22	0	1	0.49	0.23	0.42	0.20	0.43	0.24
Best Option is Strategic	0.60	0.49	0	1	0.51	0.50	0.65	0.48	0.65	0.48
Best Option is Operational	0.40	0.49	0	1	0.49	0.50	0.35	0.48	0.35	0.48
Duration (min)	44.39	15.29	5	167	42.23	12.77	46.80	11.48	44.19	19.93
Observations	305				102		100		103	

Notes: This table reports the summary statistics of baseline variables and main outcome variables across experimental groups. Baseline variables include gender, consulting background, and age. Main outcome variables include the total number of options generated, the number and proportion of options coded as strategic or operational, and whether the participant's selected "best" option was strategic or operational. Coding was performed using a fine-tuned GPT-3.5 Turbo model trained on human-coded examples (see Appendix C for details). Observations reflect the number of participants per condition.

Table 2: Impact on ideation and selection

VARIABLES	(1) Number of Options	(2) Number of Strategic Options	(3) Number of Operational Options	(4) Proportion of Strategic Options	(5) Proportion of Operational Options	(6) Best Option is Strategic
Full Treatment	1.55***	0.53*	1.02***	-0.07**	0.07**	-0.15**
	(0.43)	(0.31)	(0.27)	(0.03)	(0.03)	(0.07)
Partial Treatment	2.68***	1.55***	1.13***	0.00	-0.00	-0.01
	(0.55)	(0.36)	(0.33)	(0.03)	(0.03)	(0.07)
Constant	5.16***	3.00***	2.16***	0.56***	0.44***	0.62***
	(0.59)	(0.43)	(0.30)	(0.04)	(0.04)	(0.09)
Observations	305	305	305	305	305	305
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.08	0.07	0.06	0.02	0.02	0.01
Control Group Mean	5.68	3.23	2.45	0.57	0.43	0.65
F-test	2.97	5.79	0.09	5.65	5.65	4.47
Prob > F	0.086	0.017	0.766	0.018	0.018	0.035

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 3: Exploratory analyses related to cognitive offloading

VARIABLES	(1) Duration (min)	(2) Time Spent on Ideation (min)	(3) Proportion of Strategic Options	(4) Best Option is Strategic
Full Treatment	-4.71***	-2.59***	-0.03	-0.09
	(1.69)	(0.55)	(0.04)	(0.09)
Use LLM Daily			0.07	0.17*
			(0.04)	(0.09)
Treatment x Use LLM Daily			-0.12*	-0.20
			(0.07)	(0.14)
Constant	47.09***	6.16***	0.54***	0.51***
	(2.74)	(0.86)	(0.06)	(0.12)
Observations	202	202	202	202
Gender-Consulting-Section FE	YES	YES	YES	YES
Adjusted R-squared	0.09	0.14	0.04	0.04
Partial Group Mean	46.80	8.28	0.58	0.65

Notes: This table reports regression results for $y_{is} = \beta_0 + \beta_1 T_i^F + \gamma_s + \epsilon_i$ and $y_{is} = \beta_0 + \beta_1 T_i^F + \beta_2 Moderator_i + \beta_3 T_i^F Moderator_i + \gamma_s + \epsilon_i$. It includes the full treatment group and the partial treatment group only to compare the difference between these two conditions. All regression models include fixed effects for the randomization strata (gender-consulting-section). Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 4: Exploratory analyses related to attention dilution

VARIABLES	(1) Number of Categories Covered in Problem	(2) Number of Categories Covered in Options	(3) Average Options per Category	(4) Proportion of Strategic Options	(5) Best Option is Strategic
Full Treatment	0.70**	-0.27	-0.06	-0.16*	-0.17
	(0.30)	(0.24)	(0.08)	(0.09)	(0.17)
Overlap				-0.12	-0.14
				(0.09)	(0.18)
Full Treatment x Overlap				0.17	0.03
				(0.14)	(0.28)
Constant	4.84***	4.21***	1.56***	0.63***	0.65***
	(0.49)	(0.34)	(0.11)	(0.07)	(0.15)
Observations	202	202	202	202	202
Gender-Consulting-Section FE	YES	YES	YES	YES	YES
Adjusted R-squared	0.01	-0.01	0.04	0.03	0.03
Partial Group Mean	5.07	4.41	1.77	0.58	0.65

Notes: This table reports regression results for $y_{is} = \beta_0 + \beta_1 T_i^F + \gamma_s + \epsilon_i$ and $y_{is} = \beta_0 + \beta_1 T_i^F + \beta_2 Moderator_i + \beta_3 T_i^F Moderator_i + \gamma_s + \epsilon_i$. It includes the full treatment group and the partial treatment group only to compare the difference between these two conditions. All regression models include fixed effects for the randomization strata (gender-consulting-section). Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 5: Exploratory analyses related to cognitive drain

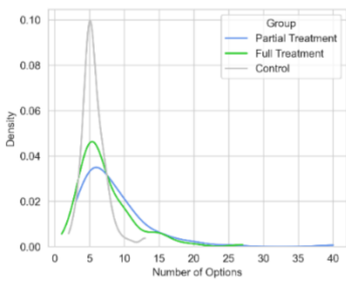
VARIABLES	(1) Time Spent on Problem Formulation (min)	(2) Average Word Count per Option	(3) Time Spent on Math Q (min)	(4) Accurate Math Q	(5) Proportion of Strategic Options	(6) Delta Strategic	(7) Strategic (Option level)
Full Treatment	-1.10	-3.48*	-0.56**	0.02	-0.05	0.09	-0.36**
	(1.66)	(2.07)	(0.26)	(0.05)	(0.04)	(0.10)	(0.17)
Consulting					0.03		
					(0.07)		
Treatment x Consulting					-0.07		
					(0.06)		
Index							0.01
							(0.01)
Full Treatment x Index							0.03
							(0.02)
Constant	29.38***	15.56***	2.74***	0.87***	0.55***	0.08	0.27**
	(2.61)	(3.41)	(0.31)	(0.07)	(0.05)	(0.07)	(0.11)
Observations	202	202	202	202	202	202	1,591
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES	NO
Adjusted R-squared	0.08	-0.01	0.03	0.03	0.03	-0.00	-
Partial Group Mean	24.19	21.42	3.26	0.82	0.58	0.08	0.58

Notes: This table reports regression results for $y_{is} = \beta_0 + \beta_1 T_i^F + \gamma_s + \epsilon_i$ and $y_{is} = \beta_0 + \beta_1 T_i^F + \beta_2 Moderator_i + \beta_3 T_i^F Moderator_i + \gamma_s + \epsilon_i$. Column 6 reports the regression results for $FirstOptionStrategic_{is} - LastOptionStrategic_{is} = \beta_0 + \beta_1 T_i^F + \gamma_s + \epsilon_i$. Columns 1-6 include fixed effects for the randomization strata (gender-consulting-section). Column 7 reports a multilevel logistic regression predicting whether an individual option is coded as strategic (at the option level). The key interaction term tests whether the likelihood of generating a strategic option changes over the course of ideation and whether this change differs between the full and partial treatment groups. All columns include the full treatment group and the partial treatment group only to compare the difference between these two conditions. Columns 1-6 report robust standard errors and column 7 reports standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

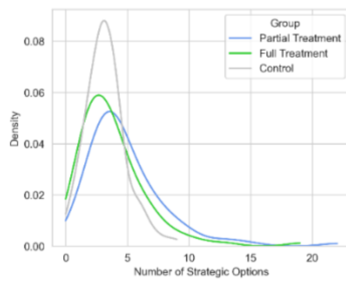
FIGURES

Figure 1: Distribution of Ideation Measures

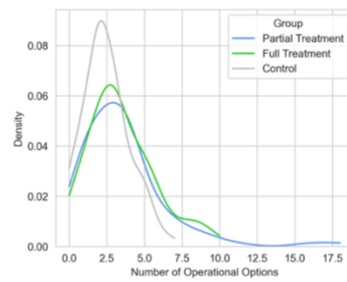
(a) Total number of options



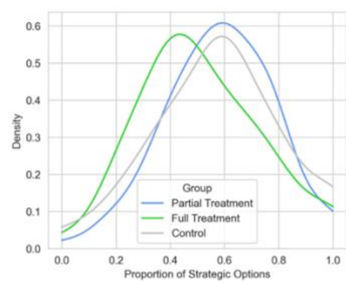
(b) Number of strategic options



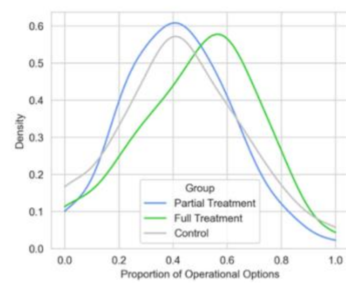
(c) Number of operational options



(d) Proportion of strategic options



(e)Proportion of operational options



Notes: These figures show the density distribution of outcome variables of the Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Panel (a) shows the distribution of the number of alternatives generated. Control group versus Full Treatment group Kolmogorov–Smirnov Test p-value: 0.014; Control group versus Partial Treatment group Kolmogorov–Smirnov Test p-value: 0.000; Full Treatment group versus Partial Treatment group Kolmogorov–Smirnov Test p-value: 0.332. Panel (b) shows the distribution of the number of strategic options generated. Control group versus Full Treatment group Kolmogorov–Smirnov Test p-value: 0.331; Control group versus Partial Treatment group Kolmogorov–Smirnov Test p-value: 0.001; Full Treatment group versus Partial Treatment group Kolmogorov–Smirnov Test p-value: 0.058. Panel (c) shows the distribution of the number of operational options generated. Control group versus Full Treatment group Kolmogorov–Smirnov Test p-value: 0.017; Control group versus Partial Treatment group Kolmogorov–Smirnov Test p-value: 0.0144; Full Treatment group versus Partial Treatment group Kolmogorov–Smirnov Test p-value: 0.712. Panel (d) shows the distribution of the proportion of strategic options. Control vs. Full Treatment K-S Test p-value: 0.027; Control vs. Partial Treatment K-S Test p-value: 0.852; Full Treatment vs. Partial Treatment K-S Test p-value: 0.02. Panel (e) shows the distribution of the proportion of operational options. Control vs. Full Treatment K-S Test p-value: 0.027; Control vs. Partial Treatment K-S Test p-value: 0.852; Full Treatment vs. Partial Treatment K-S Test p-value: 0.02.

APPENDIX A: EXPERIMENT DETAILS AND SURVEY DESIGNS

This appendix provides additional information on the experiment and survey designs. We begin by presenting the full survey flow used in the full treatment condition, which included LLM assistance during both problem formulation and ideation. We include only this version to conserve space, as the survey structure was consistent across conditions, with variations only in the timing of LLM access. Full versions for other conditions are available upon request.

Then, Figure A.1 shows the custom web interface used in the experiment. Figure A.2 shows the assignment of treatment. Figure A.3 shows the distribution of duration. Figure A.4 shows the histograms of ideation measures.

Table A.1 shows the allocation of participants across experimental conditions and classroom sections. Table A.2 shows the suggested prompts used for problem formulation and ideation. Table A.3 shows the pairwise comparisons of baseline and outcome variables across experimental groups. Table A.4 reports the main results after excluding observations with durations more than two standard deviations from the mean.

Problem

To complete this exercise, please use a Large Language Model (LLM) - an AI-based technology - which you can access here.

To start a conversation with the LLM, follow the instructions below:

- Please use your unique User ID: ${e://Field/user_id} to log in. Your initial password is the same as your User ID.
- Type your questions into the text box as if conversing with a person, but avoid overly long paragraphs due to length limits.
- After typing the question, press enter. In a moment, it will reply.
- Note that the LLM is not a search engine; verify the accuracy of its responses.
- Your conversations will be saved and analyzed in anonymized form by the research team.

Below is a suggestive prompt for identifying the causes of problems. Please note that it serves as a guideline. You might need to adjust your prompts for more precise results.

"In this task, you will be playing the role of a strategic management consultant. Your task is to identify and list strategic factors that contribute to the profit decline of an online restaurant review platform. Your response should be in the form of a detailed strategic analysis, covering various aspects such as market trends, competitive landscape, customer behavior, operational inefficiencies, and any other relevant factors affecting the platform's profitability. Provide a comprehensive breakdown of each factor

and its impact on the platform's profitability."

What are the strategic reasons that could be causing the profit decline of Rated?

How did you determine the strategic causes for Rated's profitability decline? Please detail your reasoning process.

Option

Below is a suggestive prompt for brainstorming. Please note that it serves as a guideline. You might need to adjust your prompts for more precise results.

"Identify and outline a comprehensive list of strategic options available for Rated based on the strategic reasons you previously identified. Please provide as many strategic options as possible, ensuring that each option is distinct and well-defined. Your response should encompass a wide range of potential strategic choices that Rated could consider, taking into account various aspects of the business and its competitive environment. Your strategic options should be creative, relevant, and reflective of the strategic reasons previously identified."

Based on the strategic reasons you identified earlier, what are the strategic options available for Rated? Please come up with as many strategic options as possible.

Please input only one option into each field. You can click the "+" button to add more options.

You have come up with the following strategic options. From all the strategic options that you have articulated, please select which you believe is the best one.

How did you come up with strategic options for Rated? Please describe your thought process.

Option - Survey Qs

How difficult was it for you to come up with strategic options?

Very Easy	Easy	Neutral	Difficult	Very Difficult
○	○	○	○	○

How confident are you about the options you have developed?

Very Confident	Confident	Neutral	Slightly Confident	Not Confident
○	○	○	○	○

Math

Rated is planning to launch a new analytics platform. The platform has two different subscription models: Basic and Premium. The fixed costs for developing and launching the platform are $800,000. The variable cost for providing the Basic subscription is $40 per subscription, and $80 for the Premium subscription. The price for the Basic subscription is set at $200 per year, and the Premium subscription is priced at $400 per year. In the first year, market research suggests that the mix of Basic to Premium subscriptions will likely be 2 to 1.

How many of each type of subscription must Rated sell to break even in the first year? You are not allowed to use the LLM tool provided for this question.

