Strategy Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Theory-Driven Strategic Management Decisions

Arnaldo Camuffo; , Alfonso Gambardella; , Andrea Pignataro

To cite this article:
Arnaldo Camuffo; , Alfonso Gambardella; , Andrea Pignataro (2024) Theory-Driven Strategic Management Decisions. Strategy Science 9(4):382-396. https://doi.org/10.1287/stsc.2024.0173

https://pubsonline.informs.org/journal/stsc

STRATEGY SCIENCE
Vol. 9, No. 4, December 2024, pp. 382–396
ISSN 2333-2050 (print), ISSN 2333-2077 (online)

Theory-Driven Strategic Management Decisions

Arnaldo Camuffo,[a,b,*] **Alfonso Gambardella,**[a,b] **Andrea Pignataro**[b,c]

[a] Department of Management and Technology, Bocconi University, Milan 20136, Italy; [b] ION Management Science Lab, Bocconi and Utah University, Milan 20136, Italy; [c] ION Group, London EC4R 1BE, United Kingdom

*Corresponding author

Contact: arnaldo.camuffo@unibocconi.it, https://orcid.org/0000-0001-9039-1057 (AC); alfonso.gambardella@unibocconi.it, https://orcid.org/0000-0002-8714-5813 (AG); ap@iongroup.com, https://orcid.org/0009-0008-0244-3039 (AP)

Received: February 4, 2024
Revised: August 8, 2024; September 20, 2024
Accepted: September 22, 2024
Published Online in Articles in Advance: November 1, 2024

https://doi.org/10.1287/stsc.2024.0173

Abstract. This paper studies strategic decisions under uncertainty for which past data are not available. It provides microfoundations of the theory-based view of the firm by showing that the strategic problem starts with the selection of theories rather than choosing actions and that theories are selected through experiments. The value of experimenting with theories increases with the number of theories and with their uncertainty. Moreover, uncertainty makes theories superadditive—that is, experimenting with a more uncertain theory increases the benefits of experimenting with other more uncertain theories. The paper also shows that decision makers should experiment with more "surprising" theories because in this case experiments are more informative and enable more learning. A leading example helps to illustrate our concepts throughout the paper.

Funding: A. Camuffo and A. Gambardella acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme [Grant 101021061].

Keywords: decision problem • experiments • exploration • framing • strategy • theory • uncertainty

1. Introduction

Executives often face strategic decisions where past data are either unavailable or insufficient. These "low-frequency, high-impact" decisions, such as those involving radical innovations, new business models, governance changes, mergers and acquisitions, or executive appointments, are inherently uncertain. In these cases, the formulation of the decision problem is the cornerstone of the strategic decision-making process, as it defines the future state space, the set of alternative actions, and the performance consequences for the firm executives should focus on (Nickerson and Zenger 2004, Baer et al. 2013, Nickerson and Argyres 2018).[1]

This paper develops a framework for theory-based decision making under uncertainty, emphasizing the need for executives to select and test theories before committing to actions. By microfounding the theory-based view and providing a normative protocol for theory construction and selection, this paper contributes to both strategic management literature and practical decision making. It offers a pathway for firms to sustain their competitive advantage.

Drawing on concepts from decision science and strategic management, we argue that effective strategic decision making lies in the formulation and experimentation with theories. This structured and disciplined approach, akin to scientific methods, allows decision makers to navigate uncertainty by grounding their strategies in causal logic and rigorous testing (Felin and Zenger 2017; Camuffo et al. 2020, 2024; Carroll and Sorensen 2021; Ehrig and Schmidt 2022; Zellweger and Zenger 2023).

Our theory-based approach builds upon recent developments in decision science, which advocate the use of theories for structured and informed decision making under uncertainty (Klibanoff et al. 2005; Cerreia-Vioglio et al. 2013, 2022; Hansen and Marinacci 2016; Denti and Pomatto 2022; Hansen and Sargent 2022; Karni 2022). We apply these contributions to strategic decision making in firms and use them as the basis for normative prescriptions about how to deal with uncertainty. Our key intuition is that, when there are few or no data (Choi and Levinthal 2023), the source of value creation and competitive advantage shifts from choosing the best action for known decision problems to formulating and choosing novel decision problems through theories.

We also show that the value of experimenting with theories increases with the number of theories and their uncertainty, revealing a superadditive property: experimenting with one uncertain theory enhances the benefits of experimenting with others. Furthermore, we show that decision makers should experiment with more "surprising" and less plausible theories, as these can provide greater learning and insights.

Our framework is illustrated through motivating examples such as the transformation of Luxottica into a fashion eyewear leader and the strategic foresight behind PayPal's success in digital payments. These cases underscore the practical applicability of the approach and suggest that competitive advantages can arise from well-formulated and empirically tested strategic theories.

Section 2 provides a motivating example that illustrates the concepts and principles of the framework. Section 3 focuses on the theory-building part of the framework, whereas Section 4 focuses on the theory-testing part (experiments). Section 5 concludes. Three appendices provide proofs of propositions, a generalization of the framework and a demonstration of its versatility by applying it to a different firm example, and a glossary of definitions of concepts.

2. A Motivating Example

Founded by Leonardo Del Vecchio in 1961, EssilorLuxottica is now the world leader in designing, manufacturing, and distributing ophthalmic lenses, frames, and sunglasses (Camuffo 2003). Since the 1970s, Luxottica's business model targeted individuals with eyesight defects, a large and steadily growing global market. Opticians played a key role, as they deployed ophthalmologists' prescriptions and assembled lenses and frames. Del Vecchio understood that controlling opticians meant controlling the market: larger production volumes led to lower unit costs because of economies of scale and learning curves, enabling lower prices, greater market penetration, and higher market share.

