Individual and Organizational Antecedents of Strategic Foresight—A Representational Approach*

Felipe A. Csaszar
Ross School of Business, University of Michigan
Ann Arbor, MI 48109
fcsaszar@umich.edu

Daniella Laureiro-Martínez
Department of Management, Technology and Economics, ETH Zurich
Zurich, Switzerland
dlaureiro@ethz.ch

July 18, 2018

Forthcoming in *Strategy Science*

*The authors thank the seminar participants at Columbia, Wharton, University of Michigan, University of Illinois, University of Vienna, the 2015 Workshop on Experimental Organization Science at Frankfurt, and the 2016 Meeting of the Academy of Management. The authors also thank José Arrieta, Andrea Casali, and Martin Sterchi for excellent research assistance. Any errors are the authors' own responsibility.

Individual and Organizational Antecedents of Strategic Foresight—A Representational Approach

Abstract

The ability to make predictions about strategic outcomes—what we term *strategic foresight*—is central to most theories of competitive advantage. This paper identifies individual- and organization-level antecedents of strategic foresight by analyzing an exercise taken by 358 MBA students. Among the individual antecedents, we show that two characteristics of mental representations (namely, their breadth and agreement with consensus) are positively related to strategic foresight. Comparing individual to group performance reveals that groups exhibit greater strategic foresight than do individuals. Finally, from comparing the performance of real-life groups with "statistical" groups (for which decisions are computed by averaging the predictions of individuals before they become group members), we find that the superiority of group performance is due mostly to aggregating predictions, not representations.

Keywords: strategic foresight; managerial cognition; decision-making structure; aggregation; representation; upper echelons

1 Introduction

1.1 Foresight in strategy

Foresight—the ability to predict[1]—is central to many theories of competitive advantage. For instance, according to the resource-based view, the main way in which a firm can generate above-normal profits is by being better than its competitors at predicting the value that a resource will generate once it is bought, sold, or developed (Barney 1986). Similarly, the "positioning" school of thought argues that profits stem from being able to accurately predict the future attractiveness of the industries the firm can enter (Porter 1980) while the value-based view holds that profits stem from accurately predicting the firm's opportunities to create and capture value vis-à-vis its competitors, customers, and suppliers (Brandenburger and Stuart 1996). Since these and other theories postulate that competitive advantage is caused by either luck or foresight (Ahuja et al. 2005:792) and since, of these two causes only the latter can be controlled, it follows that foresight is arguably among the most fundamental and relevant concepts in strategy. It is therefore important for research to develop a sound understanding of the antecedents of strategic foresight.

Although much management research has theorized about antecedents of strategic foresight at the individual level (Dearborn and Simon 1958, Gavetti and Levinthal 2000, Helfat and Peteraf 2015) and at the organizational level (Gavetti et al. 2007, Simon 1947/1997, Weick and Roberts 1993), the empirical research on this topic has only recently started to catch up. In fact, the first empirical evidence on how strategic foresight depends on two basic individual and organizational characteristics—mental representation accuracy and organizational structure—has only recently been provided by, respectively, Gary and Wood (2011) and Csaszar (2012). At the individual level, Gary and Wood (2011) use a business simulation to show that individuals with more accurate mental representations have better strategic foresight. At the organizational level, Csaszar (2012) uses the mutual fund setting to show how organizational structure affects the type of foresight errors made by firms. In particular, mutual funds with a more decentralized structure make fewer errors of omission whereas mutual funds with a more centralized structure make fewer errors of commission (i.e., decentralized funds overlook few good investment opportunities and centralized funds make

[1] New Oxford American Dictionary, third edition.

few investments that turn out to be bad).

1.2 Open questions

The empirical papers by Gary and Wood (2011) and Csaszar (2012) are valuable additions to the strategy literature, as they provide pioneering and compelling evidence that individual- and organization-level characteristics affect strategic foresight. These papers raise two fundamental questions—concerning generalizability and underlying mechanisms—that need to be addressed in order to continue advancing the understanding of strategic foresight.

The first of these questions is whether the findings of Gary and Wood (2011) and Csaszar (2012) generalize to settings that are more representative of the contexts in which strategic decisions are typically made. The settings employed by both papers are convenient because they are amenable to empirical observation, yet they differ from typical strategy settings in some fundamental ways. In Gary and Wood's (2011) setting, an underlying correct model is readily available—namely, the business dynamics model underlying their simulation. In many real-world strategy settings, however, such an underlying model is not available; as a result, managers' representations can differ in ways that are more crucial than whether they do or do not capture a given part of the underlying model (e.g., different managers can view a strategic problem in terms of altogether different conceptual frameworks; Csaszar and Levinthal 2016). Csaszar's (2012) mutual fund setting may also be fundamentally different from the typical strategy setting: mutual funds' investment decisions depend to a large extent on portfolio- and market-level considerations (Elton and Gruber 1995) whereas strategy decisions depend more on firm-level considerations.

The second question—which applies only if the findings of Gary and Wood (2011) and Csaszar (2012) can be generalized—is: What mechanisms, at the individual and organizational levels, contribute to strategic foresight? In terms of individual-level mechanisms, Gary and Wood (2011) proposes that foresight depends on having an accurate mental model. But given that in realistic settings the underlying model is seldom available, the accuracy of a mental representation cannot be assessed in terms of knowledge about that model. Hence it is unclear whether a suitably defined measure of accuracy would corroborate the findings of Gary and Wood (2011) or whether a mechanism other than accuracy (e.g., the depth or breadth of a mental representation) could better explain how strategic foresight is affected by such representations. In terms of organization-level mechanisms, Csaszar (2012) only observes coarse measures of the decision process (level of consensus required to make a decision and whether a fund buys or not a given stock), hence, cannot elucidate the mechanism by which the group affects foresight. It is unclear whether, for example, group foresight depends more on aggregating individuals' predictions (e.g., combining forecasts regarding each alternative) or on aggregating individuals' representations (e.g., discussing rationales supporting each alternative to achieve a common understanding).

In sum, the recent empirical evidence on strategic foresight brings up important questions regarding its generalizability and the mechanisms leading to strategic foresight at the individual and organizational levels. In this paper we address these questions by examining a simplified but representative setting.

1.3 Research approach

Our paper tackles the aforementioned questions on generalizability and mechanisms by analyzing a novel exercise taken by 358 first-year MBA students from a major US university. This exercise is representative of evaluation tasks that require strategic foresight. The decision-making exercise asks participants to predict a typical strategic outcome (which company is more likely to be successful) given typical strategic stimuli (audiovisual presentations of the companies' strategies). This setting is representative of many strategy settings that lack an underlying correct model and in which managers must attend to a stream of unstructured, complex, and uncertain information that must serve as the basis for making a decision. Examples include a manager deciding whether to hire an interviewed applicant, a venture capitalist judging whether a particular startup merits further consideration, or a CEO deciding whether to implement a plan suggested by a subordinate. In all these settings, foresight is reflected in choosing the best available alternative.

Our exercise asks participants to adopt the role of an investor who must predict which of two startups is more likely to succeed. The only knowledge that participants have about each startup is a three-minute "elevator pitch" video describing each startup's main product. The videos contain information that is indicative of the startups' future success, as each video presents what the respective startup's founder believed was the most relevant information for potential investors and because the success (or failure) of each startup happened shortly after the videos were filmed.

Strategic foresight in this setting corresponds to a participant's ability to extract information from the videos and then to use that information for predicting which startup is more likely to succeed. We explore how strategic foresight depends on participants' mental representations (which we code from a list of "pros" and "cons" that each participant reports for each startup) and also on each participant's demographic characteristics, which include age, gender, and GMAT score. In addition, we analyze the effect of the decision-making structure used by comparing the predictions of individuals with those of two-person groups. Among other robustness checks, we replicated our results in a different population (450 undergraduate students) using different videos and presenting them in different order.

This setting allows us to deal with several challenges that complicate the empirical study of strategic foresight. These challenges are better understood if one imagines using archival data to study the individual and organizational antecedents of strategic foresight. Such a study would require that we collect data on many strategic decisions, and for each decision we would need information on the mental representations of participating decision makers, the decision-making structure they used, the decision they actually made, and the outcome of that decision. Such data are seldom easy to obtain, since it is difficult to establish retrospectively just what mental representations and organizational structures were used when a given decision was made. Moreover, decisions and their outcomes are typically separated by long lags; that delay reduces the number of decisions for which there is complete information and increases the chances of introducing selection and survivor biases. In contrast, in our setting mental representations, organizational structures, decisions, and outcomes are all directly observable.

Our setting also determines the scope of our research. The scope of this paper becomes clear when one considers our exercise in the context of the strategic decision-making process illustrated in Figure 1. This process comprises four stages: (i) identifying the problem, (ii) developing alternatives to address the problem, (iii) evaluating those alternatives, and (iv) implementing the selected alternative. In our exercise, the problem is already identified (i.e., judging which startup is more likely to succeed), the alternatives have already been developed (namely, the two competing startups presented to participants), and the implementation—that is, execution of the selected startup's strategy—is beyond the control of those making the selection. Thus, our exercise addresses only one stage (the evaluation stage) of this four-stage strategic decision-making process. It is outside the scope of our paper to examine, for instance, whether managers exhibit foresight regarding

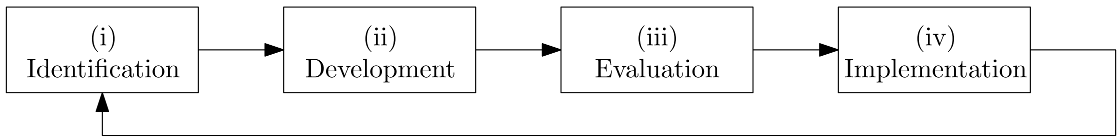

Figure 1: The strategic decision-making process (after Mintzberg et al. 1976 and Simon 1977).

the type of problems they identify (stage (i)) or regarding the alternatives they develop (stage (ii)). Likewise, we do not study how well managers can predict their own ability to implement the chosen alternative (stage (iv)). Some further scope limits of the study stem from the participants themselves (MBA students, not actual managers) and the task's extension (an in-class exercise, not a protracted decision process).