- ○ 2,000 Basic and 1,000 Premium
- ○ 2,500 Basic and 1,250 Premium
- ○ 3,000 Basic and 1,500 Premium
- ○ 3,500 Basic and 1,750 Premium

LLM Qs

How useful do you find the tool provided in completing this exercise?

- ○ Extremely useful
- ○ Very useful
- ○ Moderately useful
- ○ Slightly useful
- ○ Not at all useful

Please describe how it was useful / not useful.

How often do you use LLM (large language model) tools such as ChatGPT?

- ○ Daily
- ○ Several times a week
- ○ A few times a month
- ○ Rarely
- ○ Never

How would you rate your level of familiarity with LLM tools such as ChatGPT?

- ○ Very familiar – I understand many of its functions and limitations.
- ○ Somewhat familiar – I have a good understanding of what it can do.
- ○ Neutral – I know what it is but not much about its capabilities.
- ○ Somewhat unfamiliar – I have heard of it but don't understand how it works.
- ○ Very unfamiliar – I have little to no knowledge about it.

Figure A.1: Custom Web Interface

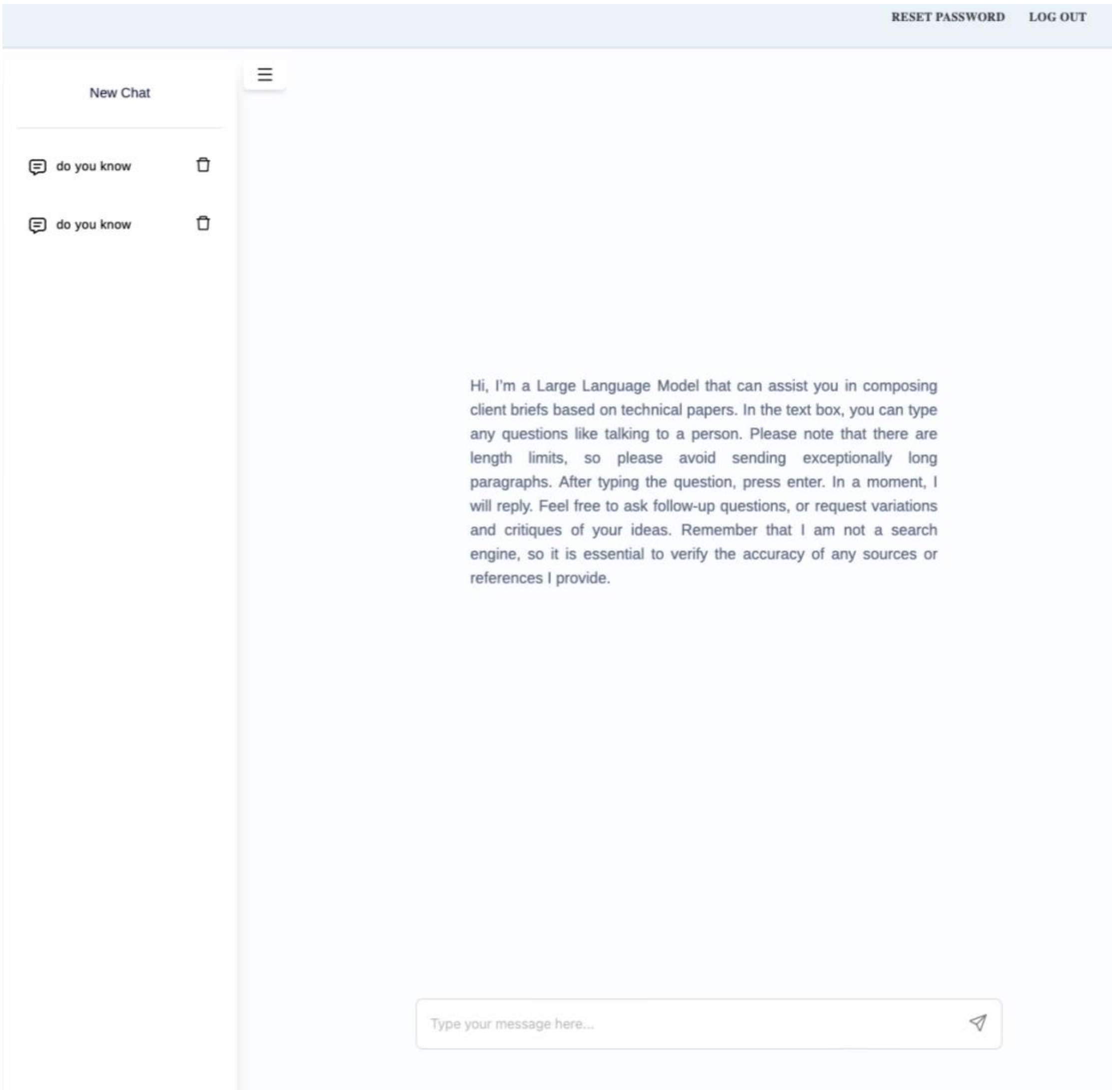


Figure A.2: Treatment Assignment

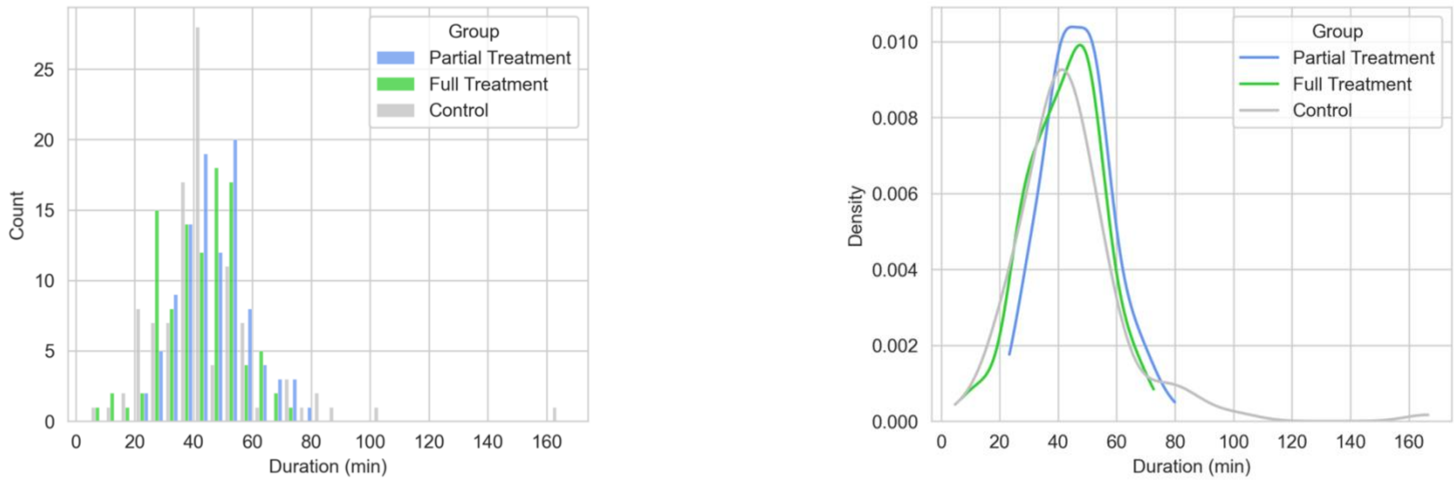
Condition	Problem Formulation	Ideation
Control Group	Human Only	Human Only
Partial Treatment	Human Only	Human + GPT
Full Treatment	Human + GPT	Human + GPT

Human Only   Human + GPT

Notes: This figure shows the assignment of GPT access across treatment groups.

Figure A.3: Distribution of Duration (in minutes)

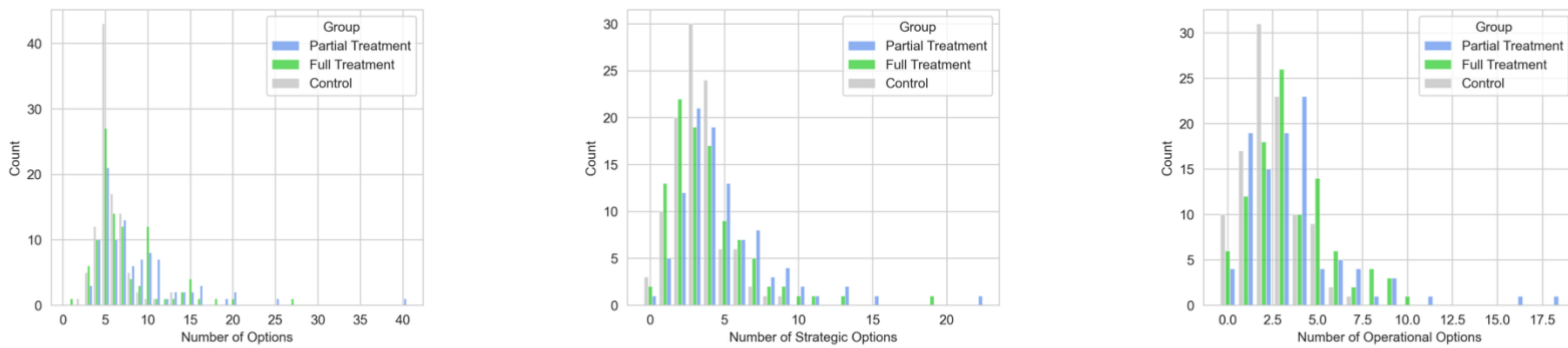
(a) Histogram (b) Density distribution



Notes: These figures show the histogram and density distribution of duration (in minutes) of the Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Control vs. Full Treatment K-S Test p-value: 0.177; Control vs. Partial Treatment K-S Test p-value: 0.019; Full Treatment vs. Partial Treatment K-S Test p-value: 0.058.

Figure A.4: Histogram of Ideation Measures

(a) Total number of options (b) Number of strategic options (c) Number of operational options



Notes: These figures show the histograms of outcome variables of the Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Panel (a) shows the distribution of the number of alternatives generated. Panel (b) shows the distribution of the number of strategic options generated. Panel (c) shows the distribution of the number of operational options generated.

Table A.1: Distribution of Participants Across Classroom Sections and Experimental Conditions

Section	Control group	Partial treatment	Full treatment	Total
1	25	23	24	72
2	25	27	26	78
3	24	23	24	71
4	29	27	28	84
Total	103	100	102	305

Notes: This table reports the distribution of the 305 participants across four sections and the three experimental conditions: control (no LLM access), partial treatment (LLM access during ideation only), and full treatment (LLM access during both problem formulation and ideation).

Table A.2: Suggested Prompts Used for Problem Formulation and Ideation

	Prompts
Problem formulation	*In this task, you will be playing the role of a strategic management consultant. Your task is to identify and list strategic factors that contribute to the profit decline of the online restaurant review platform Rated. Your response should be in the form of a detailed strategic analysis, covering various aspects such as market trends, competitive landscape, customer behavior, operational inefficiencies, and any other relevant factors affecting the platform's profitability. Provide a comprehensive breakdown of each factor and its impact on the platform's profitability.*
Ideation	*Identify and outline a comprehensive list of strategic options available for Rated based on the strategic reasons you previously identified. Please provide as many strategic options as possible, ensuring that each option is distinct and well-defined. Your response should encompass a wide range of potential strategic choices that Rated could consider, taking into account various aspects of the business and its competitive environment. Your strategic options should be creative, relevant, and reflective of the strategic reasons previously identified.*

Notes: This table displays the prompts provided to participants during the experiment. All participants, regardless of condition, received these prompts as part of the task instructions. Participants were informed that the prompts were optional and could be freely modified to suit their own approach
.

Table A.3: Pairwise Comparisons of Baseline and Outcome Variables across Experimental Groups

	(1) Partial - Full			(2) Control - Full			(3) Control - Partial		
	b	se	p	b	se	p	b	se	p
Baseline Variables									
Female	-0.07	0.07	0.278	-0.09	0.07	0.166	-0.02	0.06	0.774
Consulting	-0.01	0.07	0.837	-0.01	0.07	0.844	0.00	0.07	0.992
Age	0.06	0.35	0.856	-0.31	0.35	0.377	-0.38	0.34	0.270
Main Outcome Variables									
Number of Options	1.10	0.66	0.097	-1.63	0.44	0.000	-2.73	0.55	0.000
Number of Strategic Options	1.02	0.43	0.019	-0.60	0.32	0.064	-1.62	0.36	0.000
Number of Operational Options	0.08	0.36	0.827	-1.03	0.27	0.000	-1.11	0.32	0.001
Proportion of Strategic Options	0.07	0.03	0.019	0.06	0.03	0.060	-0.01	0.03	0.755
Proportion of Operational Options	-0.07	0.03	0.019	-0.06	0.03	0.060	0.01	0.03	0.755
Best Option is Strategic	0.14	0.07	0.044	0.14	0.07	0.041	0.00	0.07	0.994
Best Option is Operational	-0.14	0.07	0.044	-0.14	0.07	0.041	-0.00	0.07	0.994
Duration (min)	4.58	1.71	0.008	1.96	2.34	0.403	-2.62	2.29	0.255
Observations	202			205			203		

Notes: This panel presents pairwise comparisons in means between treatment conditions. Each cell reports the difference in group means (b), standard error (se), and p-value (p). These comparisons test whether treatment conditions significantly differ in terms of baseline characteristics and key outcome variables.