However, Del Vecchio envisioned alternative futures for Luxottica. He realized that eyeglasses, once seen as necessary but unattractive medical devices, had become widespread and lost their social stigma. Instead, they were increasingly viewed as symbols of distinctiveness, high income, and education. In the early 1980s, amid the global popularity of Italian style and fashion, Del Vecchio began to see eyeglasses as potential fashion accessories and started to think about a new possible world for the eyeglass industry. Instead of being an unpleasant and ugly medical device, eyeglasses could become an adornment of one's face, an aesthetic accessory to complement one's image and style used by everyone.

He theorized that if Luxottica could design and produce stylish eyeglasses, a new market for fashionable eyewear would emerge. This market would shift value from opticians to direct access to customers, who would seek eyeglasses as personal style statements. By involving fashion designers and brands, and controlling the customer experience through retail, Luxottica could command premium prices, achieve low costs through global supply chain efficiency, and sustain business growth and profitability.

This eyewear theory was novel and unexplored. Nobody had thought about (a) the existence of a global market for eyeglass frames other than providing support to lenses for correcting eyesight defects, (b) making eyeglasses an accessory to complement one's outfit and express one's style, and (c) leveraging fashion and design as a way to increase the likelihood that this new market would emerge.

Observations such as Optyl's success with Christian Dior and Safilo's acquisition of Optifashion hinted at the potential for merging fashion with eyeglasses. Del Vecchio further tested his theory through licensing agreements with Armani and acquisitions of small sunglasses producers like Briko and Persol. These experiments gradually strengthened his beliefs in the eyewear theory. Luxottica's strategy evolved into accelerated growth through licensing agreements with iconic fashion brands (e.g., Armani, Bulgari, Chanel), acquisitions of major retail chains (e.g., LensCrafters, Sunglass Hut), and iconic brands (e.g., Ray-Ban, Oakley). Today, Luxottica is the undisputed leader in the luxury eyewear market that Del Vecchio originally envisioned.

In the stylized and formalized framework of the next section, Del Vecchio envisions four states: (i) ability to design fashion glasses and high demand; (ii) inability to design fashion glasses and high demand; (iii) ability to design fashion glasses and low demand; (iv) inability to design fashion glasses and low demand. Although the four states are all possible, Del Vecchio's theory enables him to believe that the first is more likely to occur than the others.[2]

The Luxottica case illustrates how theories can help in devising innovative strategies. As discussed in the next section, theories are conceptual causal structures through which executives become persuaded that some states—within a given future state space—are more (or less) likely to occur than others. In our example, causal reasoning made the future state of interest (rise of large market for eyewear) more likely to occur *in the strategist's mind*. A theory, such as Luxottica's, makes executives believe what others do not.

In the following sections we develop a framework about how to build and evaluate theories.

3. Building Theories

3.1. Background

In a classical decision problem, decision makers do not know which state will materialize among a set of potential future states. However, if they know the possible states and their probabilities, they can compute the expected values of different actions and select the one with the highest expected value. This knowledge relies on having data from similar problems solved in the past, allowing decision makers to draw analogies. For example, Luxottica's experience with spectacles for eyesight defects meant that when the company began exporting to other countries, it understood the importance of controlling opticians. Based on its experience in earlier markets, Luxottica could estimate the probability of controlling a new national market. This, in turn, enabled a sequence of factors—from control of opticians to pricing and high demand—that led to optimal decisions.

In contrast, with low-frequency, high-impact decisions, both states and probabilities are unknown. In such cases, decision makers can use theories to formulate the decision problem and identify states and probabilities. For instance, when Del Vecchio envisioned the new eyewear market, there were few data to confirm its existence. However, through logical reasoning, he hypothesized that the availability of stylish eyeglass frames increased the likelihood of the market existing. What seemed unlikely without a theory appeared possible to him because of his assumptions and causal links.

3.2. Space of Attributes

Decision makers start to build their theories by identifying attributes of the problem, which are elements of a future state space with uncertain realizations.[3] Among the many attributes that they can choose, decision-makers focus on the attributes that they believe affect the likelihood of the occurrence of a state of interest. This state of interest is the problem that decision-makers focus on. For example, Del Vecchio was interested in the rise of a market for eyewear.

Del Vecchio became interested in this problem starting from the observation that wearing eyeglasses was seen less as the signal of disability and more as a proxy of high education and social prominence. He then thought that, if Luxottica could successfully design, produce, and market stylish eyeglasses, the rise of this market would be more likely. Del Vecchio focused on two attributes, whether Luxottica could successfully design, produce, and market fashion eyeglasses, and whether there would be a market, that is, high demand for fashion eyeglasses. Let these two attributes and their spaces be

$$X_f = \{Y, N\} \qquad X_d = \{H, L\},$$

where subscripts f and d stand for "fashion" and "demand," Y and N for *yes* and *no*, and H and L for *high* and *low* demand. A simple way of thinking about *attributes* is that they are random variables whose realizations are *answers to questions*. In our example, the questions are "will Luxottica be able to design, produce, and market fashion eyeglasses?" and "what will demand for fashion glasses be?" The latter question is the problem of ultimate interest to the decision maker, whereas the former question is the logical antecedent (the assumption or cause) that the decision maker uses to increase the belief of successfully solving it (high demand).