1.4 Our contribution

Our paper offers several findings regarding the aforementioned questions on generalizability and mechanisms. Regarding the question of generalizability, we show that the main results of Gary and Wood (2011) and Csaszar (2012)—that mental representation accuracy and decision-making structure affect strategic foresight—do indeed generalize to a setting that is more representative of actual strategic decisions.

Regarding the question of underlying mechanisms, we provide novel evidence on the individual- and organizational-level mechanisms that drive strategic foresight. At the individual level, we find that strategic foresight is greater in individuals whose mental representations are broad and accurately match the consensus. At the organizational level, comparing the performance of individuals to that of groups reveals that the latter exhibit greater strategic foresight than do the former. Additionally, from comparing the performance of real-life groups with that of "statistical" groups (for which decisions are computed by averaging individual assessments made prior to any group interaction), we find that the improved foresight exhibited by groups is due more to aggregating its members' predictions than to aggregating their representations. Furthermore, studying the individual and organizational antecedents of strategic foresight within the same setting allows us to illuminate the relative effectiveness of interventions at these two levels.

Apart from these specific findings, a more general contribution of our work is to provide a methodology that can advance the research on strategic foresight and on upper echelons. Our methodology allows to jointly measure the mental representations of individuals making strategic decisions and the quality of their decisions. This methodology can advance the nascent literature on strategic foresight, as it allows testing the very relationships this literature postulates. For example, Gavetti and Menon (2016:227) hypothesize that the depth and breadth of Charles Merrill's mental representation allowed him to devise Merrill Lynch's successful strategy. As illustrated in the current paper, our methodology can test on large samples hypotheses like the ones proposed by Gavetti and Menon (2016).[2] Our methodology can also advance the upper echelons literature, whose main tenet is that managers' cognitive characteristics affect organizational outcomes (Hambrick and Mason 1984). Yet, empirical constraints have dictated that, for the most part, this literature has not directly measured cognitive characteristics but has instead proxied them with demographic characteristics. Our method offers a way to directly measure the cognitive characteristics of teams making strategic decisions and to relate these characteristics to organizational outcomes.

The rest of the paper proceeds as follows. The next section presents an overview of the literatures that speak to the issue of strategic foresight. The subsequent section describes our method. We then present the results, and build on those results to discuss some broader theoretical and managerial implications of our research.

2 Theoretical background

In this section we use Brunswik's (1952) lens model to develop a framework in which to understand the individual- and organizational-level mechanisms that affect strategic foresight. This framework makes clear predictions regarding the antecedents of strategic foresight; that is, how strategic foresight depends on individual and organizational characteristics. The framework also helps explain the rationale underlying the design and analysis of the exercise discussed in the paper.

[2] Gavetti and Menon (2016:207) define strategic foresight as "the ability of a strategist to identify a superior course of action, especially one that is markedly different from the status quo." Our setting tests that same ability, as our participants are asked to identify which is the superior course of action (i.e., in which startup to invest), and for a startup to be successful it needs to offer a product that is different from the status quo (as otherwise competition would destroy profits).

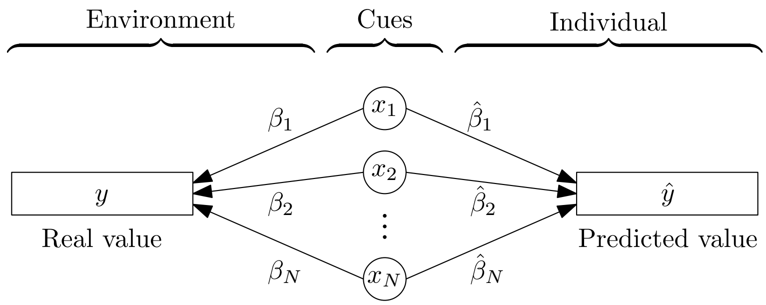

Figure 2: Brunswik's lens model.

2.1 Overview of the lens model

Brunswik (1952:19–21) proposed the lens model to account for the process by which individuals make judgments based on multiple characteristics or *cues*. Brunswik's model has influenced many empirical and theoretical studies on perception, learning, and decision making (see, e.g., Karelaia and Hogarth (2008) for a survey of 249 experimental lens model studies). The task captured by Brunswik's model is that of predicting a value when presented with some of these cues.

Brunswik's model conceptualizes the individual and the environment in a symmetric fashion: the individual is represented by a function that connects the cues to a *predicted* value, and the environment is represented by a function that connects cues to a *real* value. For instance, a manager may predict this year's profits based on the cues of last year's sales (x_1) and the economy's growth (x_2) according to $\hat{y} = 0.3x_1 + 0.1x_2$. If predictions perfectly matched reality, the manager would be able to make perfect decisions.[3] In reality, however, the weights may differ or another cue—say, inflation (x_3)—could also matter; thus the actual profits might be given by $y = 0.3x_1 + 0.2x_2 - 0.1x_3$. The lens model is customarily drawn as shown in Figure 2, with the left-hand side representing the environment and the right-hand side representing the individual.

Brunswik's model offers a natural framework for studying the processes that drive foresight. In this framework, foresight corresponds to how closely predictions match reality (i.e., the absolute difference between $\hat{y}$ and y). An individual's *mental representation* of the predicted phenomena consists of (i) the set of x's to which she attends and (ii) the set of $\hat{\beta}$'s that she employs. This

[3] In terms of the project-selection model presented by Garud et al. (1997:25), this would mean that the individual's predictive validity would be perfect (and hence the ellipse in their Figure 3.1 would become a 45-degree line). Note that our usage of the word "foresight" differs from theirs. For them, "a foresight" means accepting a good project, which they contrast with "an oversight" (an omission error). In contrast, by "foresight" we mean "ability to predict" (which affects not just the probability of accepting good and bad projects but also of rejecting them).

Brunswikian conceptualization of mental representation is consistent with a common definition of mental representation, namely "a model of reality held in the mind of an individual, who can use this representation to generate predictions" (Csaszar and Levinthal 2016:2031–2).

2.2 Individual- and organizational-level implications of the lens model

At the individual level, Brunswik's model implies that foresight increases with the mental representation's accuracy. In other words: an individual's predictions ($\hat{y}$'s) improve to the extent that she attends to relevant cues (i.e., does not miss relevant x's) and also weighs cues sensibly (i.e., uses $\hat{\beta}$'s that closely resemble the real β's).

In practice, Brunswik's model implies that foresight should increase with three characteristics of mental representations: breadth, depth, and consensus. *Breadth* is the number of relevant cues taken into account. Increasing breadth makes more likely that the mental representation will include cues that affect the real value. *Depth* is the detail with which an individual considers each relevant cue. Increasing depth improves the chances that the estimated weights will be closer to the corresponding real weights.

Because the ability to make good predictions depends on the fit between the mental representation and the environment, foresight should increase as the weights in the mental representation get closer to the "real weights." However, measuring this distance is problematic, as in realistic settings (such as predicting firm success) no one knows what the real weights are. One way to escape this problem is to use the "wisdom of the crowd" logic (Galton 1907, Surowiecki 2004) and assume that averaging cue estimates across a large number of individuals provides a good approximation of the real weight of that cue. We define *consensus* as the distance between the weights estimated by one individual and the average weights estimated by a large crowd. Hence, consensus proxies for the fidelity of a mental representation. There is empirical support for accuracy increasing with breadth, depth, and consensus (see Karelaia and Hogarth 2008 for a survey of such findings; see also Davis-Stober et al. 2014 for evidence on the robustness of crowd decision making).

At the organizational level[4], Brunswik's model implies that foresight can increase through

[4] Our usage of the term "organizational level" is consistent with Simon's (1947/1997:18–19) definition of organization as "the pattern of communications and relations among a group of human beings, including the processes for making and implementing decisions" since we examine group-level communication patterns and decision-making. Our usage is also aligned with recent work on organization design which has looked at how micro-organizations (rather than macro-organizations) affect organization-level decisions (see, e.g., Csaszar 2012, Csaszar and Eggers 2013).

two different mechanisms: aggregating predictions and aggregating representations. *Aggregating predictions* occurs when the individual group members' predictions (i.e., their $\hat{y}$'s) are combined to produce a more accurate group-level prediction. *Aggregating representations* occurs when the group members exchange ideas and learn, for example, that some members have been overlooking cues (x's) and/or attaching incorrect weights ($\hat{\beta}$'s) to some of the cues. After this exchange of ideas, the group members are in a better position to make an informed and, hence, more accurate decision. Both mechanisms can improve performance, yet how they achieve that improvement is different: aggregating predictions depends on combining $\hat{y}$'s, while aggregating representations depends on combining x's and $\hat{\beta}$'s. Therefore, the two mechanisms offer a partition of the ways in which groups can improve performance.

Both mechanisms have advantages and disadvantages. The advantage of aggregating predictions is that it is simpler, as there is little information to combine (i.e., one $\hat{y}$ per member) and combining these predictions can be done with little effort (e.g., by computing an average). In contrast, aggregating representations is more involved: the group needs to deliberate about each element in their members' representations (potentially a large number of x's and $\hat{\beta}$) and come up with a negotiated belief structure (Walsh et al. 1988). The advantage of aggregating representations is that it can produce better predictions, as the deliberation can allow the group members to improve their representations (e.g., seeing new cues and better assessing their impact).

In sum, based on Brunswik's lens model we theorize that strategic foresight may depend on three individual-level antecedents (breadth, depth, and consensus) and two group-level antecedents (aggregating predictions and aggregating representations). Yet, a priori it is not clear to what extent each of these antecedents will explicate strategic foresight. It could be that mechanisms that work well in simple forecasting tasks do not generalize to the more complex problems of strategic foresight. For instance, if achieving strategic foresight is very difficult, perhaps none of the antecedents has a significant impact. Or it could be that some of the antecedents are much more relevant than others. The method we present next sheds light on such questions.

3 Methods

3.1 Participants

Data were obtained from six sections of students taking the core strategy course at a major US business school. Three sections took the course in 2014 and three in 2015. Within each cohort, section assignments are made by the MBA office with the aim of creating sections that are equivalent in all measurable aspects (in fact, we found no statistically significant differences regarding gender composition, proportion of international students, and standardized test scores).

The pool of individuals that participated in the decision-making exercise comprised 407 MBA students. From this set we removed those who reported any familiarity with the firms presented or who provided invalid or incomplete answers. Hence the final sample consisted of 358 individuals (245 men and 113 women) whose average age was 27.43 years (SD = 2.36 years). For further details, see Tables 12 and 13 (in Appendix B).