Table A.4: Impact on Ideation and Selection (excluding outliers)

VARIABLES	(1) Number of Options	(2) Number of Strategic Options	(3) Number of Operational Options	(4) Proportion of Strategic Options	(5) Proportion of Operational Options	(6) Best Option is Strategic
Full Treatment	1.48***	0.49	1.00***	-0.06*	0.06*	-0.16**
	(0.44)	(0.32)	(0.29)	(0.03)	(0.03)	(0.07)
Partial Treatment	2.59***	1.47***	1.12***	0.00	-0.00	-0.01
	(0.56)	(0.37)	(0.33)	(0.03)	(0.03)	(0.07)
Constant	5.28***	3.08***	2.20***	0.56***	0.44***	0.61***
	(0.60)	(0.44)	(0.31)	(0.04)	(0.04)	(0.09)
Observations	293	293	293	293	293	293
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.07	0.06	0.06	0.01	0.01	0.01
Control Group Mean	5.79	3.31	2.48	0.57	0.43	0.65
F-test	2.74	5.14	0.12	4.73	4.73	4.33
Prob > F	0.099	0.024	0.733	0.030	0.030	0.038

Notes: This table reports regression results for equation (1), excluding duration outliers defined as 2SD away from the mean. All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

APPENDIX B: RATED CASE

The Company

In December of 2018, Rose Nakamura, CEO of Rated[1], found herself in intense discussions with her senior executives about the future of the business she had founded 8 years ago. During those years, Rated had become one of the largest online platforms for consumer reviews of local businesses -- expanding from its initial start in restaurants in New York to cover all local business verticals. By 2018, Rated had accumulated over 25 million reviews with a dedicated community of reviewers who wrote reviews for free, attracting more than 70 million unique visitors per month. Yet, Rated was still not turning a profit.

The Internal Context

Michael Wood, Global Head of Sales: I called this meeting because we have a serious problem. Our revenue per advertiser is down from last quarter, and it's trending downward. We are trying to get more first-time advertisers on board to make up for this, but it's only going to get us so far if they continue cancelling almost immediately.

Here's my plan. I've identified 100 big national chain brands as strategic accounts we want to break into: McDonalds, Starbucks, KFC, you name it. This is a segment that we've been underrepresented in, and they have huge budgets on longer time frames. We need to go after the big fish, not the small plankton we've been chasing in the local markets.

I need more resources to make this happen. Our salespeople are working tirelessly around the clock cold-calling businesses, and we pay them 70% of the market rate. How can we expect them to stay with us and get these accounts when there isn't enough of an incentive?

And we really need the Product team to step up. We make money by selling advertising, and our ads product is a disaster. We have barely updated it since we started, and it is nowhere near what the Googles and Facebooks are doing. They are running ads auctions, where businesses only pay when a customer interacts with the ad. We are still charging a flat rate per month of advertising, and businesses don't want to pay for that.

We also need to chuck our algorithm flagging fake reviews. We know that it flags at least a quarter of all reviews, and a lot of those are real, legitimate reviews. Business owners are furious that these reviews are being taken down, and they don't even want to talk about buying ads – they slam the phone down as soon as they hear it's us.

Imani Ahmad, Chief Product Officer: I'm with you on the ads product – we are working on it. But the algorithm? Oh sure, we would be fine as a business if we just said, anything goes on Rated. You want to get 5 stars? Get your family, friends, employees to write a bunch of reviews. Wouldn't that be wonderful? You can then advertise with us!

Our product means something. When you see a 5-star restaurant on Rated, you know it's going to be stellar. And you're never going to eat at a 1-star restaurant. That's why users flock to our platform – they

[1] The names of the company and the management team as well as some factual details have been disguised to preserve confidentiality.

trust our content. And that's why we have our Rated Star reviewers who are giving us amazing content for free. We have more reviews than any other platform out there!

Sarah Koons, Chief Technology Officer: I also don't think the algorithm is the problem. The real issue is that your salespeople are badgering businesses and being way too aggressive in pushing them to buy advertising, as if that will get them more positive reviews. This is destroying our reputation and I'm sure it is affecting our ads revenues. If anything, I think we need to reduce their incentives.

Jeff Sterling, Chief Operating Officer: I think you all have a point here. In my view, a big part of what is happening right now is the changing market – we were an early innovator in getting everyday consumers to review local businesses, but now the marketplace is getting crowded. Everyone is selling ads to local businesses, from Mugshots to Flex to Chowhub to Rezly, not to mention all of the new startups. How can we compete with a giant like Flex? They account for 70% of our traffic, and they're becoming a fiercer competitor by the day. We've been getting approached by competitors who are eyeing our business, and maybe we should think more about this. Rose, what's your take?

Rose Nakamura, CEO: I hear all of your concerns, and I think we need to take some time to think through our next steps very carefully. But one thing I want to remind you of is why we're all here. Our platform helps people find great local businesses. There is no site like ours that provides the breadth and depth of content, and this draws in even more consumers to our site. We are already the definitive guide to local businesses in most major cities in the US. It's a matter of time until businesses see that there's no better value they can get from advertising than our community of consumers who come to our site with the intent to buy. We've had a huge growth in user adoption in the last few years: our reviews are up 71 percent from last year, our unique monthly visitors are up 74 percent. The network effects are going to kick in, and our revenues will likely follow.

The Decision

As she rallied her team, Rose thought to herself, how do we figure out what to do?

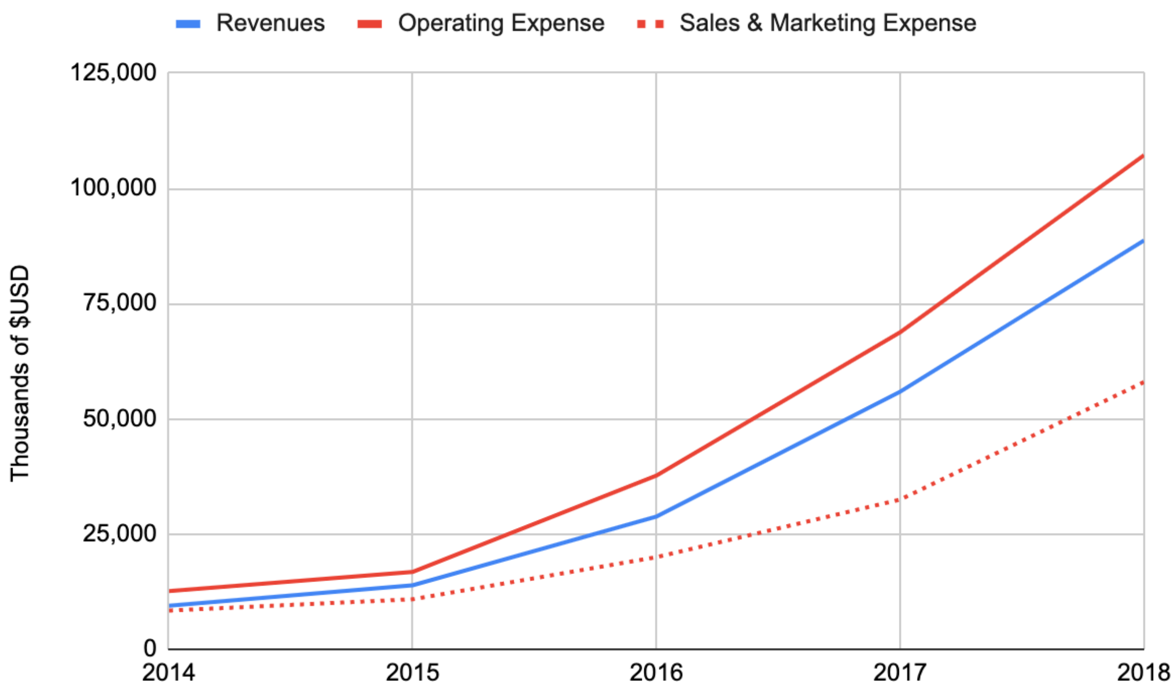
Exhibit 1: ***Internal Data***

Rated 4Q18 Consolidated Statements of Operations (as a percentage of net revenue)

	Three Months Ended Dec. 31, 2018
Consolidated Statements of Operations data	(as a percentage of net revenue)
Net revenue by product	
Local independent business advertising	83%
Chain business advertising	13%
Other services	4%
Total net revenue	100%
Costs and expenses	
Cost of revenue	7%
Sales and marketing	65%
Product development	12%
General and administrative	32%
Depreciation and amortization	5%
Restructuring and integration costs	-
Total costs and expenses	121%
Loss from operations	(21%)
Other income (expense), net	
Loss before income taxes	(21%)
Provision for income taxes	-
Net loss	(21%)

Source: Internal data

Revenues vs. Expenses



Source: Internal data

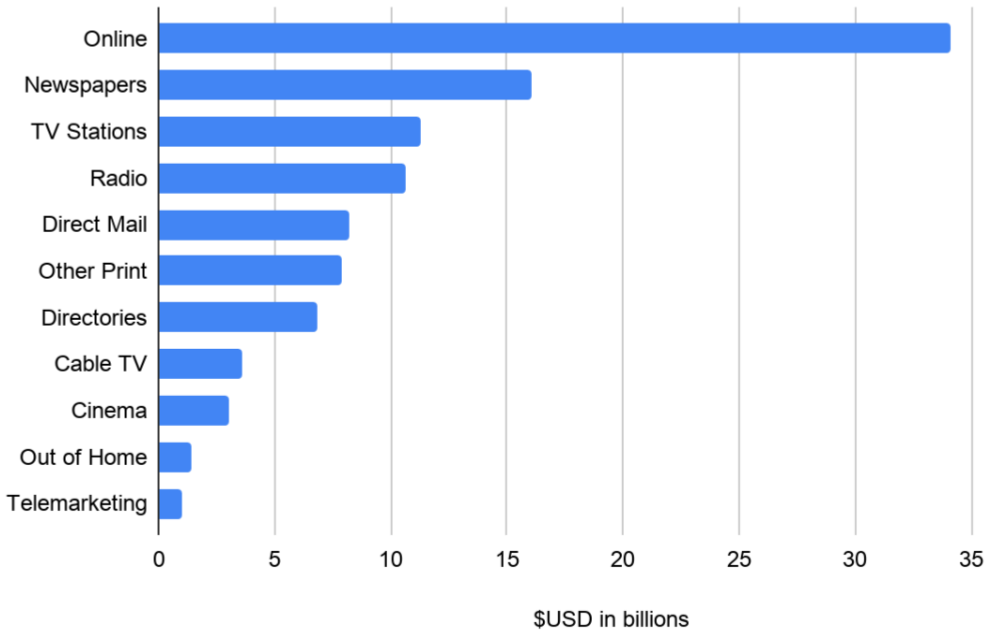
The Returns to Purchasing Rated Ads for 3 months

Category	Pageviews (% lift)	Calls (% lift)	Reservations (% lift)
Chain Business	15%	5%	0%
Local Independent Business	25%	10%	0%

Note: This table shows the results from an internal experiment over 3 months in 2018
Source: Internal data

Exhibit 2: *Market Overview*

U.S. Local Advertising Spending in 2018



Search Advertising Spending Worldwide from 2009-2018

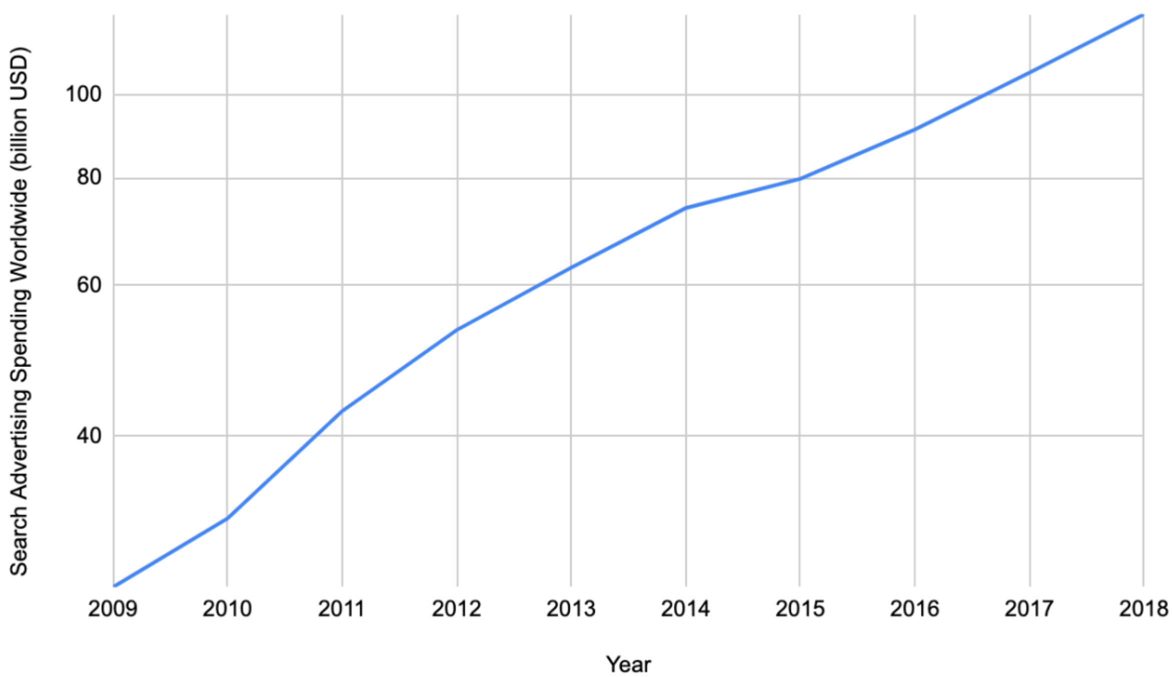


Exhibit 3: ***Competitors***

Product Coverage of Competitors

	Customer Reviews	**Business Search & Discovery**	**Food Delivery**	**Reservation System**	**Social Network**	**Vertical outside Restaurants (*)**
Rated	X	X				X
Flex	X	X				X
BookIt!	X	X				X
Umami	X	X	X			
Chowhub	X		X		X	
Rezly	X	X		X		
Mugshots	X	X			X	X

Note: "X" marks the markets that each company is in. Company names have been changed.
(*): e.g. fitness centers, beauty salons, concerts, cinemas, doctors, etc.

Source: Competitor analysis

Number of Reviews and Monthly Unique Visitors Across Selected Competitors in 2018

	Number of Reviews	**Number of Monthly Unique Visitors**
Rated	25.7M	71.4M
Flex	8.3M	*Not available*
BookIt!	75.6M	60M
Rezly	15.1M	*Not available*

Note: Company names have been changed. Selected competitors are those that are most prominent with respect to customer reviews at the time.