Attributes define a state space which is the Cartesian product of the individual attributes. In our case, the two attributes define a state space $X = X_f \times X_d$ made of four states

$$X = \{(H, Y), (H, N), (L, Y), (L, N)\}. \qquad (1)$$

When facing low-frequency, high-impact decisions, decision makers have limited information about the state space. This information comprises sparse empirical observations and "pseudo-observations," which are observations in the head of decision makers. Hence, they fit preliminary probability distributions on X stemming from "soft" rather than "hard" empirical frequencies that occur in worlds that decision makers envision in their mind.[4] Decision makers are interested in a relevant subset of states for their decision. In Luxottica's case, this is state H (high demand for eyewear) and Del Vecchio's goal is to strengthen his subjective belief that this state will occur, that is, increase $P(H)$, which is the sum of the probabilities that generate high demand:

$$P(H) = P(H, Y) + P(H, N). \qquad (2)$$

3.3. Causal Links and Models

The decision makers' goal is to become persuaded that their theory works, that is, that thanks to their theory, their future state of interest will occur. In our example, Del Vecchio conjectures about the reasons why the probability $P(H)$ is high or low. To do so, decision makers generate causal links among attributes, that is, they generate theories. Theories shape decision makers' subjective probability distributions (the beliefs) on a state space like X. More specifically, the choice of attributes and causal links strengthens or weakens these beliefs, which concentrates (or not) the probabilities on the state of interest. Stronger beliefs correspond to probabilities closer to one. Weaker beliefs correspond to probabilities closer to zero. If they expect $P(H)$ to be closer to one, decision makers are more persuaded that the state of interest occurs; vice versa if they expect $P(H)$ to be closer to zero.[5]

In Luxottica's case, Del Vecchio believes that the ability to produce high-fashion eyeglass frames raises the probability of high demand. We represent causal theories of value as Bayesian networks (Pearl 2009). In our simple case we have

$$X_f \longrightarrow X_d$$

and $P(H|Y) > P(H|N)$.

This inequality captures the essence of Del Vecchio's theory: (a) a set of relevant attributes (state space X), (b) a causal link between the antecedent attribute (fashion design) and the consequent attribute (size of the market/level of demand), and (c) subjective probability distributions on the causally linked attributes.[6] Del

Vecchio believes that X is the relevant state space for his problem, and no other attribute or causal link matters. Of course, this is an extreme characterization of the problem to sharpen our exposition. In reality he might consider other attributes or causal links. But the general point is that decision makers focus on a state space that they believe captures the relevant attributes of their problem, and they prioritize some attributes and causal links.

When they start formulating their problem, decision makers have sparse pseudo-observations in their mind. As a result, the probability distributions that they fit on X are random variables themselves because they are not based on objective observations. Let θ_{HY} and θ_{HN} be two parameters of Del Vecchio's empirical probability distribution on X that represent, respectively, $P(H|Y)$ and $P(H|N)$. In addition, let θ_Y be a parameter of the same distribution that represents $P(Y)$—that is, the probability that Luxottica is able to produce a good design. This generates the *statistical model* (Marinacci 2015, p. 1037):

$$\begin{aligned}\mathbb{E}P(H) \equiv v(\theta) &= \theta_{HY}\theta_Y + \theta_{HN}(1-\theta_Y)\\ &= \theta_{HN} + \theta_Y(\theta_{HY} - \theta_{HN}),\end{aligned} \tag{3}$$

which is the expected probability of $P(H)$ in (2), using the product of conditional and marginal probabilities in lieu of joint probabilities.

A specific set of parameter values $\theta = \{\theta_{HY}, \theta_{HN}, \theta_Y\}$ identifies one probability distribution, and therefore one expected probability of high demand $v(\theta)$. But decision makers are aware that their empirical distributions are random variables because the parameters θ do not stem from objective frequencies of the states in X. Therefore, decision makers also have, in their mind, a subjective probability distribution $\mu(\theta)$ of the parameters θ. This generates a family of probability distributions defined by the different θ in the space of θs. Note that this is a different space from X. The space X is the space generated by the realizations of the attributes of the problem. The space of the parameters θ is the space of the realizations of these parameters, which generates the space of realizations of the expected probabilities $P(H)$ of high demand, that is, $v(\theta)$.

This extra step of considering the probability distributions of the parameters θ is essential for a complete theory-based framework for firm strategies.

First, it realistically represents what actually happens. Because decision makers imagine future scenarios and develop theories about future states, the probabilities of these states are not fixed but are uncertain. Therefore, decision makers cannot rely on a single probability distribution for a state space. Instead, they must consider multiple probability distributions based on different possible values of the parameters θ. On this collection of distributions, decision makers have a probability distribution, $\mu(\theta)$.

Second, this extra step helps us better understand the nature of theories, how they are created, and their function. By acknowledging that the parameters θ themselves have a distribution, we disentangle different layers of uncertainty and priors and better represent how theories help decision makers believe that what they envision will occur.

3.4. Theories as Restrictions of the Space of Parameters

With no theory, decision makers can pick any probability distribution $\mu(\theta)$ on the space of the parameters θ and these parameters can take any value in their space (0, 1). A theory restricts the space of values that these parameters can take.

For example, Del Vecchio's theory is that $P(H|Y) > P(H|N)$, or in terms of parameters $\theta_{HY} > \theta_{HN}$. This restricts the set of feasible probability distributions of $\theta_{HY.}$ In particular, we can only employ distributions in which θ_{HY} is restricted to be in the space $(\theta_{HN}, 1)$. Thus, we define $\Theta = \{\theta : \theta_{HY} > \theta_{HN}\}$ to be the space of parameters θ under the theory's restriction, and $\mu_\Theta(\theta)$ a subjective probability distribution implied by the theory—that is, decision makers can choose any form of the probability distribution $\mu_\Theta(\theta)$, but it has to be defined on the restricted space Θ.

As an example, suppose that Del Vecchio believed that $\theta_{HN} = 0.5$ with probability 1. Also, suppose that Del Vecchio was confident about Luxottica's ability to design, produce, and market high-fashion eyeglass frames, and sets $\theta_Y = 0.8$ with probability 1.[7] Finally, suppose that θ_{HY} is uniformly distributed.

The choices of $\theta_Y = 0.8$, $\theta_{HN} = 0.5$ and of a probability distribution concentrated on them are assumptions that decision makers set based on higher-order beliefs.[8] There are no antecedent causal links that explain these choices. These assumptions define the boundaries of the theory. The statement $\theta_{HY} > \theta_{HN}$ represents the restricted set of parameters $\Theta = \{\theta : \theta_{HY} > \theta_{HN}\}$ upon which the decision makers ground a stronger belief $P(H)$. This implies that, under the theory, θ_{HY} is uniformly distributed between 0.5 and 1.