In addition to the exercise reported here, we conducted variations of it with 450 undergraduate students. The results of these other exercises are not included in this paper because the stimuli and participant type were different. However, these additional exercises served to check the robustness of our results with respect to (a) changing the startups presented for evaluation and (b) changing the order in which the video presentations are made (i.e., whether the best performing startup is shown first or second). Our results from these other exercises are consistent with the results reported here.

3.2 Task

Before settling on the current task, we explored different tasks that might allow us to measure strategic foresight in realistic business situations (among these alternatives were vignette scenarios, written business cases, and business simulations). After conducting several pilot studies, we opted to use videos created by startups to raise capital via Kickstarter.com (a website that enables firms to raise money from individuals). This type of stimuli has the advantage of being naturalistic; that is, similar to situations faced by managers in terms of content (complexity, richness of information), form (an audiovisual presentation), and purpose (picking a superior alternative).

To increase the probability of detecting individual differences in foresight, we chose two startups with diametrically opposed performance with respect to financial, technological, and commercial dimensions: one startup was able to raise more money than it requested, delivered the product on time, and had commercial success; the other startup could not raise the requested funds, did not deliver the product on time, and failed to attract customers when the product was finally available.[5] Another criterion driving the selection of these startups was that the time elapsed between when the videos were filmed and when each firm had become a total success or failure was less than one year, which increases the chances that the video presents information that is relevant to predict performance.

Before starting the exercise, participants were asked to imagine that they were investors with no liquidity constraints and who were about to assess two startups for the purpose of making an investment decision. Participants were also told that each startup would request a certain amount of money and that they would need to specify a minimum interest rate at which they would be willing to loan the money. The interest rate was described in terms of a multiplier M such that, if the startup were successful, then participants would be repaid M times the loan amount three years from the day of the exercise; otherwise, they would receive nothing. Comparing the M's that a given participant specified for each startup allows us to infer which startup the individual believes is riskier (since the riskier startup should be charged a higher interest rate). So in our exercise, exhibiting good foresight corresponds to the failed firm being charged a higher M than the successful one.

After watching each video, participants were asked to report (a) the minimum interest rate (M) at which they would invest in the startup, (b) an open-ended list of the pros and cons for the startup, and (c) some control questions.

To collect the list of pros and cons, we asked participants to write down all the elements to which they paid attention when deciding (we asked, "feel free to list as few or as many elements as you wish, as long as you considered them to decide"). Then, for each item in the list, we asked participants to assign a valence between -2 (if the item had a strong negative effect on their decision) and $+2$ (if it had a strong positive effect).

[5]The startups we selected were MindRider and SmartHerb. The videos shown to the participants are available at https://www.kickstarter.com/projects/1168534473/mindrider-a-new-mind-mapping-helmet-system and https://www.kickstarter.com/projects/mattiaslepp/smart-herb-garden-by-click-and-grow/description (last accessed on July 3, 2017).

We coded each item on these lists of pros and cons into one of ten categories; for this purpose we used the content analysis method described in Appendix D. Then, for each participant, we computed the average valence per category to derive a 10-element vector (e.g., if regarding market size, an individual mentioned the pros "large domestic market, +2" and "possible international markets, +1" then these responses were recorded as a +1.5 in the "market size" element of the vector). That vector, which is generated by coding the list of pros and cons, corresponds to the individual's mental representation of a startup—akin to the $\hat{\beta}$'s in Brunswik's model. The zeros in this vector can be interpreted as those x's in Brunswik's model that are irrelevant to the individual.

In addition to the information collected during the decision-making exercise, we have demographic information on each individual: age, gender, previous business education, and GMAT score.

3.3 Data collection procedure and manipulation

The day before the exercise, participants had to take an online practice exercise whose objective was to familiarize them with the task so they would understand their role as investors and the type of questions asked. The practice exercise gave participants information about a startup and asked them to answer the key questions of the actual exercise (i.e., regarding M and the list of pros and cons).

The actual exercise took place during the ninth session of the strategy class (fifth week of the MBA program for both cohorts) in the context of learning about strategy process. We carefully considered how to incentivize the students. All participants were told before the class and at the beginning of the exercise, that their performance would be analyzed and that they would be debriefed in the next class regarding how well they did. The prospect of having their foresight publicly scrutinized in class motivated students to take the online practice exercise and to work conscientiously during the class exercise. All students were highly engaged in the exercises; they all performed the practice exercise the day before and thoroughly followed the instructions during class. The exercise was repeatedly praised in the student evaluations of the course, further confirming participants' high motivation.

The experimental manipulation of the exercise is the decision-making structure used—that is, whether foresight is affected by working alone or in a group. To be able to isolate the effect of the decision-making structure used, the exercise involves two consecutive rounds: in Round 1, individuals work alone; in Round 2, all the sections (except for two control sections) work in groups.

In Round 1, participants watch the first startup's video and are then given seven minutes to complete a survey (regarding that firm's M, its list of pros and cons, and controls). No Internet access is allowed throughout the exercise; to facilitate enforcement of this condition, we use a paper survey rather than an online one. The same procedure is repeated for the second startup. Each video is about three minutes long.

In Round 2, participants are required to answer the same questions as in Round 1. In four of the six sections studied, participants respond to the survey in groups of two. We instructed all groups to reach a joint decision by following some basic steps. Namely, after each partner explains (in two minutes) her decision and reasoning, the partners then have three minutes to decide jointly and fill out the survey. In the two control sections, participants were asked to work alone, and given the same seven minutes to answer the survey. In these sections, participants were instructed as follows: "Now that you have experience in evaluating this type of investment, several of you may want to reconsider your investments and reasons. Think again about this investment and complete the survey again."

The exercise debriefing occurred in two stages. Immediately after the conclusion of the exercise, the instructor guided a discussion by asking what factors the students considered, how those factors were related to concepts discussed in the course, and how working in a group differed from working alone. Then, one week after the exercise, the instructor debriefed the class by analyzing aggregate statistics. At this stage, the performance of the various sections under the different conditions was discussed and linked with topics relevant to the class. For more details about the experimental procedure, see Appendix A.

3.4 Individual-level measures

Our study incorporates both individual-level and group-level measures. For the sake of clarity, here we describe only the individual-level measures and delay (until the Results section) the description of the group-level measures. The reason for this sequence is that the group-level measures are more easily understood after the individual-level measures have been analyzed. The mathematical definitions of all measures are given in Tables 15 and 16 (in Appendix C).

3.4.1 **Performance (dependent variable).** We define individual performance as a binary variable that takes the value of 1 if the successful firm is charged a lower interest rate than the failed firm. Otherwise, the *Performance* variable is set to 0.[6] This measure captures whether a participant identified the relative quality of the startups. As such, this measure fits settings like the ones described in the Introduction, in which strategists try to pick the best available alternative. Defining performance in relative terms has the additional benefit of controlling for within-individual effects (e.g., risk aversion and subjective scale differences that apply equally to both startups).

3.4.2 **Demographics.** We include four variables to account for potentially important sources of individual differences among participants: age, gender, business education, and GMAT score. We consider an individual as having a business education if his undergraduate or graduate major was in any field typically taught by business schools (e.g., management, administration, international business, marketing, operations, economics, finance). Following Gary and Wood (2011), we include each participant's GMAT score as a proxy for cognitive ability.

3.4.3 **Characteristics of the mental representation.** We use the lists of pros and cons reported by each individual to characterize their mental representations. Toward that end, we compute three measures—breadth, depth, and consensus—that characterize the mental representation of each individual with respect to each firm. We then use the average of each measure across the two evaluated firms to characterize each individual's mental representation.

Recall that in the context of Brunswik's lens model we defined breadth as the number of relevant cues observed and depth as the level of detail with which these relevant cues are observed. To operationalize these definitions, we created a list of relevant categories that reflects the contents of the course (whose main goal is to study why some companies are successful and others are not; thus, it is natural to use the course content to define relevant categories). Hence, in our operationalization, a relevant cue corresponds to a category. We created the categories using the iterative coding process described in Appendix D. This produced 10 categories. Items that do not fall into any of these categories are discarded and not taken into account to compute the measures here described. In what follows we describe the operationalization of each measure.

Breadth: Measures the number of different categories in which the reported items fall. For

[6] In the Robustness Checks section we discuss an alternative performance measure that is continuous rather than binary. Our results are qualitatively robust to that alternative specification.

example, if an individual reported two pros and one con, and the two pros were in the same category while the con was in a different category, then $Breadth = 2$. Because our coding procedure uses 10 categories, breadth can take values between 0 and 10.

Depth: Measures the level of detail with which a participant evaluated the different categories. This variable is calculated as the average number of elements considered per category. Continuing the previous example, because the individual reported three items that fell into two categories, his depth was 1.5 (= 3/2).

Consensus: Measures the extent to which an individual's mental representation is similar to the average mental representation of the other participants who undertook the exercise. The measure is defined in terms of the Manhattan distance between the individual's mental representation (i.e., the 10-element vector with the categories' valences) and the "crowd's mental representation" (i.e., the centroid, or average 10-element vector, of the mental representations of all individuals taking the exercise). To make the consensus measure grow in the right direction, we reverse the sign of this distance measure so that a greater distance to the centroid implies less consensus.

Consensus, or being close to the crowd's average opinion, can be interpreted in two ways. One interpretation is based on the wisdom-of-the-crowd argument: the average mental representation across all participants is probably a good approximation of the true relationship between cues and the startup's real value; in terms of Brunswik's model, this interpretation assumes that individuals' $\hat{\beta}$'s are randomly distributed around the real β's and so the average of multiple $\hat{\beta}$'s tends to the true β. The second interpretation is that the crowd is similar to a typical Kickstarter user; hence a high value for *Consensus* implies a better ability to predict whether a startup will succeed on Kickstarter—an outcome that is certainly relevant to predicting its overall chances of success. Under either interpretation, consensus is a measure of the accuracy of the mental representation.

We believe that using consensus to proxy for the accuracy of a mental representation is reasonable in settings, such as ours, where no underlying model is available yet participants are fairly knowledgeable about the type of project being evaluated. If these conditions do *not* apply—and so the crowd is ill equipped to evaluate projects (e.g., assessing industrial projects with respect to which participants have no relevant background)—then the distance to an expert, rather than to the crowd, would likely be a better measure of mental representation accuracy.