Source: Competitor analysis

International Presence of Competitors

	North America	Europe	Asia Pacific	ME/Africa
Rated	H	L	None	None
Flex	M	M	M	M
BookIt!	M	H	H	H
Umami	L	L	H	M
Chowhub	M	None	None	None
Rezly	H	M	L	None
Mugshots	H	H	H	H

Note: *H* indicates "High", *M* indicates "Middle", and *L* indicates "low" presence, relative to the competitor set in the region. Company names have been changed.

Source: Competitor analysis

Recent Acquisition Activities of Competitors

- Flex acquired Zigit, an expert reviews product (rather than everyday consumer reviews).
- Bookit! acquired Flywithme, a consumer travel site and hotel search engine, and Jetty, a flash sale site.
- Chowhub acquired Ordery, a restaurant and menu directory product.

Note: Company names have been changed.

Source: TechCrunch

Exhibit 4: ***Stakeholder Interviews***

Interviews with Restaurants

“There is a special place in hell for Rated. The salespeople badger you until you buy their advertising, and magically your rating goes up.”

“Why would I pay for advertising on Rated? Everyone can see that my restaurant has 4.5 stars. I don’t see any need to pay for more advertising on there, it does the job itself.”

“Rated has helped me grow my business. I just started this cafe a year ago, and thanks to Rated we’ve never had an empty seat. I didn’t have to do any marketing -- my customers’ reviews spoke for themselves.”

“I never know where I should buy advertising, and whether spending that much on advertising was worth it. Who knows if the customer who walks in through the door found out about us through Rated, Flex, or what, and if they would have still walked in without seeing our ad. But at the end of the day, I guess I advertise with Rated because a ton of customers seem to read their reviews. And I might as well get my business in front of as many eyeballs as possible.”

Interviews with Users

“Rated reviews are the only ones I trust. It’s rather comical that other platforms call theirs reviews -- most are just star ratings with no text. Rated gives me menus, photos of the place, and dishes to try -- and avoid.”

“I hate that Chowhub and all these delivery platforms use their own ratings and reviews, which don’t even mean anything. All the restaurants have 4 stars, and I always have to switch back and forth between the Rated site and Chowhub to figure out what to order. I wish I could just have the ratings from Rated there when I make choices about what to eat.”

“I love food, so I started reviewing on Rated as soon as they got started. Because of the quality of my reviews, they invited me to be a “Rated Star” -- which is reserved for the best reviewers. I love this community, and will never stop reviewing for Rated. None of the other reviews platforms has anything like this community.

“I’ve been dying to become a “Rated Star” and have been reviewing a ton in the last year. But New York is such a competitive market -- so many reviewers. I think I’m getting there, though. I just love that people read my reviews and they find them helpful.”

Source: Business and user interviews

APPENDIX C: TRAINING A FINE-TUNED GPT-3.5 TURBO MODEL

We employed a fine-tuned GPT-3.5 Turbo model trained on a dataset of human-coded options to automate the coding of the options. The training data consisted of 2,269 options coded by human coders. For each option, we asked the model to code whether it is 1) Strategic and 2) Continue. A system message was provided at both the training and deployment stages to further guide the model's behavior. This ensured consistency in how the model interpreted and coded the options, aligning its outputs with the predefined rubric. Below is the system message:

"Code each option according to the following rubric. Return only the coding results. (1) Strategic: Dummy, whether the option is strategic or not, 1 means strategic, 0 otherwise. Strategic options are long-term, high-level plans, not easy to reverse, and involve significant resource allocation. Examples include launching new products, entering new markets (e.g., geographic), M&A, exit, a complete overhaul of the platform's user interface, and strategic partnerships. Operational options are specific courses of action with specific short-term goals, focusing on day-to-day activities and tasks. Examples include cutting costs, promotion/marketing, increasing sales team pay, training sessions. (2) Continue: Dummy, whether the option is 'continue as it is' or not, 1 means continue as it is, 0 means otherwise. Examples for 'continue as it is' include explicitly mentioning 'continue as it is,' increasing customer stickiness, changing pricing system — note as a special case as in this instance as it could be seen as a change, streamlining operations, improving efficiency, or increasing profit margins without fundamentally altering its value proposition, target market, or key activities. The broader business model or strategy remains largely unchanged. (e.g., keeping the focus on local businesses as the key target market would represent 'continuing as is,' while changing the target market from local businesses to national chains would not). (3) Change Pricing of Ads: Dummy, 1 means the option indicates a change in pricing, 0 means otherwise. Changing the pricing of the ads. Change in pricing would explicitly mention phrases like 'change of pricing structure,' 'change the pricing model,' etc. (4) Exit: Dummy, 1 means the option indicates exit, and 0 otherwise. Exit means selling the business or exiting the market."

After training, we evaluated the model's performance by coding a new set of 835 options and comparing the results with human coders, which indicates that the model can be a reliable tool for this specific coding task.

Table C.1: Consistency Between Fine-tuned GPT and Human Coders

	Strategic	**Continue**	**Change Pricing**	**Exit**
Agreement	87.90%	87.07%	97.72%	99.04%
Cohen's Kappa	0.7350	0.7373	0.8792	0.8413

Below is the coding rubric used by the human coders, which was constructed by the first two authors based on the case study used in the experiment. The human coding process is described below. Each option and each set of options were coded by two research assistants independently based on the coding rubric provided by the authors. Following this, their coding results were compared to identify discrepancies. In cases of disagreement, a third assistant served as the arbitrator to resolve the issue and finalize the coding decision. In total, four research assistants were involved in the coding process. To maintain objectivity, all of them were blind to the experiment details, the hypotheses, and the conditions.

For **each option**

- Code as strategic vs. operational option:

	Description	**Examples**
Strategic option	Long-term, high-level plan, not easy to reverse, involve significant resource allocation	• Launch new product • Enter new market (e.g., geographic) • M&A • Exit • A complete overhaul of the platform's user interface • Strategic partnerships
Operational option	Specific course of action with specific short-term goals, focus on day-to-day activities and tasks	• Cut costs • Promotion / marketing • Increase sales team pay • Training sessions

- Code whether the option suggests that the company continue its current strategy:

	Examples
Continue as it is	• Explicitly mention “continue as it is” • Increase customer stickiness • Change pricing system --- note as special case as in this instance as it could be seen as a change • Streamline operations, improve efficiency, or increase profit margins without fundamentally altering its value proposition, target market, or key activities. The broader business model or strategy remains largely unchanged. (e.g., keeping the focus on local businesses as the key target market would represent “continuing as is”, while changing the target market from local businesses to national chains would not)
Not continue	• Launch new product • Enter new business segment / geographic market • Exit

For **<u>each student</u>**

- Code whether the options were mutually exclusive
  - Mutually exclusive: Choosing one option would prevent the firm from choosing any of the other options.
  - All the options need to be mutually exclusive with each other, so any presence of a non-mutually exclusive option would make the set as a whole not mutually exclusive
  - Assume the company has limited resources, launching one strategic initiative would constraint the company from launching another

	Examples
Mutually exclusive	• 1. Continue; 2. Expand; 3. Exit • 1. Expand ecosystem; 2. Focus on ad sales; 3. Sell the business • (Refer to the figure)
Not mutually exclusive	• 1. Bring in new clients; 2. Automation to cut costs • 1. Expand service offering; 2. Optimize pricing; 3. Modify compensation; 4. Reduce admin costs

Output

1. Create one spreadsheet with the coding of all strategic options specifying response ID (with the actual options text)
   a. ResponseID
   b. OptionID
   c. Option: Actual option text
   d. Strategic: Dummy, whether the option is strategic or not, 1 means strategic, 0 otherwise
   e. Continue: Dummy, whether the option is “continue as it is” or not, 1 means continue as it is, 0 means otherwise
   f. Change Pricing of Ads: Dummy, 1 means the option indicates a change in pricing, 0 means otherwise
      i. Changing the pricing of the ads
      ii. Change in pricing would explicitly mention phrases like “change of pricing structure”, “change the pricing model”, etc.
   g. Exit: Dummy, 1 means the option indicates exit, and 0 otherwise
      i. Exit means selling the business or exiting the market
   h. Note
      i. <u>Improve algorithm</u> if the option mentions this
      ii. <u>Special case of "continue as it is"</u> if not fully sure

iii. Multiple if the option seems to include multiple options
iv. Partially exit if the option mentions exit some market or some product

2. Create a spreadsheet with the following columns: response ID, whether options were mutually exclusive
    a. ResponseID
    b. Mutually Exclusive: Dummy; NA if only 1 option is reported

APPENDIX D: MEASUREMENT DISCUSSION

This appendix presents additional analyses and conceptual clarifications related to measurement and outcome coding. These include: (1) additional outcome variables related to "continue" options, and (2) an initial conceptual discussion on the breadth vs. specificity distinction in strategic option measurement.

D.1 Outcomes Related to "Continue" Options

In the main manuscript, we focus primarily on strategic option generation and selection. However, the original experiment also included coding for “continue” options—i.e., ideas that propose maintaining the current course of action. These options are conceptually distinct from strategic alternatives and provide a useful benchmark for comparison.

Additionally, we coded whether options represented a continuation of the firm's current strategic direction. Continue options maintain the existing business model and competitive approach, such as “develop retention strategies to strengthen relationships with existing clients,” and “focus on small, local businesses (not chains) as this is of the core mission.” This dimension is theoretically important because strategic persistence versus change represents a fundamental choice facing executives (Tripsas & Gavetti, 2000). While firms often struggle with inertia and fail to adapt when change is needed (Gilbert, 2005; Henderson & Clark, 1990), continuation can also be the optimal choice when current strategies remain viable (Burgelman, 2002). By examining how LLM assistance influences participants' orientation toward maintaining versus changing strategic direction, we can assess whether AI tools systematically shift decision-makers toward particular strategic postures.

We report the following variables: (1) Number of Continue Options; (2) Proportion of Continue Options; and (3) Best Option is Continue (binary indicator).

Table D.1 reports descriptive statistics. Table D.2 reports regression results using these additional outcomes.

Table D.1: Summary Statistics and between-group comparisons

Panel (a) Descriptive statistics by experimental group

	(1) Full sample				(2) Full Treatment		(3) Partial Treatment		(4) Control	
	Mean	SD	Min	Max	Mean	SD	Mean	SD	Mean	SD
Number of Continue Options	3.65	2.63	0.00	24.00	3.97	2.49	4.15	3.38	2.84	1.55
Proportion of Continue Options	0.51	0.22	0.00	1.00	0.55	0.23	0.48	0.21	0.50	0.22
Best Option is Continue	0.48	0.50	0.00	1.00	0.58	0.50	0.40	0.49	0.46	0.50
Observations	305				102		100		103	

Panel (b) Pairwise comparisons of baseline and outcome variables

	(1) Partial - Full			(2) Control - Full			(3) Control - Partial		
	b	se	p	b	se	p	b	se	p
Number of Continue Options	0.18	0.42	0.667	-1.13	0.29	0.000	-1.31	0.37	0.000
Proportion of Continue Options	-0.07	0.03	0.026	-0.05	0.03	0.094	0.02	0.03	0.604
Best Option is Continue	-0.18	0.07	0.011	-0.12	0.07	0.081	0.06	0.07	0.420
Observations	202			205			203		

Notes: Panel (a) reports the summary statistics of the outcome variables across experimental groups. Coding was performed using a fine-tuned GPT-3.5 Turbo model trained on human-coded examples (see Appendix A for details). Observations reflect the number of participants per condition. Panel (b) presents pairwise comparisons in means between treatment conditions. Each cell reports the difference in group means (b), standard error (se), and p-value (p).

Table D.2 Regression Results on Continuation-related Variables

VARIABLES	(1) Number of Continue Options	(2) Proportion of Continue Options	(3) Best Option is Continue
Full Treatment	1.10***	0.06*	0.13*
	(0.30)	(0.03)	(0.07)
Partial Treatment	1.31***	-0.01	-0.05
	(0.37)	(0.03)	(0.07)
Constant	2.63***	0.53***	0.53***
	(0.33)	(0.04)	(0.09)
Observations	305	305	305
Gender-Consulting-Section FE	YES	YES	YES
Adjusted R-squared	0.05	0.00	0.01
Control Group Mean	2.84	0.50	0.46
F-test	0.27	4.85	7.19
Prob > F	0.604	0.028	0.008

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

D.2 Conceptual Distinction: Broad versus Specific Options

One concern about measurement may be that LLM-assisted problem formulation impacts the specificity of alternatives generated. The full treatment group, using LLMs for both problem formulation and ideation, appears to have been able to develop a more comprehensive understanding of the problem. This detailed understanding might lead to more specific, actionable options that address particular aspects of the problem, specific to the case context. In contrast, those with a less detailed understanding of the problem (control or partial treatment groups) might default to proposing higher-level, seemingly more strategic solutions that lack specificity, such as "M&A" or "enter a new market."

It is important to distinguish between "strategic" and "broad". While strategic options involve long-term, high-commitment plans with significant resource implications, broad options are more general and may lack contextual specificity. In other words, what was observed in the full treatment group might not be a lack of strategic focus, but rather a different, potentially more informed kind of strategic approach that leads to specificity.

We developed a few-shot classification algorithm based on a pre-trained BERT base model (uncased) (Devlin et al., 2019) to categorize options as either "Broad" or "Specific". Specifically, we first identified a small set of pre-labeled example options as anchors for each class. Then, we computed BERT embeddings for each of the labeled sentences by averaging the last hidden layer across tokens. The resulting embeddings were used to compute centroids representing the "broad" and "specific" classes. Each participant-submitted option was then embedded and classified according to its proximity to the two centroids using Euclidean distance. If an option was closer to the "specific" centroid, it was labeled as a specific action; if it was closer to the "broad" centroid, it was labeled as a broad direction. We then calculated the percentage of broad options for each participant.