Continuing with our example, Del Vecchio's theory implies that the expected probability of high demand, which is the expected $v(\theta)$ in (3), is $V_\Theta \equiv \mathbb{E}_\Theta v(\theta)$, where the subscript Θ denotes that the expectation is taken under the probability distribution $\mu_\Theta(\theta)$ of the theory, and thus within the space of parameters Θ defined by the theory (not all possible parameters). Note that, in (3), $v(\theta)$ is the expected probability of high demand $P(H)$. This is the expected probability for a given set of parameters θ, that is, for a particular probability distribution on the state space X identified by a specific set of parameter values θ. In contrast, V_Θ is the expected value of $v(\theta)$ across a space of parameters θ according to a probability distribution $\mu(\theta)$ on these parameters.

This extra step helps us understand how decision makers think. It gives a more accurate picture of their thought process. For example, when tossing a fair coin, we get 50% heads and 50% tails if repeated many times. This is an objective probability. We have one probability distribution based on this fact. In real-life decisions, we cannot repeat situations infinitely under the same conditions to get objective probabilities. So, the probabilities are subjective and are treated as random variables with their own distribution, $\mu(\theta)$. This distribution represents what decision makers believe would happen if they could repeat the situation many times. Because they can't, they imagine this distribution in their minds.

Decision makers generate theories through these pseudo-observations in their mind. They are aware that the theory—that is, in our example, the belief $P(H|Y) > P(H|N)$—does not imply that the parameters are equal to specific values (such as $p = 0.5$ with probability 1 in the case of tossing a fair coin). Under the theory, the parameters span different values with some probability. However, the space of parameters does not span all possible values of θ, but only the values that do not contradict the theory. This space is $\Theta = \{\theta : \theta_{HY} > \theta_{HN}\}$, that is, the parameter θ_{HY} cannot be smaller than θ_{HN}, and the probability distribution is $\mu_\Theta(\theta)$, where the subscript denotes that this probability distribution—whatever form it takes—is defined over the (restricted) range of parameters defined by the set Θ.

Using the numbers of our example, the expected value of $v(\theta)$ under the theory, that is, V_Θ, is 0.7: The expected values of θ_{HN}, θ_{HY}, and θ_Y, under the theory, are, respectively, 0.5, 0.75, and 0.8, and, using (3), $V_\Theta = 0.5 + (0.75 - 0.5) \cdot 0.8 = 0.7$. If Del Vecchio's theory did not constrain θ_{HY} to be higher than θ_{HN}, the expected value of θ_{HY} would be 0.5, and not 0.75. With no restrictions from theory, the expected value of $v(\theta)$ would then be 0.5.

Note that decision makers' beliefs that their future state of interest occurs are contingent on their choice of attributes (assumptions), causal links (that restrict the parameter space), and the subjective probability distribution on the set of parameters defined by the theory. In particular, two decision makers who have the same theory will have the same restricted set of parameters Θ. However, they might have different probability distributions $\mu_\Theta(\theta)$, which generate different V_Θ.[9]

Let Θ and $\mu_\Theta(\theta)$ be, respectively, the set of parameters consistent with the theory and the subjective probability distribution of these parameters under the space of parameters defined by the theory. Each theory has an expected value, which is *the expected value of the future states of interest under the theory*, that we can write in general terms as

$$V_\Theta \equiv \mathop{\mathbb{E}}_{\theta\in\Theta} v(\theta) = \int_{\theta\in\Theta} v(\theta)\mu_\Theta(\theta)d\theta, \qquad (4)$$

where the integration runs over all the spaces of all the parameters defined by Θ. This expected value is defined on the space of parameters Θ of the theory and on the subjective probability distribution μ_Θ. Under the theory, the parameters outside Θ have probability zero. This is why, other things equal, a causal theory leads the decision maker to believe that the state of interest is *more* likely to occur than the case of no theory. The theory rules out that the parameters can take some values, and this concentrates probabilities of the subset of feasible parameters according to the theory.

We close by defining theory uncertainty. Given the range of parameters $\theta \in \Theta$, theory uncertainty is the dispersion of the probability distribution $\mu_\Theta(\theta)$ of Θ. It is measured by the standard deviation of this distribution, $\sigma_\Theta(\theta)$. Therefore, theory Θ is more uncertain than theory Γ if, given $\mu_\Theta(\theta) = \mu_\Gamma(\gamma)$, $\sigma_\Theta(\theta) > \sigma_\Gamma(\gamma)$. In this case theory Γ has second-order stochastic dominance on Θ.

3.5. Priors on Theories and Null Hypothesis

Under the theory, parameters outside the set Θ have a probability of zero. This detail is crucial. Building a theory implies selecting some attributes and causal links and discarding all the others.

Decision makers might be sure they have picked the "true" attributes and causal links. Alternatively, they have doubts about their selection of attributes and causal links and acknowledge their theories might be "imprecise" or "wrong." For example, they know that they might have mistakenly forgone or considered irrelevant attributes they knew (Hanna et al. 2014). Alternatively, they might be aware of attributes they do not know of, yet (Karni and Vierø 2013, 2017).

In other words, decision makers have a prior on whether the attributes and causal links they have selected are "right."

Notably, we're not just talking about a space conditioned on the parametric restrictions of the theory. μ_Θ is not one part of a broader probability distribution that takes some values when we condition it under the parametric restrictions of the theory, and some other values for the set parameters in contradiction with the theory. It means, instead, that decision makers have a prior on whether their theory is true—that is, on whether the "true" space of parameters is Θ. Let this prior be a probability $\omega \in (0, 1)$.