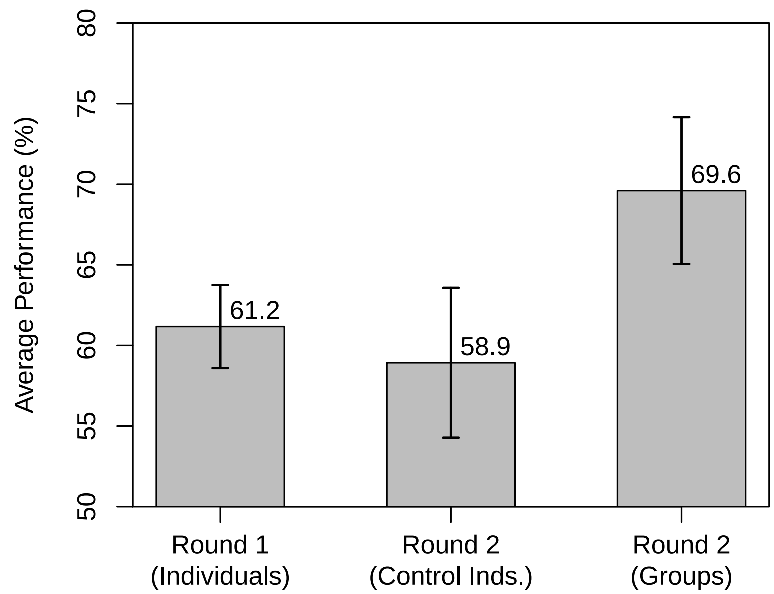

Figure 3: Overview of the results. Error bars represent ±1 standard errors.

4 Results

Before delving into the detailed regression analyses developed in this section, it is useful to get a sense of the data by looking at Figure 3. This figure shows the average performance of participants in Round 1 (in which they worked alone) and in Round 2 (in which some worked alone and others worked in groups). Two facts are apparent in the figure. First, the performance of *individuals* is about 60% irrespective of whether they are undertaking the task for the first or second time (the difference between the first and second bars in Figure 3 is not statistically significant). Second, and more importantly, the performance of *groups* is about 70%, which is substantially higher than the performance of individuals. In the rest of this section we study what lies behind these averages—that is, what determines the foresight of specific individuals and groups—and how the performance advantage of groups can be explained.

4.1 Individual-level foresight results

Table 1 reports the correlations, means, and standard deviations for all individual-level variables. The *Gender* indicator variable was coded as 1 = female and 0 = male, and the *Business education* indicator was set to 1 only for individuals with previous business education (and to 0 otherwise).

	1	2	3	4	5	6	7	8
Variable								
1. Performance								
2. Age	−0.04							
3. Gender	-0.10^{+}	-0.18^{***}						
4. GMAT	0.07	−0.03	-0.18^{***}					
5. Business education	0.03	−0.01	−0.05	−0.07				
6. Breadth	0.02	−0.02	0.07	0.04	−0.07			
7. Depth	−0.02	−0.06	0.04	−0.07	−0.02	-0.22^{***}		
8. Consensus	0.12^{*}	−0.05	0.03	-0.10^{+}	−0.04	-0.54^{***}	0.26^{***}	
Descriptive statistics								
Mean	0.61	27.43	0.32	702.47	0.46	4.45	1.68	−0.80
SD	0.49	2.36	0.47	41.78	0.50	1.32	0.50	0.16

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 1: Descriptive statistics and intercorrelations for individual-level variables in Round 1.

The significantly negative correlation (−0.22) between depth and breadth is consistent with the classical trade-off between these two variables (i.e., answering with both depth and breadth is unlikely due to time and cognitive constraints).

Since the dependent variable (performance) is binary, we use logistic regressions to estimate our models (see Tables 2 and 4). To facilitate the interpretation of effect sizes, variables with unfamiliar or disproportionate ranges are entered into the regressions as standardized scores (marked with "z-score" in the regression tables). All other variables enter the regressions in their natural units: *Age* is in years; *Gender* and *Business education* are binary (as just described); and *Breadth* and *Depth* are count variables. To control for possible section-specific differences, all the regressions in Table 2 include section fixed-effects (i.e., one dummy per section). Examples of section-specific differences could be: students being more tired in the sections that occur later in the day, slight differences in how the course is taught across sections and cohorts, and other section dynamics that could impact the students' knowledge or attentiveness.

Model 1 in Table 2 presents the effect of each individual characteristic on performance. In this model, no coefficient is significant at the 5% level. Model 2 introduces the variables we use to capture mental representations. Breadth and consensus have significantly positive effects on foresight (with respective coefficients of 0.25 and 0.53; all are significant at the 5% level or better). It is interesting that depth does not have a significant effect on foresight—simply providing more pros

	Model 1	Model 2
Demographic characteristics		
Age	−0.06	−0.05
	(0.05)	(0.05)
Gender	-0.48^{+}	-0.54^{*}
	(0.25)	(0.25)
GMAT (z-score)	0.13	0.17
	(0.12)	(0.12)
Business education	0.12	0.22
	(0.23)	(0.23)
Mental representation		
Breadth		0.25^{*}
		(0.11)
Depth		−0.19
		(0.24)
Consensus (z-score)		0.53^{***}
		(0.15)
Intercept	1.77	0.81
	(1.37)	(1.60)
Observations	358	358
Number of parameters	4	7
Nagelkerke's pseudo-R^2	0.07	0.12
Log likelihood	−229.23	−221.99

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Table 2: Regressions predicting individual foresight.

and cons per category is not helpful. This result is consistent with the research on unit weighting schemes (Einhorn and Hogarth 1975) and other "improper" decision making models (Dawes 1979), which has shown that in many settings, just getting a directional assessment of the effect of cues is enough to make accurate predictions.[7] The result is also consistent with the observation by Gavetti et al. (2005:709) that "beyond a modest level of depth, performance is not sensitive to depth."

From a practical standpoint, we should like not only to know the direction and statistical significance of the regressors but also to develop a sense of their relative relevance. Such analysis must be performed carefully, as not all regressors are measured using the same scale and because of the nonlinear nature of the logistic regression. To get an idea about effect sizes, one can look at the effect of varying a regressor on an individual whose characteristics are all at the population

[7] Such settings are sometimes described as having a "flat maximum" (von Winterfeldt and Edwards 1982).

average. Such analysis shows that increasing breadth by one unit would increase foresight by 5%, and increasing consensus by one standard deviation would increase foresight by 10%. Hence, both significant regressors in Model 2 have a considerable effect on foresight.

In sum, our analysis reveals that individual-level strategic foresight increases with breadth and consensus. In the Discussion section we elaborate on how it may be possible to manipulate these antecedents to achieve greater strategic foresight.

4.2 Group-level foresight variables

Before presenting the results on group foresight, we shall introduce the measures used in our analysis. To analyze the foresight of groups, we created variables that capture (i) the characteristics of the individuals in each group and (ii) the characteristics of the list of pros and cons reported by each group in Round 2. The first set of variables is akin to the demographic variables used in the individual regressions, and the second set of variables is akin to the mental representation variables used in the individual regressions.

4.2.1 Variables characterizing the group's members. To account for the demographics of each group, we calculated each group's average age, gender, GMAT, and business education. These variables parallel the demographic variables used in the individual regressions.[8] We also created variables to account for group heterogeneity (e.g., age difference, gender difference) and group maximum for each demographic characteristic (e.g., age of the group's oldest member, the highest GMAT score). To streamline our presentation and because no additional insights emerged from incorporating these variables, the regressions described here use only the average group characteristics.

In addition to the demographic variables, we include a measure of the foresight of the group's members. We compute this measure by plugging into the best-fitting model of individual-level foresight (Model 2 in Table 2) the characteristics of each group member, which yields a variable that captures each individual's predicted foresight. Because this variable is derived from a logistic regression, its value can range from 0 to 1 and corresponds to the probability that an individual with given characteristics will make the right decision. We use the average of the predicted foresight

[8]In the Robustness Checks section we discuss other ways of characterizing group demographics (e.g., using dummy variables to distinguish between groups in which zero, one, and two members have a business education). Our results are qualitatively robust to these alternative specifications.

	1	2	3	4	5	6	7	8	9
Variable									
1. Performance									
2. Average age	0.03								
3. Average gender	−0.13	-0.26^{**}							
4. Average GMAT	−0.01	0.04	−0.14						
5. Average business education	−0.09	−0.03	-0.18^{+}	−0.10					
6. Average predicted foresight	0.29^{**}	−0.12	−0.14	0.04	0.01				
7. Group breadth	0.03	0.07	−0.12	−0.08	0.12	0.12			
8. Group depth	0.03	−0.09	−0.12	0.06	−0.03	0.08	−0.14		
9. Group consensus	0.06	-0.23^{*}	0.10	0.11	-0.19^{+}	0.04	-0.67^{***}	0.29^{**}	
Descriptive statistics									
Mean	0.70	27.56	0.32	699.92	0.47	0.64	4.41	1.58	−0.85
SD	0.46	1.82	0.26	28.98	0.33	0.13	1.21	0.38	0.17

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 3: Descriptive statistics and intercorrelations for group-level variables.

exhibited by group members as an overall measure of their quality; we call this measure *Average predicted foresight.*

4.2.2 Variables characterizing the group's representation. We used the list of pros and cons reported by each group in Round 2 to create measures analogous to those computed for the mental representation of each participant in Round 1 (i.e., breadth, depth, and consensus). Because these group-level measures are computed on the basis of the pros and cons *jointly* reported by the group members, we refer to these three measures collectively as the "groups' representation" (we refrain from calling these measures the "groups' mental representation," as this would unnecessarily ascribe a "mind" to groups).

4.3 Group-level foresight results

Table 3 gives the correlations, means, and standard deviations for our group-level variables.

Table 4 presents three models that introduce variables much as in Table 2. Table 4 contains one additional model to demonstrate the effect of the *Average predicted foresight* variable described above. The dependent variable is the same as in the individual-level regressions: *Performance* = 1 if the successful firm is charged a lower interest rate than the failed firm (and equals 0 otherwise). As in Table 2, all the models in Table 4 include section fixed-effects.