Table D.3 shows that while both treatment groups tended to generate a higher proportion of broad options ($p = 0.001$ for the full treatment group, $p = 0.1$ for the partial treatment group) compared to the control group, their coefficients were very similar (F-test $p = 0.66$). This might suggest that both treatment groups demonstrated a tendency towards broader strategic options, but with limited differences between them.

Table D.4 shows examples from participant-submitted options to illustrate the differences of the strategic vs. operational distinction from the broad vs. specific articulation. Table D.5 presents the regression results of the full sample.

Table D.3: Pre-classified Sentences used in the Few-shot Classification Algorithm

Broad	Investment in talent development
	Technological improvement
	Acquisition or merger opportunities
Specific	Develop a rating system for businesses based on their social impact
	Launch subscription-based services or premium membership programs
	Focus on local independent businesses

Notes: These examples were manually constructed to train the few-shot classification algorithm. They serve as representative anchors for the two conceptual categories—specific actions and broad directions—and were used to compute semantic centroids based on BERT embeddings.

Table D.4: Strategic vs. Operational and Broad vs. Specific Options

	Strategic	Operational
Broad	Revise Monetization Model	Investment in talent development
Specific	Expand into Asia Pacific market and enhance presence in European market	Improve SEO and offer the service to local businesses

Notes: This table presents examples drawn from participant-submitted options in our data. It illustrates the differences of the strategic vs. operational distinction (based on the option's scope, reversibility, and resource commitment) from the broad vs. specific articulation (based on the level of contextual detail).

Table D.5: Impact on Solution Specificity

VARIABLES	(2) Proportion of Broad Options
Full Treatment	0.11***
	(0.04)
Partial Treatment	0.10**
	(0.04)
Constant	0.23***
	(0.06)
Observations	305
Gender-Consulting-Section FE	YES
Adjusted R-squared	0.06
Control Group Mean	0.09
F-test	0.19
Prob > F	0.660

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

APPENDIX E: POTENTIAL MECHANISMS

This appendix provides further detail on the operationalization of key measures and supplementary analyses related to the mechanisms discussed in the main text: cognitive offloading, attention dilution, and cognitive drain. These analyses draw on behavioral patterns, interaction logs, and semantic analyses to provide convergent evidence for the proposed explanations.

E.1 Cognitive Offloading

E.1.1 Problem formulation similarity

Using sentence embeddings derived from a pre-trained BERT model and computing cosine similarity between each participant's problem statement and the central tendency (median embedding) of their group

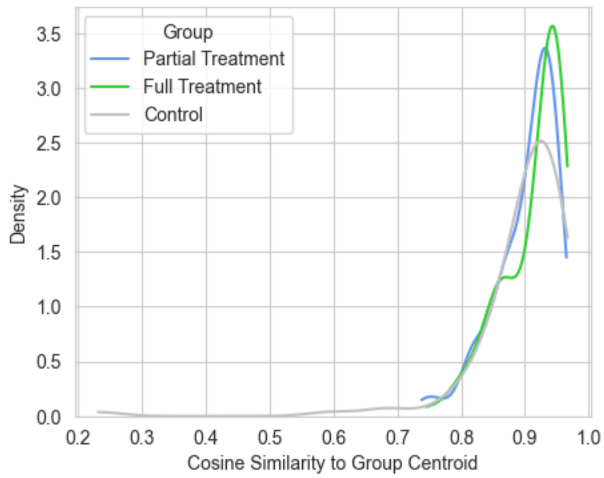
To assess the extent to which participants within each condition generated similarly formulated problem statements, we employ a semantic similarity analysis based on pre-trained transformer-based language models. Specifically, we use BERT (Bidirectional Encoder Representations from Transformers) to convert textual problem statements into high-dimensional vector representations and compare these representations within and across groups.

Each participant's problem formulation is tokenized and passed through the bert-base-uncased model. We compute the mean of the last hidden layer for all tokens in a sentence, yielding a single dense vector representing the semantic content of each problem statement. These vectors serve as semantic embeddings that capture the contextual meaning of each problem statement.

To quantify group-level convergence in problem framing, we calculate a median embedding for each experimental group. The median (rather than the mean) is used to reduce sensitivity to outliers. These median embeddings serve as group centroids, representing the prototypical problem framing within each condition.

Next, we compute the cosine similarity between each participant's embedding and their group's centroid. Cosine similarity is a widely used metric for comparing high-dimensional vectors, capturing the angle between them in embedding space (independent of magnitude). Higher values indicate stronger semantic alignment with the group's typical problem framing. Figure E.1 shows the distribution of the cosine similarity distance between the problem formulation embedding and the group centroid embedding.

Figure E.1: Distribution of the Problem Cosine Similarity by Experimental Groups



Notes: This figure shows the density distribution of the cosine similarity distance between the problem formulation embedding and the group centroid embedding across experiment groups: Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Control vs. Full Treatment K-S Test p-value: 0.132; Control vs. Partial Treatment K-S Test p-value: 0.68; Full Treatment vs. Partial Treatment K-S Test p-value: 0.063.

E.1.2 Reliance on LLM-Generated Ideas

To assess the extent to which participants relied on LLM-generated ideas, we calculated the proportion of participant-submitted options that significantly overlapped with LLM-generated content. For each option, we defined it as "matched" if it shared at least a 15-gram sequence with any of the LLM outputs presented earlier in the task. We then compared the proportion of options matched to the LLM output for each participant. On average, the full treatment condition has 32.7% options matched to the LLM output, while the partial treatment condition has 55.6% (p=0.0176).

E.1.3 Interaction patterns

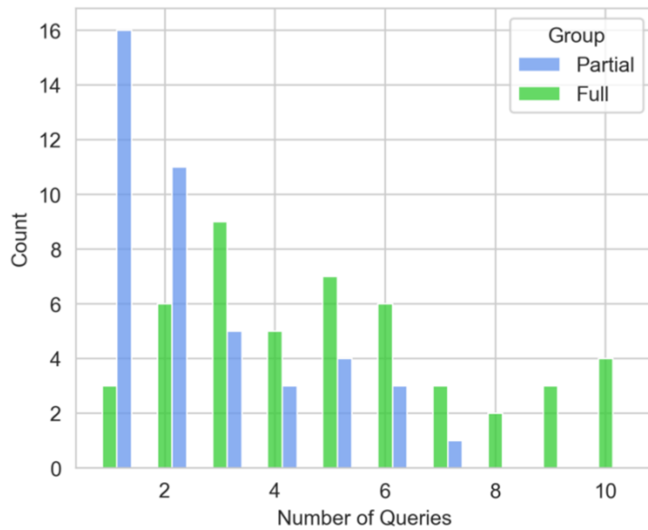
We recorded all interaction logs between participants and the LLM (88 participants in total, with 47 from the full treatment group and 41 from the partial treatment group). Figure E.1 shows the distribution of the number of queries by experimental groups.

Queries were categorized into seven mutually exclusive types based on their content and origin. Problem queries were those that drew on suggestive prompts or survey items related to problem formulation. Option queries similarly used prompts or survey items that ask LLMs to generate options. Reflection queries stemmed from survey questions that asked participants to articulate their reasoning behind problem formulation or option generation. Selection queries involved asking the LLM to evaluate and choose the best among the proposed options. Other survey queries originated from general survey items unrelated to ideation, such as questions about confidence, perceived difficulty, or a math task. Contextual information queries sought background or industry-specific insights. Any remaining queries that did not fall into these predefined categories were classified as other.

Table E.1 presents the comparison of types of queries between the full treatment group and the partial treatment group. Table E.2 shows some examples of prompts participants used other than the suggestive prompts provided.

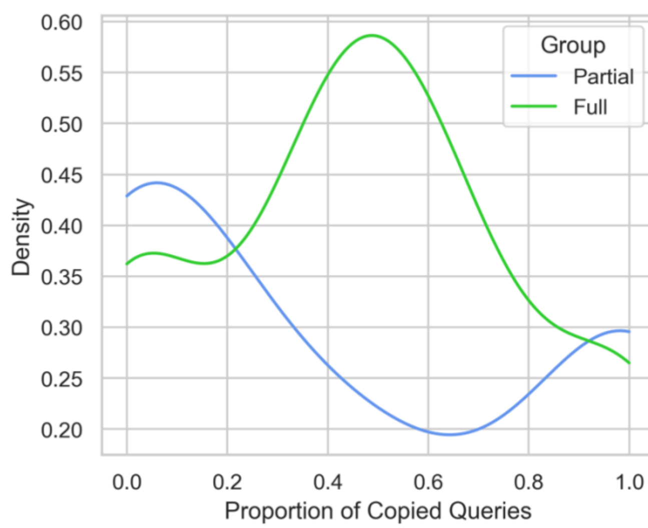
Figure E.2 shows the distribution of the number of queries by experimental groups. Figure E.3 shows the distribution of Proportions of copied queries by group. Copied queries include both the suggestive prompts and the survey questions. On average, participants in the full treatment group copied 46.4% of their queries, compared to 42.7% in the partial treatment group. The difference between the means is small and not statistically significant (t-test $p = 0.659$). However, the distributional difference (K-S test $p = 0.090$) implies that copying behavior may be more polarized in the partial treatment group. More people either copied very little or almost everything. The full treatment group's behavior is more clustered around the mean.

Figure E.2: Distribution of the Number of Queries by Experimental Groups



Notes: This figure shows the histogram of the distribution of the number of queries by experimental groups, collected from 88 participants, 47 from the full treatment and 41 from the partial treatment.

Figure E.3: Distribution of Proportions of Copied Queries by Experimental Groups



Notes: This figure shows the density plot of the distribution of the proportion of copied queries by experimental groups, collected from 88 participants, 47 from the full treatment and 41 from the partial treatment.

Table E.1 Types of Queries in the Full Treatment Group and the Partial Treatment Group

Query Type	Full Treatment Mean	Partial Treatment Mean	T-statistics	p-value
Problem	1.72	0.10	8.98	<0.001
Option	1.09	1.95	-3.44	0.001
Reflection	0.77	0.05	5.34	<0.001
-- Problem reflection	0.51	0.00	6.40	<0.001
-- Option reflection	0.21	0.05	2.36	0.021
-- Other reflection	0.04	0.00	1.43	0.160
Selection	0.11	0.00	1.94	0.058
Other Survey Questions	0.32	0.049	2.43	0.018
Contextual Information	0.68	0.19	2.40	0.019
Other	0.26	0.073	1.64	0.106

Note: Each LLM query submitted by participants was manually coded into one of nine mutually exclusive categories. The table reports the average raw count of queries per participant by type, separately for the full and partial treatment groups. T-tests compare the means across the two groups to assess differences in query behavior.

Table E.2 Examples of Prompts Participants Used

Participant Index	Experimental Group	Query Type	Prompts
1	Full treatment	Problem	Can you please provide input regarding the potential causes of decline in profitability of an online restaurant review platform (Rated)? Please segment the answers by analyzing the following things: market trends and growth, competitors landscape, competitive positioning, quality of platform, customer behaviour, operational inefficiencies, etc. Please provide your answer in different bullet points per idea
2	Full treatment	Option	How do we decrease general and administration cost
3	Full treatment	Contextual info	what is Rated competitive advantage?
4	Partial treatment	Option	List of strategic options for Rated
5	Partial treatment	Option	can rated enter chained customer busienss?
6	Partial treatment	Contextual info	How familiar are you with "rated"?

Note: This table presents a selection of participant-submitted prompts directed to the LLM, categorized by query type and experimental group.

E.2 Attention Dilution

E.2.1 Coding of Problem Categories and Option Categories

Using a self-developed dictionary shown in Table E.3, we identified the problem categories covered in the participant's response to the survey question "What are the strategic reasons that could be causing the profit decline of Rated?"

Columns 2 to 11 of Table C.3 report the impact of LLM-assisted problem formulation on specific categories. Results indicate that LLM-assisted problem formulations are more likely to cover categories related to competition, financial, positioning, and user considerations. Specifically, the use of LLM increased the likelihood of mentioning competition-related issues by 16 percentage points (p = 0.017), financial concerns by 13 percentage points (p = 0.055), positioning strategies by 10 percentage points (p = 0.09), and user-related factors by 16 percentage points (p =0.014). This pattern suggests that LLM assistance encourages participants to consider a more diverse and comprehensive set of strategic factors in their problem formulations. Interestingly, we observe no significant impact on other categories such as cost, growth, product, or business model.

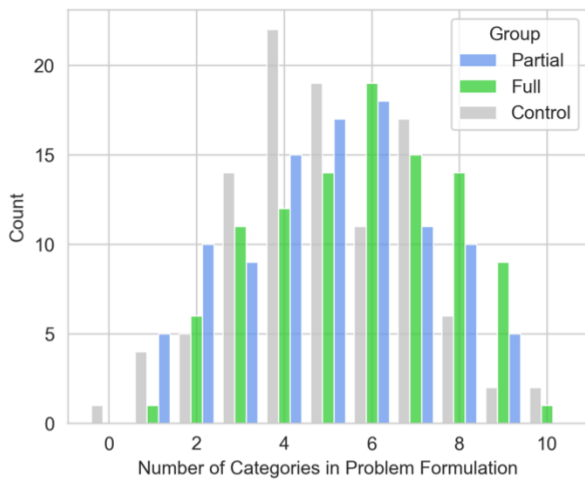
Similarly, we coded the options for these categories using a self-developed dictionary shown in Table E.3. We then calculated the average options per category as:

$$Avg\ option\ per\ category = \frac{\#\ of\ options\ in\ total}{\#\ of\ categories\ covered\ in\ options}$$

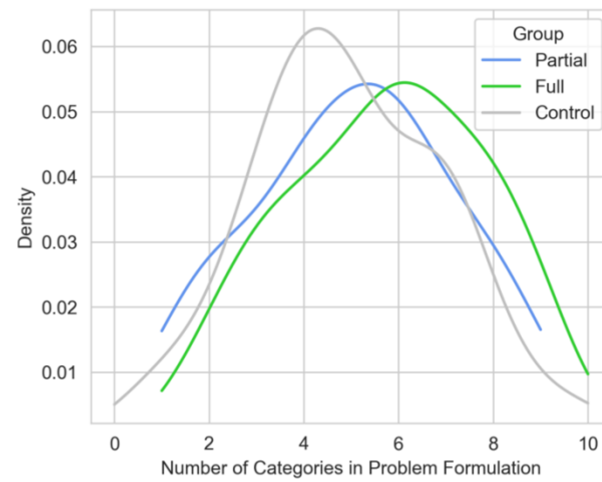
Figures E.4 through E.8 present descriptive analyses of category distributions in both problem formulation and option generation. Tables E.4 and E.5 present the regression models estimating the impact of treatments on the categories covered in problems and options.