If $\omega = 1$, decision makers have no doubt about their theories. But if they have doubts about them ($\omega < 1$), then their beliefs on the theory also depends on a "null hypothesis." This is a different configuration of attributes and causal links—that is, of parameters—against which decision makers compare their theory. In Luxottica's example, this corresponds to the probability of high demand for eyewear under other theories that decision makers can think of. For instance, Del Vecchio's null hypothesis could be that $\theta_{HY} = \theta_{HN}$. In this case, he questions the "fashion" attribute and the causal link to "high demand" (he doubts whether $\theta_{HY} > \theta_{HN}$ is true). More broadly, null hypotheses concern different beliefs, causal links, or attributes that decision makers discard or are not aware of, and that can affect the probability of the states of interest differently than what is predicted by the theory. The fact that decision makers question their theories (i.e., $\omega < 1$) and, hence, have a null hypothesis incorporates in our framework counterfactual thinking, which is an essential component of causal reasoning (Pearl and Mackenzie 2018).

Interesting special cases of decision makers' theories are *necessary* or *sufficient conditions*. In Del Vecchio's theory $P(H|Y) > P(H|N)$: (a) if $X_f = Y$ happens, then $X_d = H$ happens; that is, $X_f = Y$ is a *sufficient condition* for $X_d = H$; (b) if $X_f = Y$ does *not* happen, then $X_d = H$ does *not* happen; that is, $X_f = Y$ is a *necessary condition* for $X_d = H$; and (c) $X_d = H$ happens if and only if $X_f = Y$ happens; that is, $X_f = Y$ is a *necessary and sufficient condition* for $X_d = H$. In the first case, decision makers theorize that a good design implies high demand—that is, $P(H|Y) = 1$—and other attributes may also generate high demand—that is, $P(H|N) > 0$. In the second case, they recognize that H cannot occur without Y—that is, $P(H|N) = 0$—but may not occur even if $X_f = Y$—that is, $P(H|Y) < 1$—which means that other attributes can increase the probability of H. In the third case, $P(H|Y) = 1$, and no other attributes can generate the new market; that is, $P(H|N) = 0$.

In these cases natural null hypotheses are $P(H|Y) \neq 1$ or $P(H|N) \neq 0$—that is, the conditions are not necessary or sufficient. A theory stated in terms of necessary and sufficient conditions, with $\omega = 1$, would be a deterministic theory, and a special case of our framework with no uncertainty in the causal structure or null hypothesis. In principle, these theories need no experiments, as we shall discuss in the next section. In our framework, null hypothesis and $\omega < 1$ reflect the fact that decision makers are aware that their theories may be wrong, and therefore need to learn about them through experiments.

Let $\tilde{\Theta}$ be the set of parameters under the null hypothesis. If the null hypothesis is $\theta_{HY} = \theta_{HN}$, while holding everything else true, this set is $\tilde{\Theta} = \{\theta : \theta_{HY} = \theta_{HN}\}$. Using the data of our earlier example, the expected value of $v(\theta)$ under the null hypothesis would be $V_{\tilde{\Theta}} \equiv \mathbb{E}_{\tilde{\Theta}} v(\theta) = 0.5$, where the subscript $\tilde{\Theta}$ indicates that the expected value, and the subjective probability distribution $\mu_{\tilde{\Theta}}$, is defined over the parameter space $\tilde{\Theta}$. If decision makers believe, instead, that other elements of the theory cannot be held true, they expect a different $V_{\tilde{\Theta}}$. This will involve alternative beliefs or causal links, producing different outcomes that may in turn stem from different processes ranging from informed guesses to a more structured logic. Moreover, if decision makers believe that there could be attributes that, at the moment of their decision, they cannot imagine, their prediction about $V_{\tilde{\Theta}}$ might involve behavioral elements, such as whether they are pessimistic, optimistic, or neutral about these unknown future states.

Analogously to (4), the expected probability of the states of interest under the null hypothesis is

$$\mathop{\mathbb{E}}_{\theta \in \tilde{\Theta}} v(\theta) \equiv V_{\tilde{\Theta}} \equiv \int_{\theta \in \tilde{\Theta}} v(\theta) \mu_{\tilde{\Theta}}(\theta) d\theta, \tag{5}$$

where the restriction on the set of parameters is now $\tilde{\Theta}$, and $\mu_{\tilde{\Theta}}$ is the probability distribution of the parameters under the null hypothesis.

3.6. Expected Value of Theory

The *expected value of the theory* is then

$$V = \omega V_{\Theta} + (1 - \omega) V_{\tilde{\Theta}}. \tag{6}$$

This is the weighted average of the expected probability of the targeted events—$P(H)$ in our Luxottica example—under the theory, and under a broad expectation of $P(H)$ across any other future scenarios in which the theory is not true, using as weights the belief in the theory or the null hypothesis. In our numerical example, $V = \omega \cdot 0.7 + (1 - \omega) \cdot 0.5$.

Throughout our discussion, we focus on the subjective probabilities of decision makers. We assume that when they create and choose new theories, it is difficult, if not impossible, for them to estimate the value of their possible actions based on these theories. This is because, at an early stage, the connections between actions, consequences, payoffs, and states are still uncertain. Decision makers could develop theories about these relationships too, but doing so would be extremely demanding and challenging.

From a modeling perspective, including payoffs would require changing the current "explore-then-commit" approach used in this paper and instead adopting a different approach, like multiarm and multistage bandits (Garivier et al. 2016), to model theory development, selection, experimentation, actions/investments, and returns all at once.

In the current approach, we assume that decision makers first determine whether a future state space is likely (probabilities) by focusing on "plausible" strategic problems that could be valuable. Only after this do they consider payoffs, which involves modeling the uncertain relationships between actions, consequences, and states.

For example, Thiel and Masters (2014) argue that envisioning something new ("going from 0 to 1") is logically antecedent to and fundamentally different from "going from 1 to n." The idea is that first decision makers need to create what is possible (what novel state space they believe will occur) and then build it and scale it up, making experiments and decisions based on payoffs.