The main takeaway from Table 4 is that, in contrast to our individual-level regressions, there is just a single statistically significant regressor: average predicted foresight. In other words, the

	Model 1	Model 2	Model 3
Demographic characteristics			
Average age	-0.03	0.09	0.11
	(0.13)	(0.14)	(0.14)
Average gender	-1.57^{+}	-0.67	-0.70
	(0.95)	(1.03)	(1.06)
Average GMAT (z-score)	-0.06	-0.22	-0.22
	(0.24)	(0.26)	(0.26)
Average business education	-0.93	-0.78	-0.73
	(0.70)	(0.73)	(0.74)
Average predicted foresight		9.68^{**}	9.26^{*}
		(3.61)	(3.72)
Group representation			
Group breadth			0.09
			(0.27)
Group depth			0.03
			(0.63)
Group consensus (z-score)			0.18
			(0.36)
Intercept	2.34	-6.15	-6.82
	(3.69)	(4.87)	(5.29)
Observations	102	102	102
Number of parameters	4	5	8
Nagelkerke's pseudo-R^2	0.08	0.18	0.19
Log likelihood	-59.69	-55.55	-55.40

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Table 4: Regressions predicting group foresight.

quality of a group's predictions depends—above all—on the quality of that group's members.

It is noteworthy that *Average predicted foresight* explains a sizable portion of the variance in group foresight, as the largest increase in R^2 occurs when this variable is introduced (the value of R^2 jumps from 0.08 in Model 1 to 0.18 in Model 2). Apart from being significant and explaining a substantial amount of the variance, average predicted foresight is highly relevant: its coefficient (9.26 in Model 3) is several times larger (in absolute value) than any of the model's other coefficients.

4.4 Why do group and individual foresight differ?

The analysis of the group-level regressions showed that an antecedent of group foresight is individual foresight: groups with "better" individuals perform better. But these regressions do not answer the question of why groups have, on average (as evidenced by Figure 3), better foresight than individuals.

To shed light on this question, we now look at how the group responses (in Round 2) differ from the responses of the members of the group (in Round 1). More specifically, we compare groups against "statistical" groups, whose responses are defined by mathematically combining the responses of individuals before they become group members. For example, if in Round 1 the group members set M's of 10 and 20 for firm 1 and M's of 30 and 40 for firm 2; then we say that the "statistical group" corresponding to these members assigned M's of 15 and 35 to firms 1 and 2, respectively. In this case, the statistical group's "decision" would be to prefer firm 1 over firm 2. This method of analysis allows inferring group processes and measuring their added value.[9] Below we use this method to analyze the two main pieces of data we collect from the surveys: predictions (M's) and representations (derived from the lists of pros and cons).

The main finding of the ensuing analyses is that the group has different effects on the predictions than on the representations. While a group's prediction usually converges close to the average of its members' predictions, a group's representation is far less convergent. Moreover, groups seem to only improve the accuracy of predictions, not the accuracy of representations.

4.4.1 Effect of groups on predictions. The two-by-two matrix in Table 5 compares the decisions made by groups and statistical groups. The table's main diagonal (upper left cell to lower

[9] Contrasting real versus statistical groups has a long history in the psychological research on groups. An early use of this method is Gordon (1924); see also Lorge et al. (1958) and Hogarth (1978) for more references to early uses of this method. In previous research, statistical groups are sometimes called "statisticized" or "staticized" groups.

		Decision of Statistical Group		
		Incorrect ($M_{\text{failed}} \leq M_{\text{successful}}$)	Correct ($M_{\text{failed}} > M_{\text{successful}}$)	Total
Decision of Group	Incorrect ($M_{\text{failed}} \leq M_{\text{successful}}$)	21	10	31
	Correct ($M_{\text{failed}} > M_{\text{successful}}$)	7	64	71
	Total	28	74	102

Table 5: Matrix comparing the decisions made by real and statistical groups.

right cell) shows the cases in which real and statistical groups made identical decisions: in 21 cases, both group types chose wrongly; in 64 cases, both chose correctly. The other diagonal (upper right cell to lower left cell) shows the cases in which the decisions by real and statistical groups differed: in 10 cases, the statistical groups chose correctly while the real groups did not; in 7 cases, the converse held.

The fact that a large majority of the cases in Table 5 are located on the main diagonal ($\frac{21+64}{102} = 83\%$) means that groups and statistical groups acted similarly. Because statistical groups average their member's M's, it follows that a first approximation of what the real groups do is to average their members' M's. In other words, the groups' decisions are predicted reasonably well by an "averaging" social decision scheme (Davis 1973).[10] Multiple processes could produce a group M that is close to an average; a plausible one is that group members may negotiate the value of M and end up compromising midway given that the members have similar power (both are classmates) and similar supporting evidence (both just saw the video).

Further evidence that the group behaves as if it was averaging their members M's is that in 88% of the cases, the groups pick an M that is "bracketed" (Larrick and Soll 2006) by the M's the group members had reported (e.g., imagine that the two members of a group report M's of 10 and 20; a bracketed group M would be any number in the 10–20 range, inclusive).

Table 6 more directly assesses the extent to which the groups behave as if they were averaging their members' M's. This table compares the M's of the 102 groups vis-à-vis the M's of the corresponding statistical groups. The main observation from this table is that the M's reported by both groups are not statistically significantly different. This provides further support to the idea

[10]Note that unlike a "decision rule," a social decision scheme describes the outcome of a group decision but not the actual process leading to the outcome (Kerr and Tindale 2011:17). In other words, the social decision scheme conveys that the group behaves "as if" it was averaging M's, but does not imply that the group actually calculates an average.

	Group (average M)	Statistical group (average M)	t-test (p-value)
Firm 1	30.91	28.21	0.43
Firm 2	18.39	15.67	0.31

Table 6: Comparison of M's reported by groups and statistical groups. The last column presents the p-value of the two-tailed t-test to determine whether the two sets of M's are different.

that the behavior of the groups can be well approximated by an average.

A conjecture from Table 6 is that groups appear to behave more conservatively that their members (i.e., groups pick higher M's than statistical groups). Such behavior is consistent with the finding that consensus decision-making leads to more conservative decisions as group size increases (Sah and Stiglitz 1986, Csaszar 2012). One can precisely test this conjecture by comparing each M provided by an individual with the M provided by the group to which this individual belongs (this produces 408 comparisons, as there are 102 groups, two individuals per group, and two firms). This can be tested with a non-parametric test such as the Wilcoxon test (one-sided, paired) or a permutation test. Both the Wilcoxon and the permutation tests provide strong support for groups being more conservative than its members (with p-values of 0.027 and 0.001 for each test, respectively).[11]

In sum, the analysis of how the group transforms the individuals' M's show that group behavior tends to converge toward an M that is close (but slightly higher) than the average of the M's of the members (the best fit for the data is: group's M = average of its members' M plus 9%). Such behavior is consistent with the idea that the members have similar power and supporting evidence and that groups are more conservative than individuals.

4.4.2 Effect of groups on the representations. Table 7 shows that groups tend to assign valences that are more conservative (i.e., more negative) than the valances assigned by the statistical group. This pattern of results is consistent with the observations regarding M (discussed in the context of Table 6). As before, a non-parametric test comparing all individuals' valences to the corresponding valence produced by groups offers support to this observation (p-values below 0.01 for both the Wilcoxon and permutation tests).

[11] A t-test cannot be used in this case, as including both firms breaks the normality assumption of this test. The new test has more statistical power than the t-tests in Table 6, as it pools the data for both firms and does not average members' M's.

	Group (average valence)	Statistical group (average valence)	t-test (p-value)
Firm 1	-0.12	-0.07	0.05
Firm 2	0.20	0.23	0.19

Table 7: Comparison of valences reported by groups and statistical groups. The last column presents the p-value of the one-sided t-test to determine whether the statistical group reports lower valences than the group. Categories without a valence are assumed to be zeros. The N of each t-test is 1020 ($= 102$ groups $\times$ 10 categories).

	Group does not see category	Group sees category
Both members see category	16%	84%
Only one member sees category	47%	53%
No member sees category	91%	9%

Table 8: Probability that a category is considered by the group as a function of how many of the group members paid attention to that category.

Analogously to the bracketing analysis of the M's, we performed a bracketing analysis of the valences. Out of the 829 cases in which the group and the two members report a valence for the same category, the group valence is bracketed 58% of the time. Hence, in 42% of these cases, the group valences diverge from what the individuals reported. This contrasts with the high bracketing of the M's (88%), and suggests that group deliberation produces group representations that differ substantially from the members' mental representations.

To get a more detailed sense of how groups' and individuals' representations differ, Table 8 looks at the probability that the group includes a category as a function of how many of its members had included that category in their representation. An overall observation from this table is that the group differs substantially with respect to the categories observed by the individuals. The first row shows that when both members see a category, there is a 16% chance that the group does not include it. Conversely, the last row shows that when no member sees a category, the group still has a 9% chance of including it. And the middle row shows that when only one of the group members pays attention to a category, there is roughly a 50% chance that that category appears in the group's representation. Taken together, tables 7 and 8 show that the group performs some substantive editing of the valences and categories observed by its members.

The analysis of Table 8 is reminiscent of analyses in the hidden profile literature (Stasser and Titus 1985, 2003), whose principal finding is that groups oversample shared information (i.e., groups

	Group	Statistical group	t-test (p-value)
Breadth	4.41	4.33	0.62
Depth	1.58	1.73	0.01

Table 9: Comparison of groups and statistical groups with respect to the breadth and depth of their representations. The last column presents the p-value of the two-tailed t-test to determine whether the two means are significantly different.

pay more attention to information that members hold in common before discussion rather than to information that only some of the members know). At first sight, our results are consistent with that finding, as we find that shared categories are more likely to be included than unshared ones (84% versus 53%, according to Table 8). But this similarity is only superficial, as there is a crucial difference: in the hidden profile literature, the experimenter endows group members with different sets of *veridical* cues about a problem (e.g., members know different real characteristics of job market candidates). Thus, the best solution is achieved if the group can pool all the cues to form a fuller picture of the problem. In contrast, in our setting we do not endow members with veridical cues, but the members come up with the cues they believe are relevant (i.e., there is no assurance that every pro and con a member raises is true). Hence, the fact that in our experiment shared categories are more likely to be included cannot be attributed to "oversampling" shared information. A simpler explanation is that cues that multiple members see are more likely to be relevant.