Figure E.4: Distribution of the Number of Categories in Problem Formulation

(a) Histogram

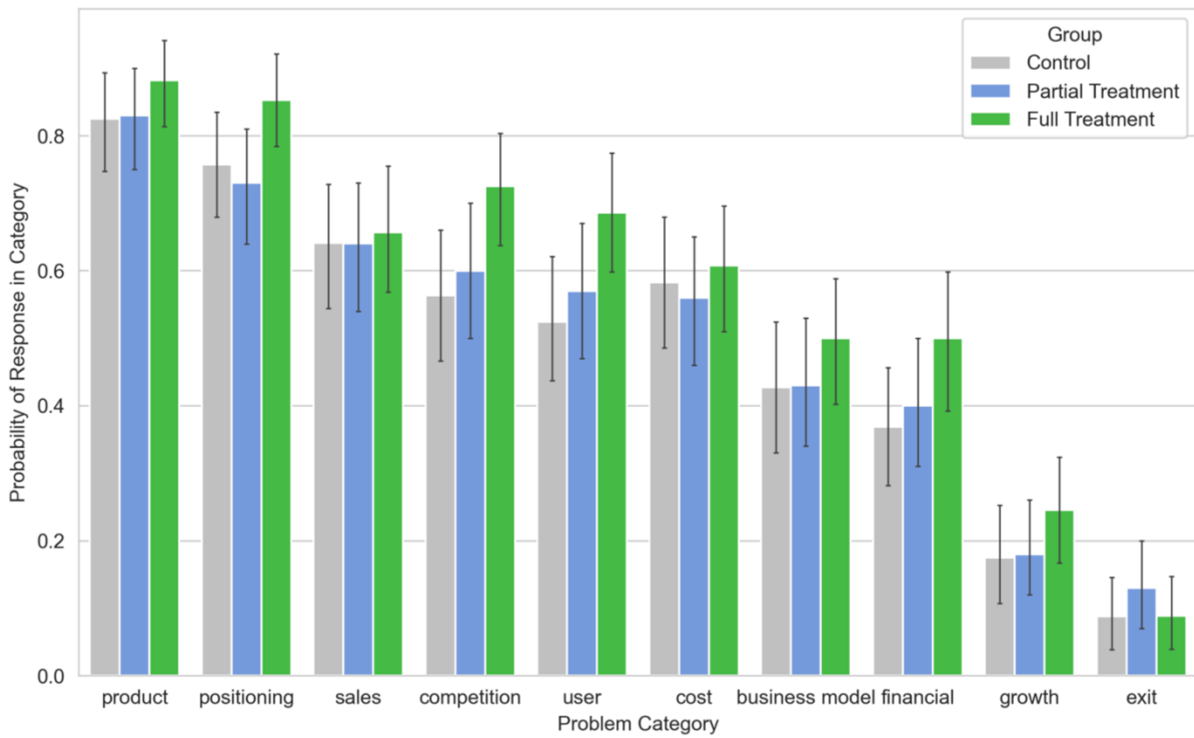


(b) Density distribution



Notes: These figures show the histogram and density distribution of the number of categories in problem formulation of the Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Control vs. Full Treatment K-S Test p-value: 0.028; Control vs. Partial Treatment K-S Test p-value: 0.937; Full Treatment vs. Partial Treatment K-S Test p-value: 0.332.

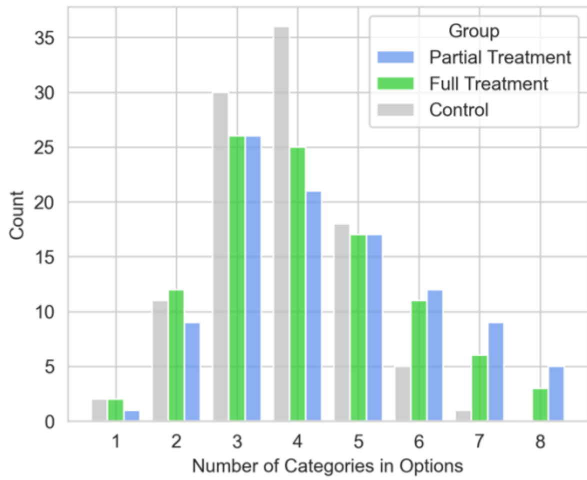
Figure E.5: Comparison of Category Distributions



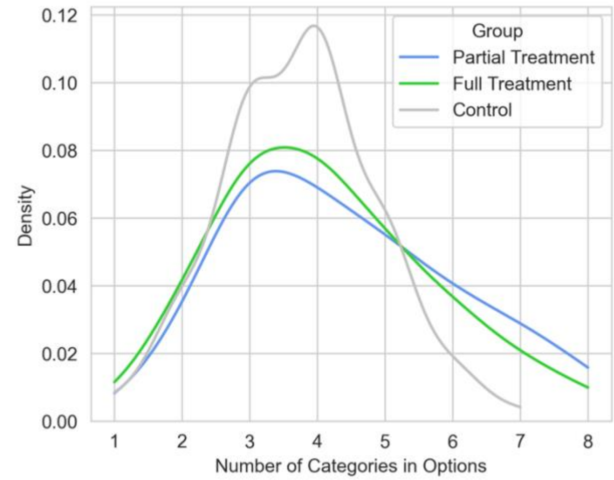
Notes: This figure plots the result of the comparison of the categories across the experimental groups analysis. The x-axis shows the categories. The y-axis shows the average probability that the problem formulation covers that category.

Figure E.6: Distribution of the Number of Categories in Options

(a) Histogram

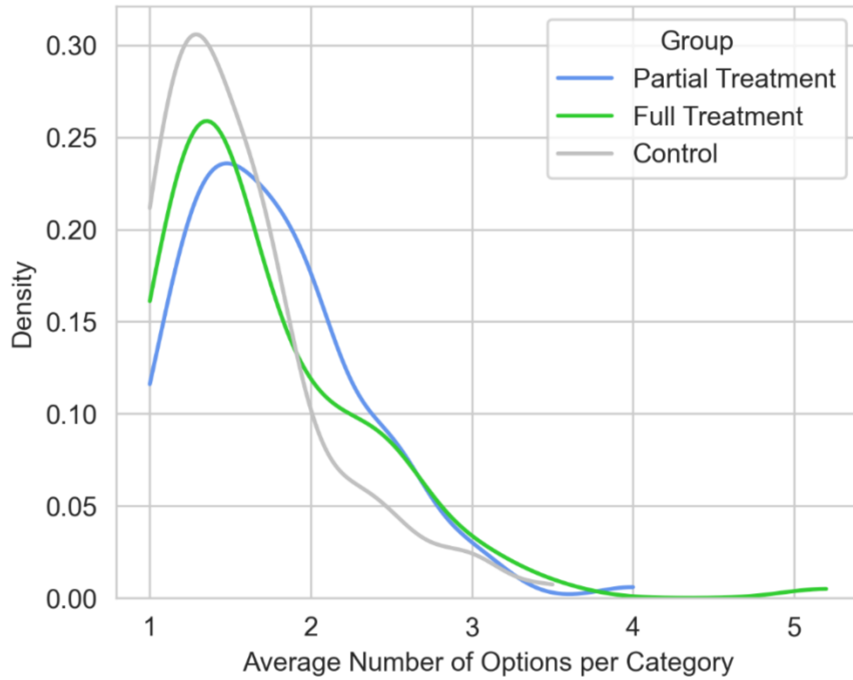


(b) Density distribution



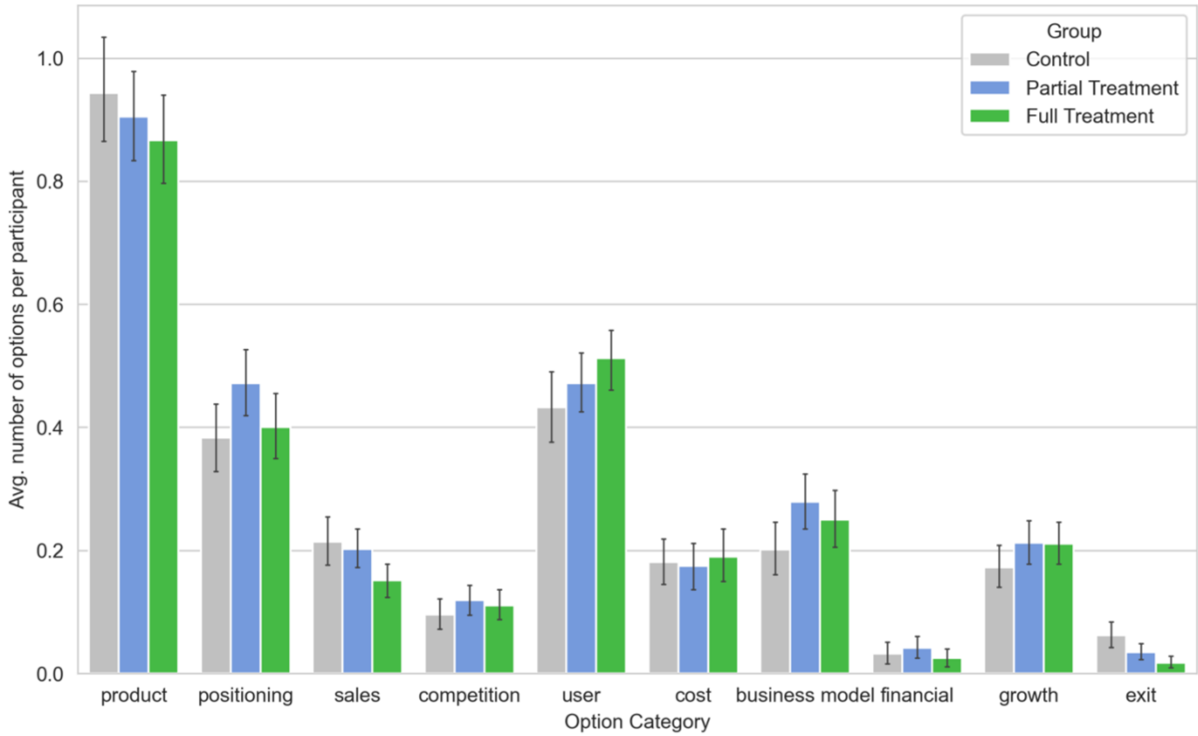
Notes: These figures show the histogram and density distribution of the number of categories in options of the Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Control vs. Full Treatment K-S Test p-value: 0.258; Control vs. Partial Treatment K-S Test p-value: 0.026; Full Treatment vs. Partial Treatment K-S Test p-value: 0.957.

Figure E.7: Density Distribution of the Average Number of Options per Category



Notes: These figures show the density distribution of the average number of options per category of the Full Treatment group (in green), the Partial Treatment group (in blue), and the Control group (in grey). Control vs. Full Treatment K-S Test p-value: 0.251; Control vs. Partial Treatment K-S Test p-value: 0.009; Full Treatment vs. Partial Treatment K-S Test p-value: 0.185.

Figure E.8: Comparison of Category Distributions



Notes: This figure plots the result of the comparison of the categories across the experimental groups analysis. The x-axis shows the categories. The y-axis shows the average number of options per participant that covers the category.

Table E.3: Categories and Keywords

Category	**Keywords**
Competition	competition, competitor, competitive, compete
Growth	growth, expansion, scaling, acquir, expand, joint venture, jv, strategic alliance, strategic partnership
Exit	sell, exit, close, divest, sale of the business
Financial	profit, profitability, margin, financial
Cost	cost, expense, charge, spending, efficienc, g&a, general administration, operation
Sales	sales, incentive, promotion, loyalty
Product	product, ads, advertising, advertise, adtech, feature, technology, algorithm, review, reservation, booking, delivery, app, api, network, data analytics, quality
Positioning	positioning, position, market, segment, target, differentiation, local, chains, verified rating
User	user, engagement, customer, client, restaurant
Business Model	monetiz, business model, revenue model, strategy, plan, pay per click, subscription, freemium, premium, streams of revenue, revenue streams, pricing, capitalize

Table E.4: Impact on Problem Formulation

VARIABLES	(1) Count	(2) Competition	(3) Growth	(4) Exit	(5) Financial	(6) Cost	(7) Sales	(8) Product	(9) Position-ing	(10) User	(11) Business Model
Full Treatment	0.82***	0.16**	0.08	-0.00	0.13*	0.02	0.02	0.06	0.10*	0.16**	0.08
	(0.29)	(0.07)	(0.06)	(0.04)	(0.07)	(0.07)	(0.07)	(0.05)	(0.06)	(0.07)	(0.07)
Partial Treatment	0.13	0.03	0.01	0.04	0.04	-0.02	0.00	0.01	-0.02	0.04	-0.00
	(0.29)	(0.07)	(0.05)	(0.04)	(0.07)	(0.07)	(0.07)	(0.05)	(0.06)	(0.07)	(0.07)
Constant	4.64***	0.52***	0.18**	0.07	0.19**	0.61***	0.55***	0.80***	0.77***	0.54***	0.43***
	(0.37)	(0.09)	(0.07)	(0.05)	(0.08)	(0.09)	(0.09)	(0.07)	(0.08)	(0.09)	(0.09)
Observations	305	305	305	305	305	305	305	305	305	305	305
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.02	0.00	0.00	-0.02	0.01	0.00	0.02	-0.02	0.03	0.03	-0.02
Control Group Mean	4.95	0.56	0.17	0.09	0.37	0.58	0.64	0.83	0.76	0.52	0.43
F-test	5.38	3.51	1.34	0.78	1.92	0.43	0.07	1.01	4.42	3.15	1.32
Prob > F	0.021	0.062	0.248	0.379	0.166	0.514	0.793	0.316	0.036	0.077	0.251