Focusing on probabilities (the expected values of theories V) evokes Arrow-Debreu's general equilibrium model in which, in the primitive stage of an economy, individuals are endowed with securities associated with states that earn one dollar if they realize and zero otherwise (Arrow and Debreu 1954). Similarly, when exploring the unknown, decision makers focus on the expected values of theories V (i.e., concentrating subjective probabilities) that can be interpreted as state prices à la Arrow-Debreu.

4. Experiments

4.1. Experiments with Theories

Decision makers learn about their theories by conducting experiments. Experiments are deliberate attempts to collect additional information about states. In our framework, where decision makers develop theories to build priors of future states and have priors on theories, experiments can be informative about a theory, about a theory's capacity to generate a future state of interest, and about the probability of a future state of interest under the theory.

Decision makers can conduct thought experiments or real experiments. Thought experiments involve applying thought processes (like deduction, induction, abduction, analogy, counterfactuals, categorization, and first principles) to test beliefs, whereas real experiments involve collecting real data or observations. Decision makers can produce these observations either from quantitative data analyzed using statistical tools or by drawing qualitative information from phenomena.

For example, in Luxottica's case, Del Vecchio's observations that the use of eyeglasses no longer came with a social stigma but was a potential element of distinctiveness, and the increasing popularity of Italian fashion design, can be thought of as a series of conceptual experiments through which Del Vecchio formed and tested beliefs about a theory of Luxottica as a fashion eyewear company. Then, as discussed in Section 2, he conducted real experiments. He first observed early attempts to link fashion to eyeglasses by competitors, and then conducted early acquisitions and alliances with fashion brand companies. They were all initial attempts to learn about the theory by acquiring information before deciding whether to commit.

Experiments are crucial in building theories of value and selecting among them because, thanks to the additional information they provide, they allow decision makers to update the expected value of their theories. Such updates can occur through the updates of the probability distribution $\mu_{\Theta}(\theta)$ or $\mu_{\tilde{\Theta}}(\theta)$ of the parameters under the theory, or the prior ω that the theory is true.

Using the updated probability distributions $\mu'_{\Theta}(\theta)$ and $\mu'_{\tilde{\Theta}}(\theta)$ in (4) and (5), experiments yield the updated V'_{Θ} and $V'_{\tilde{\Theta}}$, which update V' in (6). For instance, Del Vecchio could find that $\theta_{HY} = \theta_{HN}$ is more likely than he expected (the "fashion" attribute is irrelevant), and update $\mu'_{\Theta}(\theta)$ and $\mu'_{\tilde{\Theta}}(\theta)$ accordingly. This will lower V'_{Θ} and raise $V'_{\tilde{\Theta}}$. The experiments can be on any subset of parameters of the theory. They can be joint experiments, when they focus on more parameters at the same time.

Experiments can also update the belief ω that the theory is true. According to Ortoleva (2012), when decision makers face important contradicting evidence with what he calls *dynamic coherence*, which is a set of coherent beliefs that they hold, they will update their prior on the problem they are working on. These coherent beliefs are analogous to our theories in that Ortoleva has in mind logical links contradicted (or possibly supported) by evidence.

In our framework, this implies that, when the distance $\|\mu' - \mu\|$ between the two distributions is higher than a threshold μ^*, decision makers also change ω to ω'. Intuitively, this says that if the experiment does not change expectations about parameters radically, decision makers only change their likelihood about the parameters. Beyond the threshold, V_{Θ} and possibly $V_{\tilde{\Theta}}$ change not only because of μ but also because of ω.

For example, suppose that Luxottica obtained feedback from the alliance with Armani that suggested that a high-fashion design was slightly more (or less) important than they expected. This would only affect μ, and through it, from (4), V_{Θ}, and then V from (6). If, instead, the feedback was radically different from what was expected, decision makers like Luxottica could strengthen or weaken their belief ω on the theory, and give more or less weight to V_{Θ} or $V_{\tilde{\Theta}}$ in the expression for V. In both cases the ultimate effect is the same, a change in V. However, the channels are different in that in one case it is an adjustment of the weights of the probability distribution given the theory and the null hypothesis, whereas in the other case it is a stronger or weaker belief on whether the focal theory is the one to focus upon. Moreover, based on the discussion above, a different prior on the theory is likely to produce a stronger change in V. As we will see in the next section, this puts greater weight on the fact that decision makers commit to a different problem and an alternative theory.

After the experiment, the *expected value of the theory* becomes

$$V' = \omega' V'_{\Theta} + (1 - \omega') V'_{\tilde{\Theta}}, \tag{7}$$

where the update ω' occurs only if the evidence is sufficiently contradictory to induce changes in beliefs.

4.2. Experiments Against Alternative Theories

In our framework, decision makers imagine future states and theories; there is no objective reality to discover. Therefore, they can't use an objective standard to evaluate their theories' expected value. When running experiments on a theory, they can only compare it with other theories they imagine. If the results are unfavorable, they consider these alternatives, which are also subjective. Decision makers can only experiment with different plausible theories and state spaces, then choose the most plausible one. This means they can never be sure their chosen theory or state space is the "right" one. The goal is to envision the future, not to find an absolute truth. The alternative theories allow for counterfactual thinking but are different from the null hypothesis, which is another way of configuring parameters within the same state space. An alternative theory comprises a different state space characterized by different attributes, causal links, and a different future state of interest.[10]

Luxottica's alternative theory was about its existing successful business on spectacles for eyesight defects, rather than eyewear. Del Vecchio realized that eye surgery was picking up, and might have jeopardized its traditional business. He set attributes $X_s = \{D, \overline{D}\}$ and $X_g = \{E, C\}$ for whether eye surgery will diffuse (D) or not ($\overline{D}$), and whether demand for traditional glasses expands (E) or contracts (C). The causal link is $X_s \rightarrow X_g$, and the restriction implied by the theory is $P(C|D) > P(C|\overline{D})$ because high surgery contracts demand, against the null hypothesis that the two probabilities are equal.