Table 9 compares groups and statistical groups with respect to their depth and breadth. The only statistically significant difference is with respect to depth: groups have less depth than the average depth of their members (p-value of 0.01). The decrease in depth implies that groups simplify the representation of their members. This could happen because talking about each others' list of pros and cons may prompt individuals to prune out items that are too similar, poorly stated, or have weak support. The 7-minute limit of the group deliberation may also drive members to prioritize what items they discuss. This decrease in depth does not necessarily imply a group process loss (Steiner 1972; see also Kerr and Tindale 2004:625–627 for a survey), as the individual and group regressions have indicated that depth does not significantly affect performance.

Finally, to try to assess whether the group's valences are more accurate than the valences assigned by the group members, Table 10 compares groups and individuals with respect to their

	Group (distance)	Statistical group (distance)	t-test (p-value)
Firm 1	0.67	0.65	0.88
Firm 2	0.83	0.83	0.58

Table 10: Comparison of groups and statistical groups with respect to their distance to the crowd's centroid. The last column presents the p-value of the one-sided t-test to determine whether the groups' representation are closer to the centroid than the individuals' representation. The N of the t-tests is 204 (= 102 groups $\times$ 2 individuals per group).

distance to the crowd's centroid (akin to how the *consensus* measure is defined). Interestingly, groups' and individuals' representations do not differ in their accuracy (i.e., none of the p-values in Table 10 are significant). This result is consistent with the fact that in the group-level regressions (Table 4), the consensus variable was also not significant.

4.4.3 Summary of the effect of groups. To conclude on how groups affect foresight, we contrast the effect of the group on predictions and representations. One similarity on how predictions and representations are affected is that both become more conservative (i.e., the group produces higher M's and lower valences than its members). This effect is consistent with the research on consensus decision making (Sah and Stiglitz 1986, Csaszar 2012). But with the exception of this similarity, groups affect predictions and representations differently. While a group's prediction usually converges close to the average of its members' predictions (only 12% of the group predictions are not bracketed), a group's representation is far less convergent. Group representations often remove categories members had considered; add categories members had neglected; and when keeping a category both members considered, oftentimes pick a valence outside its members' range (42% are non-bracketed).

In terms of performance, group's predictions are better than individuals' predictions (as shown in Figure 3). This happens because the decision scheme used by groups—averaging predictions—is an effective way to improve accuracy (Hogarth 1978, Larrick and Soll 2006, Hastie and Kameda 2005, Davis-Stober et al. 2014). In contrast, group's representations do not seem to be better than individuals' representations (as explained in the context of Table 10).

An explanation for why groups improve predictions but not representations has to do with the cognitive cost of aggregating both types of information. Aggregating two predictions is cognitively inexpensive: it involves combining two numbers and there is a commonly agreed-upon procedure for doing so (averaging). In contrast, aggregating two representations is cognitively much costlier, as it requires communicating and processing abundant and messy information (two lists of pros and cons) and there is no quick and commonplace process for doing so other than debating the merits of each other's items. This lack of a simple and effective aggregation process might be the reason why groups' representations are not improved. In sum, in our setting, the evidence points in the direction of groups deriving their improved foresight from aggregating predictions, not representations.

4.5 Robustness checks

We performed a number of additional analyses to check the robustness of our results. Overall, these analyses confirmed the qualitative robustness of all the results described so far. Our robustness checks included: (i) using a continuous (rather than binary) performance measure as the dependent variable; (ii) using finer-grained measures of group composition (dummy variables rather than averages); (iii) checking the models' goodness of fit; and (iv) estimating all regressions using a probit (rather than a logit) model.

For our first robustness check, we devised a continuous dependent variable that measures the relative appeal of the two firms. This measure was defined as the difference in interest rate that participants charged firm 1 and firm 2 *divided by* the sum of those rates: $\frac{M_1-M_2}{M_1+M_2}$). The denominator of this measure adjusts for the overall size of the rates (e.g., $M_1 = 10$ and $M_2 = 20$ together imply a greater preference for firm 1 than do $M_1 = 80$ and $M_2 = 90$). To test for the effect of using this dependent variable, we ran models equivalent to the models whose results are reported in Tables 2 and 4 while using the continuous measure as our dependent variable; these models were estimated using ordinary least-squares regression, rather than logistic regression, because the continuous dependent variable is not constrained to fall in the 0–1 range. All results were qualitatively consistent with those already reported.

In our main analyses we preferred using the binary dependent variable because the continuous one makes the results more dependent on participants who assess firms in more disparate ways—regardless of whether or not those assessments are justified. For example, if firm 1 is slightly better than firm 2 then someone assigning interest rates of (respectively) 1 and 100 would affect the regression more than someone assigning the likely more realistic rates of 20 and 25. In our setting it cannot be determined when the assigned rates become unrealistically disparate, so the more conservative approach is to use a binary dependent variable.

As a second robustness check, we tried different ways of characterizing the composition of groups. In particular: instead of describing groups in terms of their average age and business education, we created dummy variables to differentiate between groups that had zero, one, or two members with that given characteristic. None of these new variables was statistically significant.

We also checked the goodness of fit of our models by performing the likelihood ratio test and the Hosmer–Lemeshow test. According to the likelihood ratio test, the full models (i.e., Model 2 in Table 2 and Model 3 in Table 4) fit significantly better than do the corresponding null models containing only a constant term ($p < 0.0001$). Similarly, the Hosmer–Lemeshow test gives no indication of poor fit for the full models. (This test depends on the number of quantiles used to group the data; for extra assurance, we ran the test using not only the customary value of 10 quantiles but all integer values from 5 to 15.) As a final robustness check we re-ran all our models using a probit rather than a logit specification. Again, all results remained qualitatively robust.

As in all empirical studies, there is a potentially large number of aspects that are not statistically controlled for. However, two characteristics of our experimental setting reduce such concerns. First, as mentioned above, asking participants to make relative assessments controls for within-individual effects. Second, the possibility that participants may prefer one startup over the other not because of foresight but because of a bias (e.g., gender preferences) seems unlikely. In particular, it is unlikely that such bias would be correlated with our independent variables in the same way as foresight is (e.g., there is no reason for why participants preferring male founders would also exhibit high consensus and breadth).

5 Discussion

Because the consequences of strategic decisions lie in the future, strategy unavoidably depends on foresight; it follows that achieving a better understanding of foresight affects both the theory and practice of strategy. In terms of theory, such understanding contributes to the micro foundations of the strategy theories that hinge on strategic foresight. In terms of practice, such understanding helps to identify the managerial levers that can be used to affect strategic foresight and, by extension, firm performance. We expand on these contributions in what follows.

5.1 Theoretical contributions

This paper began by recalling some important questions that pertain to the generalizability of—and the mechanisms proposed in—the emerging literature on strategic foresight (Gary and Wood 2011, Csaszar 2012). We now re-examine these questions to demonstrate how our research sheds light on them and yields further contributions.

With regard to generalizability, we show that the main results of Gary and Wood (2011) and Csaszar (2012) generalize to settings that are more representative of strategic decision making. The main result of Gary and Wood (2011)—that increasing mental representation accuracy increases strategic foresight—is corroborated by our finding that "consensus," which serves as a proxy for the accuracy of mental representations, has a strong positive effect in individual-level models (see Table 2). Additionally, the main result of Csaszar (2012)—that organizational structure affects strategic foresight—is corroborated by our finding (illustrated in Figure 3) that foresight is improved by using even the simplest possible structure: a group of two.

As for underlying mechanisms, our first question was: In a setting where the underlying model is not available, what drives individual foresight? We show that foresight in such cases is explained by characteristics of the mental representation (breadth and consensus). Our second question regarding this topic was: If groups have better foresight than individuals, then what mechanism explains that difference? We show that, in our setting, the superiority of group performance is explained by aggregating predictions rather than by aggregating representations.

Our results suggest that firms can influence their own strategic foresight. More specifically, a firm can improve its foresight by employing managers whose mental representations are broad and accurate and by leveraging the ability of groups to make better predictions. Because the effect of the mechanism underlying these better predictions—error cancellation—increases with group size, using larger groups (perhaps even crowds) to make strategic decisions may lead to still more improvements in foresight.

5.2 Managerial implications

The work reported here suggests that some simple interventions can improve strategic foresight considerably. For instance, among the most relevant interventions are increasing breadth and using groups. Managers can improve their breadth by taking strategy content courses and they can learn to exploit group decision-making by taking strategy process courses. The low cost of these interventions and their potentially high payoffs argue for the significant effect that strategy education can have on firm performance. Firms could also intervene on these dimensions by hiring managers that have more career variety (Crossland et al. 2014) and by using more heterogeneous top management teams (Hambrick et al. 1996).

The fact that in our setting groups seem to derive their performance advantage from aggregating predictions (not representations) suggests that managers should learn to recognize the situations where deliberation does not add value to either (i) focus on aggregating predictions or (ii) devise ways to improve group deliberation. A simple intervention that has been shown to improve group deliberation is to frame the discussion as a problem-solving task rather than as a judgment task for which there is no right answer (Stasser and Stewart 1992).

Another managerial implication of our work concerns the *non*significance of the GMAT variable. This lack of significance might mean that, among the type of MBA students we study, a GMAT score (which is usually interpreted as an overall measure of cognitive ability) makes little difference in strategic foresight. Hence strategy consulting firms, as well as other firms recruiting for strategy positions, would do well to focus their selection processes on individual characteristics that—unlike GMAT—are actually predictive of strategic foresight.

5.3 Further research

Future work could use the research method developed here to address some important questions. For instance: How do different teaching methods (e.g., case studies versus lectures) affect strategic foresight? We remark that the case study method has been the primary mode of strategy instruction since the field's early days (Copeland 1958:27), yet no research has examined whether this teaching method does in fact lead to better strategy decisions. Using a similar approach to the one that we have employed, one could study this question by assessing the before–after change in foresight across groups of similar students taught using different teaching methods. Other questions that could be addressed using a similar approach include: How does learning about different strategy frameworks (e.g., Porter's "five forces," the VRIN framework)—or taking different strategy courses—affect foresight? Likewise, one could ask what is the effect of getting an MBA degree on strategic foresight.