Notes: This table reports regression results for equation (1). “Count” indicates the number of categories covered by the problem formulation. All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table E.5: Impact on Option Categories

VARIABLES	(1) Count	(2) Competition	(3) Growth	(4) Exit	(5) Financial	(6) Cost	(7) Sales	(8) Product	(9) Position-ing	(10) User	(11) Business Model	(12) Avg. Options per Cat
Full Treatment	0.38*	0.02	0.20*	-0.08*	0.04	0.13	-0.16*	0.32	0.25**	0.42***	0.23***	0.15*
	(0.20)	(0.09)	(0.11)	(0.04)	(0.04)	(0.10)	(0.09)	(0.23)	(0.11)	(0.14)	(0.08)	(0.08)
Partial Treatment	0.67***	0.09	0.36***	0.02	0.08*	0.15	0.05	0.53**	0.49***	0.27**	0.22**	0.21***
	(0.21)	(0.09)	(0.12)	(0.05)	(0.05)	(0.10)	(0.11)	(0.25)	(0.14)	(0.12)	(0.09)	(0.08)
Constant	3.63***	0.26**	0.56***	0.04	0.01	0.36***	0.50***	1.86***	0.63***	0.42***	0.31***	1.44***
	(0.22)	(0.11)	(0.16)	(0.04)	(0.04)	(0.11)	(0.12)	(0.26)	(0.17)	(0.14)	(0.10)	(0.08)
Observations	305	305	305	305	305	305	305	305	305	305	305	305
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.01	0.00	0.01	0.03	-0.01	0.01	0.04	0.02	0.03	0.07	-0.00	0.05
Control Group Mean	3.74	0.31	0.52	0.14	0.05	0.49	0.58	2.23	0.50	0.44	0.27	1.56
F-test	1.45	0.61	1.52	5.07	0.59	0.03	4.47	0.62	2.42	0.82	0.01	0.45
Prob > F	0.230	0.435	0.218	0.025	0.445	0.870	0.035	0.431	0.121	0.366	0.932	0.503

Notes: This table reports regression results for equation (1). "Count" indicates the number of categories covered in the options. All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

E.2.2 Overlap Between Problem and Option Categories

Figure E.9 illustrate how we define and calculate overlap. For each participant i, let A_i denote the set of categories mentioned in the problem formulation (left circle), and B_i the set of categories mentioned in the options (right circle). Their intersection $C_i = A_i \cap B_i$ captures categories that are carried through from problem formulation to ideation (overlap). The overlap between problem and option categories is defined as the share of problem categories that are pursued in options:

$$Overlap_i = \frac{\#\ of\ categories\ covered\ in\ both\ problem\ and\ options\ (|C_i|)}{\#\ of\ categories\ covered\ only\ in\ problem(|A_i|)}$$

The set difference $A_i - C_i$ comprises problem-only categories. These are issues raised in problem formulation but not acted upon in options. This measure captures the drop-off from initial framing. The set difference $B_i - C_i$comprises option-only categories. These are issues introduced during ideation without prior mention in the problem formulation. This measure captures the expansion beyond initial framing.

Table E.6 shows that, compare to control group, the full treatment (LLM in problem formulation + ideation) increases the absolute number of categories carried from problem to options (Column 1, 20.4% of the control mean), with no significant change in the share of problem categories that are pursued or dropped (Columns 4 and 5). More categories are carried through in level terms, but alignment as a proportion of the problem frame is unchanged. Partial treatment (LLM in ideation only) expands beyond the problem frame: it raises the number of option-only categories (Column 3, 29.8% of the control mean) and the proportion of options outside problem formulation (Column 6). Direct comparisons between treatments (F-tests) show the partial treatment produces a higher raw count and share of options outside the problem frame (Columns 3 and 6), while the full treatment tends to have more drop-off than the partial treatment (Column 2). Overall, LLMs used during problem formulation anchor subsequent ideation within the stated frame (greater absolute carry-through), whereas using LLMs only during ideation promotes exploration beyond the initial frame.

Figure E.9: Illustration of overlap

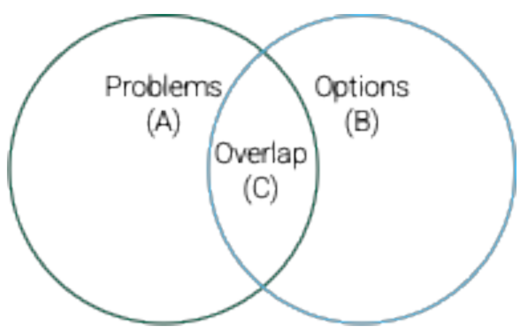


Table E.6: Impact on Overlap (reference group: control)

VARIABLES	(1) Size of overlap	(2) Problem-only categories (size)	(3) Option-only categories (size)	(4) Proportion of problem categories pursed (Overlap)	(5) Proportion of problem categories dropped	(6) Proportion of options outside problem categories
Full Treatment	0.51**	0.31	-0.13	0.01	-0.01	-0.05
	(0.20)	(0.24)	(0.15)	(0.03)	(0.03)	(0.04)
Partial Treatment	0.30	-0.18	0.37**	0.05	-0.05	0.03
	(0.19)	(0.23)	(0.18)	(0.04)	(0.03)	(0.04)
Constant	2.33***	2.31***	1.31***	0.55***	0.44***	0.35***
	(0.23)	(0.30)	(0.21)	(0.05)	(0.05)	(0.05)
Observations	305	305	305	305	304	305
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.02	0.00	0.02	-0.00	-0.00	0.01
Control Group Mean	2.50	2.46	1.24	0.52	0.47	0.32
F-test	0.93	4.22	7.38	1.42	1.42	3.91
Prob > F	0.335	0.041	0.007	0.235	0.235	0.049

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table E.7: Impact on Overlap (reference group: partial treatment)

VARIABLES	(1) Size of overlap	(2) Problem-only categories (size)	(3) Option-only categories (size)	(4) Proportion of problem categories pursed (Overlap)	(5) Proportion of problem categories dropped	(6) Proportion of option-only categories
Full Treatment	0.21	0.49**	-0.48***	-0.04	0.04	-0.07*
	(0.22)	(0.24)	(0.18)	(0.04)	(0.04)	(0.04)
Constant	2.36***	2.48***	1.85***	0.55***	0.45***	0.42***
	(0.30)	(0.41)	(0.31)	(0.06)	(0.06)	(0.06)
Observations	202	202	202	202	202	202
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.02	0.01	0.04	0.00	0.00	0.04
Partial Group Mean	2.79	2.28	1.62	0.57	0.43	0.35

Notes: This table reports regression results for $y_{is} = \beta_0 + \beta_1 T_i^F + \gamma_s + \epsilon_i$. It includes the full treatment group and the partial treatment group only to compare the difference between these two conditions. All regression models include fixed effects for the randomization strata (gender-consulting-section). Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

E.3 Cognitive Drain

E.3.1 Quality of ideas over time

We analyzed option-level data to investigate whether the strategic quality of participants' options declines over time. Each participant generated multiple options in the task, with each option indexed in the order it was produced. Each option is coded as Strategic or not.

To assess whether the likelihood of generating a strategic idea declines over time, and whether this temporal pattern varies by treatment condition, we estimate a multilevel logistic regression model. The key independent variables include the sequential index of the idea and a binary indicator for assignment to the full treatment group, along with their interaction term. A random intercept is included at the respondent level to account for within-subject correlation in idea quality. Column 7 in Table 6 in the main text presents the results.

We also look at the first option and the last option, and see if the first option is more likely to be strategic than the last option for the full treatment group. We calculated

$$Delta\ strategic = first\ is\ strategic - last\ is\ strategic$$

and regressed it on the treatment group indicators using robust standard errors. Column 6 in Table 6 in the main text presents the results.

APPENDIX F: PRE-REGISTRATION DIFFERENCES

The experiments were pre-registered with the AEA RCT Registry. The key differences between the paper and the pre-registration are:

- We pre-registered 336 students based on expected enrollment, excluding two students initially categorized as "partners." Because these students had similar characteristics and eligibility as the main sample, we later included them, bringing the assigned sample to 338. Of these, 305 participated (including the two additions), and 33 were absent.
- An attrition analysis comparing observable characteristics of participants and non-participants shows limited imbalance (Table F.1). Only prior consulting experience differs significantly across groups at the 5% level. Results remained consistent if we excluded the two additions (Table F.2), when controlling for consulting experience (Table F.3) and when additionally controlling for gender and age (Table F.4).
- We pre-registered three additional outcomes that are not included in the current analysis: (1) how participants prioritized problems when multiple were identified — participants were asked to identify the root cause of the problems without prioritizing them; (2) whether the set of options were mutually exclusive (binary); and (3) whether the best-chosen option reflected an "exit" or "expand" strategy (binary). These were initially registered as primary outcomes but were not analyzed due to a shift in focus toward more theoretically grounded measures of strategic focus. We report the regression results for (2) and (3) in Table F.5.
- We also pre-registered the number of "continue" options as an outcome variable. To maintain focus in the main text on broader patterns of strategic reasoning, we report results for this variable in Appendix D, where we examine whether LLM assistance increases participants' tendency to maintain the status quo rather than propose substantive change.
- The pre-analysis plan describes secondary outcomes, some of which we noted at the time as subject to feasibility. We were indeed not able to obtain peer evaluations of option quality and a binary variable on how detailed the option is.
- The pre-analysis plan describes that the alternatives generated by participants would be coded by two independent coders. Instead, the options were coded using a fine-tuned GPT-3.5 Turbo model trained on previously human-coded data due to budget constraints.

Table F.1: Attrition Analysis

	(1) Full		(2) Absence		(3) Participated		(4) Diff (3) – (2)		
	Mean	SD	Mean	SD	Mean	SD	b	se	p
Female	0.33	0.47	0.45	0.51	0.32	0.47	-0.14	0.09	0.114
Consulting	0.33	0.47	0.48	0.51	0.31	0.47	-0.17	0.09	0.049
Age	30.08	2.51	30.12	2.94	30.07	2.47	-0.05	0.46	0.915
Observations	338		33		305		338		

Notes: Columns (2) and (3) report means and standard deviations by attrition status. Column (4) reports mean differences and p-values from t-tests.

Table F.2: Impact on Ideation and Selection

VARIABLES	(1) Number of Options	(2) Number of Strategic Options	(3) Number of Operational Options	(4) Proportion of Strategic Options	(5) Proportion of Operational Options	(6) Best Option is Strategic
Full Treatment	1.53***	0.51	1.02***	-0.07**	0.07**	-0.15**
	(0.44)	(0.32)	(0.27)	(0.03)	(0.03)	(0.07)
Partial Treatment	2.65***	1.53***	1.13***	0.00	-0.00	0.00
	(0.55)	(0.36)	(0.33)	(0.03)	(0.03)	(0.07)
Constant	5.18***	3.01***	2.16***	0.56***	0.44***	0.62***
	(0.59)	(0.43)	(0.30)	(0.04)	(0.04)	(0.09)
Observations	303	303	303	303	303	303
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.08	0.06	0.06	0.02	0.02	0.01
Control Group Mean	5.69	3.26	2.44	0.58	0.42	0.64
F-test	2.97	5.81	0.09	5.66	5.66	4.47
Prob > F	0.086	0.017	0.767	0.018	0.018	0.035

Notes: This table reports regression results for equation (1), excluding 2 additions after pre-registration. All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table F.3: Impact on Ideation and Selection (with consulting control)

VARIABLES	(1) Number of Options	(2) Number of Strategic Options	(3) Number of Operational Options	(4) Proportion of Strategic Options	(5) Proportion of Operational Options	(6) Best Option is Strategic
Full Treatment	1.55***	0.53*	1.02***	-0.07**	0.07**	-0.15**
	(0.43)	(0.31)	(0.27)	(0.03)	(0.03)	(0.07)
Partial Treatment	2.68***	1.55***	1.13***	0.00	-0.00	-0.01
	(0.55)	(0.36)	(0.33)	(0.03)	(0.03)	(0.07)
Consulting	-0.33	-0.82	0.49	-0.06	0.06	-0.08
	(1.01)	(0.65)	(0.83)	(0.09)	(0.09)	(0.18)
Constant	5.16***	3.00***	2.16***	0.56***	0.44***	0.62***
	(0.59)	(0.43)	(0.30)	(0.04)	(0.04)	(0.09)
Observations	305	305	305	305	305	305
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.08	0.07	0.06	0.02	0.02	0.01
Control Group Mean	5.68	3.23	2.45	0.57	0.43	0.65
F-test	2.97	5.79	0.09	5.65	5.65	4.47
Prob > F	0.086	0.017	0.766	0.018	0.018	0.035

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section) and control for consulting experience (binary indicator). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table F.4: Impact on Ideation and Selection (with full control)

VARIABLES	(1) Number of Options	(2) Number of Strategic Options	(3) Number of Operational Options	(4) Proportion of Strategic Options	(5) Proportion of Operational Options	(6) Best Option is Strategic
Full Treatment	1.57***	0.52*	1.05***	-0.07**	0.07**	-0.15**
	(0.44)	(0.31)	(0.28)	(0.03)	(0.03)	(0.07)
Partial Treatment	2.70***	1.54***	1.16***	0.00	-0.00	-0.01
	(0.56)	(0.37)	(0.33)	(0.03)	(0.03)	(0.07)
Female	0.46	-0.23	0.69	-0.02	0.02	0.09
	(0.87)	(0.55)	(0.55)	(0.06)	(0.06)	(0.14)
Consulting	-0.96	-0.56	-0.41	-0.02	0.02	-0.18
	(1.05)	(0.58)	(0.93)	(0.10)	(0.10)	(0.20)
Age	-0.06	0.01	-0.07	0.00	-0.00	-0.00
	(0.09)	(0.06)	(0.05)	(0.01)	(0.01)	(0.01)
Constant	6.93**	2.68	4.25**	0.42**	0.58***	0.70*
	(2.81)	(1.84)	(1.73)	(0.17)	(0.17)	(0.38)
Observations	305	305	305	305	305	305
Gender-Consulting-Section FE	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.08	0.06	0.06	0.02	0.02	0.00
Control Group Mean	5.68	3.23	2.45	0.57	0.43	0.65
F-test	2.97	5.76	0.09	5.65	5.65	4.46
Prob > F	0.086	0.017	0.761	0.018	0.018	0.036

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section) and control for gender (binary indicator), consulting experience (binary indicator), and age. We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table F.5:Additional Outcome Variables

VARIABLES	(1) Mutually Exclusive	(2) Number of Change Pricing Options	(3) Number of Exit Options
Full Treatment	-0.01	-0.14*	-0.13***
	(0.03)	(0.08)	(0.04)
Partial Treatment	-0.04*	-0.01	-0.12***
	(0.02)	(0.09)	(0.04)
Constant	0.02	0.49***	0.13***
	(0.01)	(0.10)	(0.05)
Observations	305	305	305
Gender-Consulting-Section FE	YES	YES	YES
Adjusted R-squared	0.01	0.02	0.04
Control Group Mean	0.04	0.52	0.16
F-test	2.91	2.55	0.52
Prob > F	0.089	0.111	0.472

Notes: This table reports regression results for equation (1). All regression models include fixed effects for the randomization strata (gender-consulting-section). We also report the F-test statistics and corresponding p-values to assess whether the coefficients for full treatment and partial treatment are statistically different. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

APPENDIX G: FOLLOW-UP EXPERIMENT

To enrich the theory developed through the main experiment, we conducted a supplementary team-level study involving 98 high-level executives enrolled in an executive education program for senior managers. Participants were randomly assigned to 24 teams of four to five members. Given the limited statistical power at the team level, the objective of this supplementary study was not to replicate individual-level findings. Rather, it aimed to generate qualitative insight on how LLMs interact with the strategic decision-making process in real-world, team-based contexts.