Using different parameters and priors, decision makers retrieve Q, the unconditional expected value of the alternative theory in the same way as we obtained V in (6). They then compare V and Q and pick the more plausible theory—that is, the theory with a more concentrated probability that the state of interest will occur. In our example, this depends on whether $V > Q$ or vice versa. Experiments on the first theory update V to V', whereas experiments on the second theory update Q to Q'. Decision makers can run experiments on one or the other theory, and they can run more than one experiment till they are sufficiently confident to commit to one or the other. As we will also discuss below, the optimal stopping time depends on the cost of the experiments and the discount factor because running experiments takes time.[11]

4.3. Bayes Plausible Experiments

We distinguish between two types of experiments: *Bayes Plausible Experiments*, which we discuss in this section, and *Biased Experiments*, which we discuss in the next section.

Bayes Plausible Experiments rely on the assumption $\mathbb{E}V' = V$ and $\mathbb{E}\omega' = \omega$, and equivalently for the alternative theory ($\mathbb{E}Q' = Q$, and similarly for its belief analogous to ω). Kamenica and Gentzkow (2011) coined this term for experiments based on this assumption, which is standard in the literature on information design. According to this assumption, the expected value of the theory after the experiment equals—in expectation—its current expected value. The logic is that, before running the experiment, decision makers are aware of all the relevant information about their problem. Therefore, before running the experiment, they have no reason to expect that, after the experiment, the outcome will weigh to a greater extent one or the other side of the current expected value. Only the experiment reveals, ex post, new information that updates beliefs in one or the other direction.

Moscarini and Smith (2001) have studied the optimal stopping time when decision makers carry out a sequence of experiments of this kind. If experiments are costly and the value of the information that they produce is discounted, decision makers stop experimenting when the experiment yields a probability of the desired event sufficiently close to zero or one. In this case, decision makers become sufficiently certain that the event will occur or not. Otherwise it is optimal to acquire more information by running a new experiment.

In Appendix A we extend the Moscarini and Smith (2001) model to two theories with expected values V and Q. At each stage of their dynamic program, decision makers have four options: commit to theory Θ or the alternative theory, or run an experiment on one or the other theory to acquire more information about it and update V or Q. When V is very high with respect to Q, decision makers commit to Θ and obtain V; when V is very small compared with Q, decision makers commit to the alternative theory and obtain Q. In between it is worth running an experiment on either Θ or the alternative theory till the updated ratio between V and Q tilts in one or the other direction.

The assumption that the expected outcome of the experiment is equal to the prior before the experiment is crucial to obtain this result. The intuition is that if V is relatively high compared with Q, the experiment is unlikely to make V much higher (because V is bounded by one). Simply put, if you are nearly certain about an event, you cannot become much more certain. But if a positive update of V cannot be much higher, in order to preserve $\mathbb{E}V' = V$, the negative update of V cannot be much smaller. The extra information provided by the experiment is then limited because the ratio of V to Q will not change much. Therefore, it may then not be worth incurring the cost of the experiment to obtain delayed information. As a result, decision makers do not run the experiment, commit to Θ, and obtain V. The logic is symmetric when V is much smaller than Q, in which case decision makers commit to the alternative theory and obtain Q.

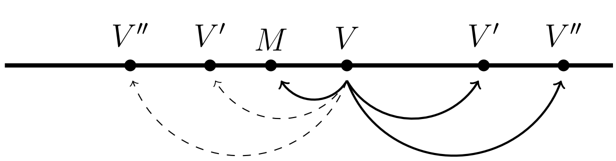
Appendix A also shows that the value of experimenting with more uncertain theories is superadditive—that is, experimenting with a more uncertain theory increases the expected value of experimenting with a more uncertain alternative theory. This is an important result that highlights the value of embracing uncertainty to create value. In the remainder of this section, we first discuss the intuition of this result, and then its implications. Appendix A provides a formal proof.

As far as the intuition is concerned, let's start by focusing on the decision to experiment with theory Θ. Decision makers know that, after the experiment, when they observe the update V', either V' is sufficiently high or sufficiently low compared with Q, in which case they commit to theory Θ or the alternative theory, or they will run a new experiment on Θ or the alternative theory, depending on the conditions discussed in Appendix A. The outcome of the experiment on Θ is bounded from below: Before running the experiment, decision makers hold V, and after the experiment either they obtain V', if V' is higher than all other three options, or they obtain the best of these three alternative options. But this means that the outcome of the experiment is more favorable than $\mathbb{E}V'$ because when V' is higher than the other three options, the outcome of the experiment is V'; when it is lower, it will be the best alternative option, and therefore it will be higher than V'. As a result, the expected outcome of the experiment is higher than V. Whether the value of the experiment is worth the cost of the experiment or the delayed information is to be seen, but this establishes our earlier statement that the existence of alternative options rules out that it is never worth running an experiment.

The lower bound to the outcome of the experiment also makes experiments on more uncertain theories more valuable. Intuitively, a more uncertain theory implies that the experiment will yield updates V' more distant from the expected value V both above and below V. Figure 1 below illustrates the point.

In Figure 1 an experiment on a theory that, after the experiment, yields potential updates V'' is more uncertain than a theory that after the experiment yields potential outcomes V' closer to the expected outcome V of the experiment. The greater uncertainty is inherent in the fact that the expected value is a less precise predictor of the outcome of the experiment. At the same time, the more uncertain experiment is more valuable because if the experiment yields a favorable outcome, V'' is greater than V', and if it yields an unfavorable outcome decision makers settle on the best alternative option, whose outcome is represented in the figure by M. But this means that the upside of the experiment on the more uncertain theory is higher, and the downside is the same, which makes the experiment on the more uncertain theory more valuable.