It is both noteworthy and rather puzzling that, even though a primary goal in the field of strategy is to improve firm performance, hardly any research has attempted to measure rigorously the field's actual contribution to that performance. Clearly, further work along the lines explored here is warranted.

Apart from studying the effect of teaching methods, our method could also be used to explore the effect of experience on strategic foresight. For example, using a repeated game design one could look at how the mental representations of strategists change with experience. Here, interesting areas to explore include better understanding how the mental representations of novices and experts differ (Shanteau 1992) and the extent to which strategists learn from mistakes (Garud et al. 1997:29). Finally, future research could test the generalizability of our results to situations that are even more realistic (e.g., more extended cases in terms of information available, group size, and time to deliberate).

5.4 Conclusion

Research on strategic foresight is in a nascent stage, so questions abound at both the individual and organizational levels of analysis. The strategy field now has only fragmentary answers to the questions of what makes a great strategist and how groups affect the likelihood of making good strategy decisions. Organizations are constantly faced with choices about *who* should make strategic decisions (as when selecting a CEO) and *how* such decisions should be made (as when determining how best to organize a strategy meeting or consulting project). However, there is seldom a sound theoretical grounding for these choices.

The aim of our work is to contribute toward the establishment of such theoretical grounding. Toward that end we have developed a method to measure strategic foresight and to disentangle the effects of the group and the individual on strategic foresight. Our paper's overarching goal has been to achieve a better understanding of the processes through which foresight emerges and thereby to improve the quality of strategic decisions.

References

Ahuja, G., R. W. Coff, P. M. Lee. 2005. Managerial foresight and attempted rent appropriation: Insider trading on knowledge of imminent breakthroughs. *Strategic Management Journal* **26**(9) 791–808.

Barney, J. B. 1986. Strategic factor markets: Expectations, luck, and business strategy. *Management Science* **32**(10) 1230–1241.

Barr, P. S., J. L. Stimpert, A. S. Huff. 1992. Cognitive change, strategic action, and organizational renewal. *Strategic Management Journal* **13** 15–36. Special Issue.

Brandenburger, A. M., H. Stuart. 1996. Value-based business strategy. *Journal of Economics & Management Strategy* **5**(1) 5–24.

Brunswik, E. 1952. *The Conceptual Framework of Psychology*. University of Chicago Press, Chicago, IL.

Copeland, M. T. 1958. *And Mark an Era: The Story of the Harvard Business School*. Little, Brown and Company, Boston, MA.

Crossland, C., J. Zyung, N. J. Hiller, D. C. Hambrick. 2014. CEO career variety: Effects on firm-level strategic and social novelty. *Academy of Management Journal* **57**(3) 652–674.

Csaszar, F. A. 2012. Organizational structure as a determinant of performance: Evidence from mutual funds. *Strategic Management Journal* **33**(6) 611–632.

Csaszar, F. A., J. P. Eggers. 2013. Organizational decision making: An information aggregation view. *Management Science* **59**(10) 2257–2277.

Csaszar, F. A., D. A. Levinthal. 2016. Mental representation and the discovery of new strategies. *Strategic Management Journal* **37**(10) 2031–2049.

Davis, J. H. 1973. Group decision and social interaction: Theory of social decision schemes. *Psychological Review* **80**(2) 97–125.

Davis-Stober, C. P., D. V. Budescu, J. Dana, S. B. Broomell. 2014. When is a crowd wise? *Decision* **1**(2) 79–101.

Dawes, R. M. 1979. The robust beauty of improper linear-models in decision-making. *American Psychologist* **34**(7) 571–582.

Dearborn, D. C., H. A. Simon. 1958. Selective perception: A note on the departmental identifications of executives. *Sociometry* **21**(2) 140–144.

Duriau, V. J., R. K. Reger, M. D. Pfarrer. 2007. A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods* **10**(1) 5–34.

Einhorn, H. J., R. M. Hogarth. 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* **13**(2) 171–192.

Elton, E. J., M. J. Gruber. 1995. *Modern Portfolio Theory and Investment Analysis*. Wiley, New York.

Galton, F. 1907. Vox populi. *Nature* **75**(1949) 450–451.

Garud, R., P. R. Nayyar, Z. Shapira. 1997. Technological choices and the inevitability of errors. R. Garud, P. R. Nayyar, Z. Shapira, eds., *Technological Innovation: Oversights and Foresights*, chap. 3. Cambridge University Press, New York, 20–40.

Gary, M. S., R. E. Wood. 2011. Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal* **32**(6) 569–594.

Gavetti, G., D. A. Levinthal. 2000. Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly* **45**(1) 113–137.

Gavetti, G., D. A. Levinthal, W. Ocasio. 2007. Neo-Carnegie: The Carnegie School's past, present, and reconstructing for the future. *Organization Science* **18**(3) 523–536.

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

Gavetti, G., A. Menon. 2016. Evolution cum agency: Toward a model of strategic foresight. *Strategy Science* **1**(3) 207–233.

Gordon, K. 1924. Group judgments in the field of lifted weights. *Journal of Experimental Psychology* **7**(5) 398–400.

Hambrick, D. C., T. S. Cho, M. J. Chen. 1996. The influence of top management team heterogeneity on firms' competitive moves. *Administrative Science Quarterly* **41**(4) 659–684.

Hambrick, D. C., P. A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* **9**(2) 193–206.

Hastie, R., T. Kameda. 2005. The robust beauty of majority rules in group decisions. *Psychological Review* **112**(2) 494–508.

Helfat, C. E., M. A. Peteraf. 2015. Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal* **36**(6) 831–850.

Hogarth, R. M. 1978. A note on aggregating opinions. *Organizational Behavior and Human Performance* **21**(1) 40–46.

Karelaia, N., R. M. Hogarth. 2008. Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin* **134**(3) 404–426.

Kerr, N. L., R. S. Tindale. 2004. Group performance and decision making. *Annual Review of Psychology* **55**(1) 623–655.

Kerr, N. L., R. S. Tindale. 2011. Group-based forecasting? A social psychological analysis. *International Journal of Forecasting* **27**(1) 14–40.

Larrick, R. P., J. B. Soll. 2006. Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science* **52**(1) 111–127.

Lorge, I., D. Fox, J. Davitz, M. Brenner. 1958. A survey of studies contrasting the quality of group performance and individual performance, 1920–1957. *Psychological Bulletin* **55**(6) 337–372.

Mintzberg, H., D. Raisinghani, A. Théorêt. 1976. Structure of unstructured decision-processes. *Administrative Science Quarterly* **21**(2) 246–275.

Porter, M. E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press, New York.

Sah, R. K., J. E. Stiglitz. 1986. The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review* **76**(4) 716–727.

Shanteau, J. 1992. How much information does an expert use? Is it relevant? *Acta Psychologica* **81**(1) 75–86.

Simon, H. A. 1947/1997. *Administrative Behavior*. 4th ed. The Free Press, New York.

Simon, H. A. 1977. *The New Science of Management Decision*. Revised ed. Prentice-Hall, Englewood Cliffs, NJ.

Stasser, G., D. Stewart. 1992. Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology* **63**(3) 426–434.

Stasser, G., W. Titus. 1985. Pooling of unshared information in group decision-making: Biased information sampling during discussion. *Journal of Personality and Social Psychology* **48**(6) 1467–1478.

Stasser, G., W. Titus. 2003. Hidden profiles: A brief history. *Psychological Inquiry* **14**(3–4) 304–313.

Steiner, I. D. 1972. *Group Process and Productivity*. Academic Press, New York.

Surowiecki, J. 2004. *The Wisdom of Crowds*. Doubleday, New York.

von Winterfeldt, D., W. Edwards. 1982. Costs and payoffs in perceptual research. *Psychological Bulletin* **91**(3) 609–622.

Walsh, J. P., C. M. Henderson, J. Deighton. 1988. Negotiated belief structures and decision performance: An empirical-investigation. *Organizational Behavior and Human Decision Processes* **42**(2) 194–216.

Weber, R. P. 1990. *Basic Content Analysis*. 2nd ed. SAGE Publications, Newbury Park, CA.

Weick, K. E., K. H. Roberts. 1993. Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly* **38**(3) 357–381.

A Appendix: Experimental procedure

Table 11 summarizes the sequence of events that took place during the study for each of the two conditions. To start, precise instructions were read and a timer was projected in front of the class. Each of the events was carefully timed and all participants complied with the given instructions. The procedure for the two conditions is the same with the difference being whether during the second round participants worked again individually or in groups.

Time	**Individual condition**	**Group condition**
Round 1		
$1'$	- Instructions: “Follow questions precisely, write clearly, and do not talk to others.”	- Same as in individual condition.
$3'$	- Distribute individual survey.	- 〃
$4'$	- Show first video.	- 〃
$7'$	- Fill out first part of survey.	- 〃
$4'$	- Show second video.	- 〃
$7'$	- Fill out second part of survey.	- 〃
$3'$	- Collect surveys.	- 〃
Round 2		
$3'$	- Distribute clean copies of individual survey.	- Distribute copies of group survey to pairs of individuals.
$2'$	- Instructions: “Now that you have experience in valuating this type of investment, several of you may want to reconsider your investments and reasons. Think again about these two investments and fill out again the surveys.”	- Instructions: “Each partner will explain to the other his/her list of pros and cons [$2'$ each], then both jointly fill out the survey [last $3'$].”
$7'$	- Fill out first part of survey.	- Fill out first part of survey.
$7'$	- Fill out second part of survey.	- Fill out second part of survey.
$3'$	- Collect surveys.	- Collect surveys.

Table 11: Sequence of steps followed by participants.

B Appendix: Individuals and groups in the sample

Table 12 provides details about the individuals omitted from the analysis, for various reasons, as indicated in the main text. Table 13 does likewise for the groups that undertook our experimental exercise.

	Section						Total
	1	2	3	4	5	6	
Potential individuals	73	70	73	63	65	63	407
(−) Knew case	2	2	4	1	8	1	18
(−) Did not understand	0	0	0	0	0	1	1
(−) Incomplete	8	4	6	6	2	4	30
(=) Final individuals	63	64	63	56	55	57	358

Table 12: Individual participants in the study.