Similar to the main experiment, we randomly varied when the custom web interface (based on gpt-4o-mini, temperature is set to 0.7) was available (Figure G. 1): half of the teams received access for both the problem formulation and ideation stages (full treatment), while the other half gained access at the ideation stage (partial treatment). Unlike the main experiment, which used a standardized, pre-distributed case, this supplementary study drew on strategic challenges that participants were facing in their own companies. Each team worked on a problem contributed by one of its members, which spanned topics as diverse as competitive repositioning, declining profitability, or market expansion (Table G.1 provides further details). The teams collaboratively progressed through three key stages: problem formulation, alternative generation, and selection, following a similar structure to the main experiment. The group task lasted approximately two hours. Immediately after the group task, participants completed an individual online survey capturing their reflections on the team process and their perceptions of the LLM's contribution.

Figure G.1 Custom Web Interface

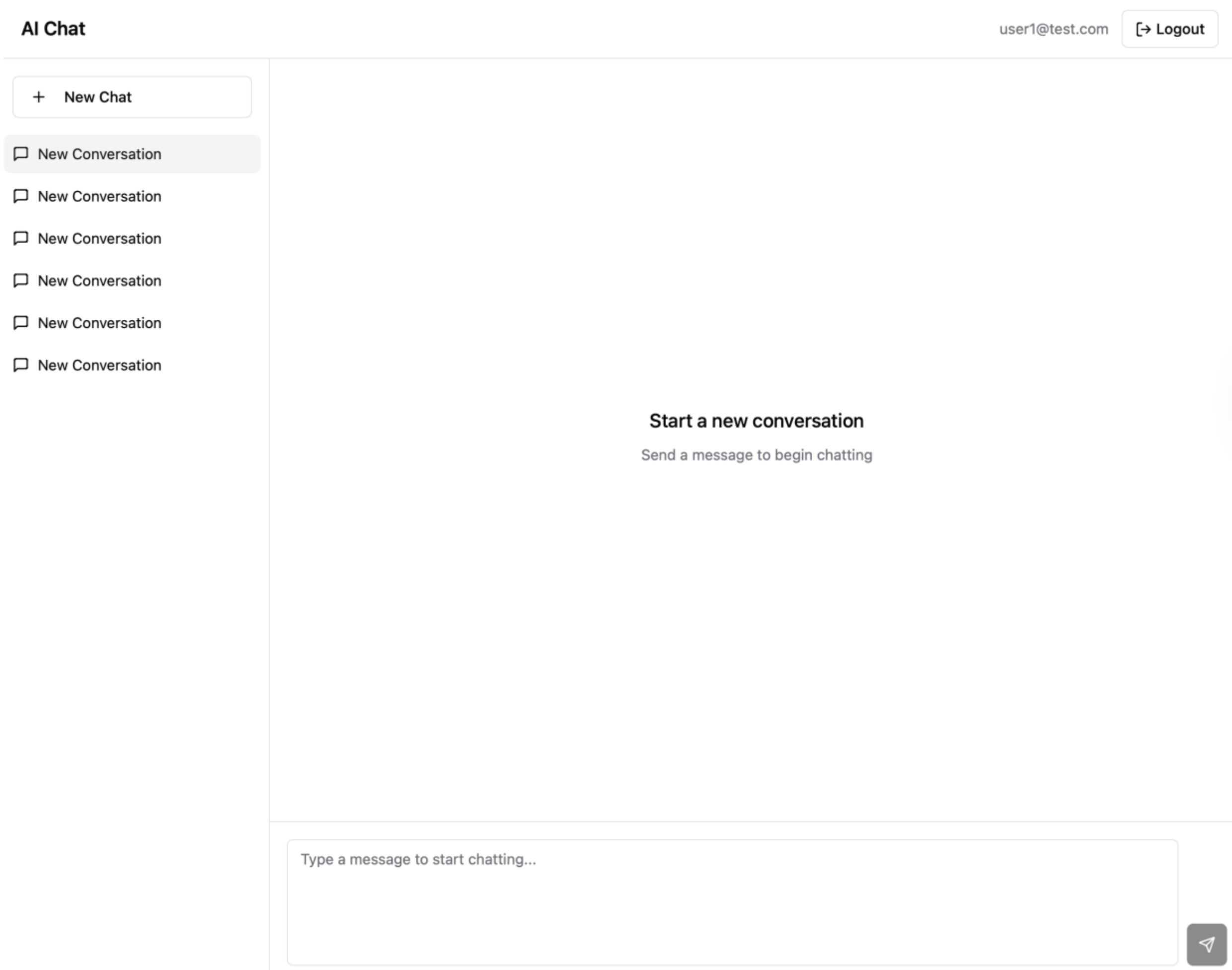


Table G.1: Strategic Challenges presented to each Group

Treatment	Group	Strategic Challenge
Full treatment	1	A digital streaming platform undergoing brand transformation is attempting to reposition itself. It is competing against a well-established legacy brand while working under significant constraints in time, budget, and team alignment.
Full treatment	2	How can an energy company sustain output and remain relevant amid aging assets and the global shift to renewables? Key issues we are currently facing include reservoir pressure depletion, rising costs of oil and gas production, aging infrastructure such as pipelines, surface facilities, and equipment life, as well as the superannuation of skilled manpower within the company. Additionally, the constant fluctuation of oil prices has made production decisions in our maturing fields increasingly difficult. We are also experiencing competitive pressure from new entrants. At present, renewable energy comprises only a small portion of our portfolio.
Full treatment	3	A creator economy platform is trying to develop a sustainable growth model in a market where the user demand is promising but the creator ecosystem is underdeveloped.
Full treatment	4	The startup is struggling to define and execute a go-to-market strategy that is supported by both external stakeholders (such as ship owners, charterers, and financial institutions) and internal stakeholders. Leadership is currently considering how to market the offering, build credibility, and win the trust of external stakeholders, while also determining how to drive adoption among internal stakeholders.
Full treatment	5	The company is facing lower growth rates and need to reach aggressive revenue targets. There is an existing business growing at a decling rate due to market maturity, hence the pressure to find scalable new revenue streams.
Full treatment	6	A key competitor gained traction in a region central to our user-generated content (UGC) base. This resulted in a significant user decline, leading to a drop in engagement and related performance metrics.
Full treatment	7	A cloud data platform is struggling to identify and convert small and medium business customers in a cost-sensitive market. Despite the product's technical strength, the company is unable to generate sufficient sales pipeline or conversions in this segment.
Full treatment	8	How can we respond to a well-funded competitor's aggressive market entry, whose deep discounts have challenged our premium brand positioning.
Full treatment	9	The company is facing economic downturn, contract delays, and shareholder pressure. Revenue and earnings projections are threatened by customer hesitancy and uncertainty about market direction.
Full treatment	10	A sports management company is facing stagnation in their core on-ground branding business due to limited inventory growth, brand fatigue, and shifting media consumption and spending habits driven by the rise of digital advertising. How can they leverage their exclusive rights to unlock new growth verticals and evolve into a multi-vertical sports marketing powerhouse over the next five years?
Full treatment	11	A simulation software company with a low-cost, lean product is struggling to gain share from feature-rich incumbents. Market penetration has plateaued due to product limitations and internal resource bottlenecks.
Full treatment	12	One of our lines of business, Indirect Tax, is struggling with revenue growth and thinning margins. As tax compliance has become increasingly automated and commoditized, the practice is now just breaking even. However, it remains strategically important for us to retain it in order to offer full-service, end-to-end solutions to clients.

Partial treatment	13	The company acquired a content management services provider in APAC, which is rapidly losing clients after founder exister post acquisition. How to turn the business around and navigate people challenges attached to the acquisition?
Partial treatment	14	As a streaming platform, we are distributing through telco bundles. We have scale, but limited control, low customer stickiness, and low lifetime value. How can we monetize in the long-term, especially as subscriber growth is expected to plateau or decline with a shrinking share of wallet?
Partial treatment	15	A technology startup is struggling to acquire and build trust with potential customers who are considering international expansion but are hesitant to engage a new, relatively unknown tech partner.
Partial treatment	16	With global disruptions to paper supply and a steep drop in readership due to misinformation about newspapers being Covid carriers, the costs began to far exceed revenues for a daily newspaper company, threatening the viability of several city editions.
Partial treatment	17	Whether and how to launch B2B SaaS product in Australia region after strong early signals, but complex supply chains and lead times.
Partial treatment	18	We have been in the natural diamonds trading and jewellery retail for over 30 years. Lately, lab grown diamonds are a major issue due to improved process of manufacture and cost effectiveness. They are eroding the market for natural diamonds. How should we respond?
Partial treatment	19	A tech firm is grappling with how to transition to an AI-first strategy without clear customer demand or internal roadmap. There is ambiguity about what AI solutions the market needs and how internal teams should prioritize development.
Partial treatment	20	Two leaders in the government consulting space, instrumental in driving sales, departed the firm. In addition, government spending declined, leaving over 50 professionals in the government consulting team underutilized. The future of the government consulting business looked increasingly uncertain.
Partial treatment	21	A company must decide how to acquire a mission-critical technology: whether to build it internally or buy externally.
Partial treatment	22	As an early entrant and market leader in the Indian Railway Signaling sector, we are facing a high cost of imported components, which made us vulnerable to new entrants offering cheaper alternatives. How can we reclaim market share and restore margins?
Partial treatment	23	The company is facing the challenge of how to expand its energy network rapidly and efficiently in a market where growth opportunities are unevenly distributed across regions. It must decide how to balance near-term commercial certainty with long-term strategic positioning, while navigating operational constraints and competitive pressures.
Partial treatment	24	How to revive growth in our HNI segment, where our market share had stagnated despite having strong products? We were relying on underperforming distributor channels, had limited internal bandwidth to build direct relationships, and the HNI market itself was fragmented with low brand loyalty.

Table G.2: Summary Statistics of the Follow-up Study

	(1) Full sample					(2) Full Treatment			(3) Partial Treatment			(4) Difference (3)-(2)		
	N	mean	SD	min	max	N	mean	SD	N	mean	SD	b	se	p
Individual-level														
Gender	96	0.29	0.46	0	1	48	0.29	0.46	48	0.29	0.46	0.00	0.09	1.000
Consultant	96	0.04	0.20	0	1	48	0.04	0.20	48	0.04	0.20	0.00	0.04	1.000
Team-level														
Number of Options	24	4.83	1.31	3	8	12	5.08	1.62	12	4.58	0.90	0.50	0.54	0.360
Number of Strategic Options	24	3.38	1.71	0	7	12	3.58	1.98	12	3.17	1.47	0.42	0.71	0.563
Number of Continue Options	24	1.88	1.48	0	5	12	1.92	1.68	12	1.83	1.34	0.08	0.62	0.894
Proportion of Strategic Options	24	0.69	0.31	0	1	12	0.68	0.33	12	0.69	0.29	-0.01	0.13	0.958
Proportion of Continue Options	24	0.40	0.31	0	1	12	0.39	0.34	12	0.40	0.29	-0.01	0.13	0.947
Duration (min)	24	123	31	20	163	12	125	38	12	121	23	4.41	12.80	0.734

Notes: This table reports the summary statistics of baseline variables and main outcome variables across experimental groups, as well as the pairwise comparisons in means between treatment conditions.

References

Burgelman, R. A. (2002). Strategy as Vector and the Inertia of Coevolutionary Lock-in. *Administrative Science Quarterly*, *47*(2), 325–357. https://doi.org/10.2307/3094808

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

Gilbert, C. G. (2005). Unbundling the Structure of Inertia: Resource Versus Routine Rigidity. *Academy of Management Journal*, *48*(5), 741–763. https://doi.org/10.5465/amj.2005.18803920

Henderson, R. M., & Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, *35*(1), 9–30. https://doi.org/10.2307/2393549

Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, *21*(10–11), 1147–1161. https://doi.org/10.1002/1097-0266(200010/11)21:10/11<1147::AID-SMJ128>3.0.CO;2-R