	Section				Total
	1	2	5	6	
Potential groups	36	35	31	30	132
(−) Contained eliminated individual(s)	9	7	8	6	30
(−) Did not understand	0	0	0	0	0
(−) Incomplete	0	0	0	0	0
(=) Final groups	27	28	23	24	102

Table 13: Groups in the study.

C Appendix: Description of variables

Table 14 describes the variables based on the survey or the demographic data. Table 15 and Table 16 provide mathematical definitions of (respectively) the individual- and group-level measures used in the analyses.

Variable Name	**Description**
Multiplier (M)	Multiplier assigned by individual i (or group g) to firm f in Round r. In Tables 15 and 16 this is referred as $M[\text{ind}=i, \text{firm}=f, \text{round}=r]$ and $M[\text{grp}=g, \text{firm}=f, \text{round}=r]$, respectively.
Scores per category	List of scores mentioned by individual i (or as group g) regarding firm f on category c in Round 1. In tables 15 and 16 this is referred as $\text{Scores}[\text{ind}=i, \text{firm}=f, \text{cat}=c]$ (or $\text{Scores}[\text{grp}=g, \text{firm}=f, \text{cat}=c]$). The number of elements in this list is denoted $\#\text{Scores}[\cdot]$.
Age	Age of individual i (in years).
Gender	1 if individual i is female; 0 otherwise.
GMAT	GMAT score of individual i. Students who took the GRE instead of the GMAT had their scores transformed via the formulas provided by the Educational Testing Service (ETS).
Business education	1 if individual i has a business-related major; 0 otherwise.

Table 14: Description of survey data and demographic data.

Variable name (and symbol)	**Description**	**Mathematical definition**
Individual performance IndPerf[i]	1 if individual i set a lower multiplier (i.e., charged a lower interest rate) to firm 2 than to firm 1; 0 otherwise.	$\mathbb{1}\left[M[\text{ind}{=}i, \text{firm}{=}2, \text{round}{=}1]{<}M[\text{ind}{=}i, \text{firm}{=}1, \text{round}{=}1]\right]$
Average score per category AvgScore[i, f, c]	Average scores that individual i gave to category c for firm f in Round 1.	$\dfrac{\sum \text{Scores}[\text{ind}{=}i, \text{firm}{=}f, \text{cat}{=}c, \text{round}{=}1]}{\#\text{Scores}[\text{ind}{=}i, \text{firm}{=}f, \text{cat}{=}c, \text{round}{=}1]}$
Breadth Breadth[i]	Average number of categories that individual i considered when making the investment decision for each firms in Round 1.	$\dfrac{1}{2}\sum_{f=1}^{2}\sum_{c=1}^{10}\mathbb{1}\left[\#\text{Scores}[\text{ind}{=}i, \text{firm}{=}f, \text{cat}{=}c, \text{round}{=}1]{>}0\right]$
Number of items NumItems[i]	Average number of pros and cons that individual i considered when making the investment decision for each firm in Round 1.	$\dfrac{1}{2}\sum_{f=1}^{2}\sum_{c=1}^{10}\#\text{Scores}[\text{ind}{=}i, \text{firm}{=}f, \text{cat}{=}c, \text{round}{=}1]$
Depth Depth[i]	Average number of elements per category (i.e., pros and cons) considered by individual i for each firm in Round 1.	$\dfrac{\text{NumItems}[i]}{\text{Breadth}[i]}$
Crowd score CrowdScore[f, c]	Average score (across all individuals in the sample) for firm f on category c in Round 1.	$\dfrac{1}{N}\sum_{i=1}^{N}\text{Scores}[\text{ind}{=}i, \text{firm}{=}f, \text{cat}{=}c, \text{round}{=}1]$
Consensus Consensus[i]	Degree of closeness between individual i's scores and the crowd's scores. The initial minus sign is inserted for ease of interpretation.	$-\dfrac{1}{20}\sum_{f=1}^{2}\sum_{c=1}^{10}\Big\lvert\text{AvgScore}[\text{ind}{=}i, \text{firm}{=}f, \text{cat}{=}c, \text{round}{=}1] - \text{CrowdScore}[\text{firm}{=}f, \text{cat}{=}c]\Big\rvert$
Predicted foresight PredForesight[i]	Predicted foresight for individual i.	Foresight of individual i as predicted by Model 2 in Table 2

Table 15: Individual-level variables.

Variable name (and symbol)	**Description**	**Mathematical definition**
Group performance GrpPerf[g]	1 if group g set a lower multiplier (i.e., charged a lower interest rate) to firm 2 than to firm 1; 0 otherwise.	$\mathbb{1}[M[\text{grp}{=}g, \text{firm}{=}2, \text{round}{=}2]{<}M[\text{grp}{=}g, \text{firm}{=}1, \text{round}{=}2]]$
Average predicted foresight AvgPredForesight[g]	Average of the predicted foresight for the two individuals of each group. Their foresight is calculated using the coefficients of the individual-level regression.	$\dfrac{\text{PredForesight}[i_A] + \text{PredForesight}[i_B]}{2}$

Table 16: Group-level variables.

D Appendix: Coding of mental representations

Content analysis was used to codify the lists of pros and cons. The participants, who were allowed absolute freedom in phrasing their pros and cons, produced a collection of 9,251 different short sentences. We codified these short sentences using Weber's content analysis protocol (Weber 1990, Duriau et al. 2007), which comprises the following eight steps: (1) define the recording unit, (2) define the categories, (3) test coding on sample of text, (4) assess reliability, (5) revise the coding rules, (6) return to step 3 until the coders achieve sufficient reliability, (7) code all the text, and (8) assess the achieved reliability.

In our application of Weber's protocol, the recording unit is the sentence (step 1). A preliminary list of categories was produced by the authors after studying the whole collection of short sentences (step 2). Next, two independent coders who were blind to the study's objectives were familiarized with the rules of coding, with the categories, and with examples of pros and cons that belong to each category. The coders classified a sample of 50 pros and cons (step 3). After, the authors and the coders revised the coding rules twice until their application was clear (steps 4–6), which produced ten categories plus an additional residual category. The ten categories reflected topics that are relevant to strategy. Then, the two coders coded the whole text independently (step 7), achieving an inter-rater agreement of 95.5% (step 8). Finally, and following Barr et al. (1992:22), the coders discussed all the cases they had classified differently until a mutually satisfactory category was agreed. The residual category ended up with only six items—which, after discussion with the coders, were each reclassified as belonging to one of the ten main categories. Table 17 lists the final categories along with descriptive statistics for their use in Round 1. Table 18 shows examples of sentences and the main topics that were coded under each category.

Category	Firm	#Individuals	#Items	Mean	SD
1. Industry structure	1	161	262	−0.42	0.86
	2	170	293	−0.30	0.87
2. Market size	1	249	370	−0.07	1.12
	2	275	415	0.93	1.03
3. Imitability & time to market	1	143	176	0.51	0.91
	2	190	246	0.84	0.97
4. Costs	1	112	120	−0.65	0.63
	2	150	172	−0.83	0.88
5. Operations	1	119	143	−0.25	0.81
	2	135	175	0.77	0.88
6. Value to customer	1	324	822	−0.18	1.09
	2	313	825	0.49	1.05
7. Nonmarket	1	52	62	0.17	0.52
	2	62	69	0.97	0.59
8. Marketing	1	121	162	−0.42	0.71
	2	104	129	0.57	0.65
9. Business model	1	137	169	0.48	0.88
	2	108	128	0.97	0.77
10. Funding	1	138	213	−0.00	0.84
	2	124	177	0.48	0.81

Table 17: Descriptive statistics for the ten categories used to codify the lists of pros and cons. For each category and firm, the table reports (for Round 1): number of individuals using the category, number of items, mean valence, and standard deviation (SD) of the valence.

Category	*Sample sentences from the surveys* • **Main course topics covered by the category**
1. Industry structure	*'Entry barriers'*, *'Threat of new entrants'*, *'Internal rivalry'*, *'How can we be protected against competition'*, *'Availability of substitutes'* • Five forces, entry barriers, industry rivalry, substitutes, supplier power, buyer power
2. Market size	*'Growth potential'*, *'Is there any demand for this?'*, *'Who are we targeting?'*, *'Global appeal'*, *'Many different types of customers'*, *'Limited target group'* • Market size, market trends, growth potential, expansion opportunities
3. Imitability & time to market	*'Patents'*, *'Can this be copied?'*, *'Hard to replicate technology'*, *'Easy to imitate with investment'*, *'Proprietary design'*, *'Seems really cool and unique'* • Preemption, imitability, uniqueness, differentiation, intellectual property, patents
4. Costs	*'Scale economies'*, *'Cost structure'*, *'Mass production'*, *'Development costs could be very high'*, *'High initial startup cost to manufacture'* • Economies of scale, minimum efficient scale, learning curve
5. Operations	*'Manufacturing concerns'*, *'Does technology even work? How?'*, *'What materials are needed to manufacture?'*, *'Future obsolescence'*, *'Product reliability'* • Manufacturing concerns, technological risks, past products, current stage of development
6. Value to customer	*'Willingness to pay'*, *'Ease of use for costumers'*, *'The ease of exchanging information from the user'*, *'Safety features'*, *'Design and look'* • Price, product attributes, usability, usefulness/utility/usage, health concerns, safety concerns, novelty, design factors, willingness to pay, network effects
7. Nonmarket	*'Environmental impact'*, *'Is it eco-friendly?'*, *'Government interest'*, *'Safety and regulation issues'*, *'Socio-environmental aspects'*, *'Benefits to society'* • Regulation, socio-environmental concerns
8. Marketing	*'Advertising'*, *'Management clarity of communication'*, *'Trustworthiness of spokesperson'*, *'Appeal of pitch'*, *'Clearly explained technology and how it works'* • Marketing/communication/advertising, founder characteristics, team characteristics
9. Business model	*'Synergies with other products'*, *'Business model (can do razor/blade)'*, *'Ability to apply tech to other products'*, *'Potential for other uses'* • Distribution & sales channels, product synergies
10. Funding	*'Return on investment'*, *'Backed by an association'*, *'Potential other investors already engaged'*, *'Small investment compared to potential grow'* • Net present value, likelihood of success, exit opportunities

Table 18: Examples of sentences and main course topics coded under each category.