Figure 1. Uncertain Theories



Finally, to establish superadditivity, the value of the alternative option M is the best among the three values of committing to the alternative theory, or running an experiment on the current or alternative theory. But if the alternative theory is also uncertain, the value of experimenting with it is higher, which increases the value of the alternative option M in Figure 1. Appendix A shows formally that this increase in M raises the value of an experiment with the theory that yields V'' more than the value of an experiment on the theory that yields V'. As a result, a more uncertain alternative theory raises the value of experimenting with a more uncertain theory.

The greater value of experimenting with more uncertain theories supports the conjecture about the value of testing more "surprising" theories. In addition, our framework is compatible with testing less plausible theories—"contrarian beliefs" (Felin and Zenger 2017)—that is, theories such that, initially, $V < Q$. Thus, it can be optimal to test not only more uncertain theories but also less plausible theories. Luxottica tested its more unconventional theory. Del Vecchio had a good understanding of his traditional business. Eyewear was novel and unknown. Del Vecchio started by observing its competitor Safilo's seemingly minor acquisition of Optifashion, a pioneering company in high-fashion glasses, and then signed initial deals with fashion brand companies to experiment with the new market. More generally, many large high-tech companies today experience hypergrowth because they experiment with many uncertain theories, which generates more learning and, potentially, more value.

This discussion raises the question of what limits decision makers in developing more surprising theories. Besides differences in the decision makers' ability to imagine such theories, more uncertain theories often have lower expected values because they seem too unconventional and implausible. However, as long as the expected values are not drastically different, our framework suggests that experimenting with more uncertain theories can be valuable. Clearly, understanding where good and surprising theories in business strategy originate is an open question for future research.

4.4. Biased Experiments

Decision makers could be aware of forgone or unknown attributes or, more generally, that their theories might not be true ($\omega < 1$) (Karni and Vierø 2013, Hanna et al. 2014). In this case, they conduct *biased* experiments, that is, experiments designed to elicit *biased* signals (Gans 2023).

Biased experiments are designed so that they are in favor of or against one of the theories, contingent on whether the anticipated unknown events are positive or negative (Karni and Vierø 2017). They are characterized by a sample or experimental conditions overwhelmingly favorable or unfavorable to a theory. The rationale is that if a positively biased experiment with an unlikely theory produces updates smaller than the alternative theory, it reinforces the beliefs about the commitment with the alternative theory (Adner and Levinthal 2024). Similarly, if a negatively biased experiment produces updates higher than the lower-probability theory, it reinforces the beliefs about the more likely theory. Experiments characterized by Heisenberg effects (Shelef et al. 2024) or conducted to elicit signals about entangled theories (Barney et al. 2024) are types of biased experiments.

In our framework, decision makers run biased experiments when, according to an equivalent Bayes Plausible Experiment, V and Q are sufficiently distant to commit to one or the other theory. In this case, they add a further element of doubt, asking themselves whether there could be unforeseen contingencies that break the assumption $\mathbb{E}V' = V$.

There are four potential types of biased experiments. Suppose that V is very small compared with Q, and decision makers commit to the alternative theory. In this case, they may run a positively biased experiment with theory Θ or a negatively biased experiment with the alternative theory. The question is whether the positively biased experiment with theory Θ confirms the commitment to the alternative theory—that is, even a positively biased experiment on Θ does not provide enough information in favor of theory Θ. Similarly, a negatively biased experiment against Q informs decision makers on whether the alternative theory is really superior. The logic is symmetric if V is very large compared with Q and Bayes Plausible Experiments commit to Θ.

Biased experiments are therefore designed to test ω, that is, whether the state space is true or false. They either radically question theories (i.e., they are designed to collect a signal that "kills" the theory) or radically support them (i.e., they are designed to collect an extremely favorable signal). If this happens, decision makers update ω and can change radically the expected value of their theories after the experiment.

Finally, decision makers need to decide on which theory to run the biased experiment among the four options discussed above. This depends on the ability of decision makers to design informative biased experiments, by identifying which bias to introduce in the analysis and whether they can find a more informative bias if they test Θ or the alternative theory.

5. Conclusion

This paper argues that strategic decision making under uncertainty starts with choosing theories before choosing actions and that decision makers experiment with theories until the present value of experimenting with them is larger than their previous value. Our approach addresses the challenges faced by executives making "low-frequency, high-impact" decisions where past data are unavailable or insufficient.

We show that strategic decision making should begin with the selection of theories rather than immediate action choices. Executives should experiment with theories, using a structured and disciplined approach akin to scientific methods, until the value of further experimentation becomes negligible. The process of experimenting with theories is particularly valuable when the theories are numerous and uncertain.

Moreover, experimenting with sufficiently uncertain theories is superadditive; that is, uncertain theories are complementary, or experimenting with one uncertain theory enhances the benefits of experimenting with other uncertain theories, thus maximizing the learning and insights that can be derived from these experiments.

Furthermore, we underscore the importance of grounding strategic decisions in causal logic and rigorous testing of theories. This structured approach facilitates more informed decisions under uncertainty and improves the likelihood of successful outcomes. Interestingly, we also find that decision makers should not shy away from testing more "surprising" and less plausible theories. Despite their initial lower expected value, these theories can provide significant learning opportunities and potentially higher long-term benefits, reinforcing the notion that unconventional thinking can be a source of competitive advantage.

Prior beliefs and null hypotheses play a crucial role in framing theories and conducting experiments. Incorporating counterfactual thinking into the decision-making process enhances causal reasoning and the robustness of strategic choices. This aspect of our framework ensures that decision makers remain critical and reflective about their assumptions, thereby improving the quality of their strategic decisions.

The paper contributes to the theory-based view of strategy in four ways. First, it microfounds the concept of theory by articulating the relationship between conceptual causal structures and beliefs using Bayesian networks. Second, it provides a normative protocol for theory construction and selection. Third, it clarifies the causal nature of theories of value by highlighting when conditional probabilities reflect causation and not correlation and by using countertheories (null hypotheses) and alternative theories to incorporate counterfactual thinking. Fourth, it illustrates how decision makers form prior beliefs about future states in nonergodic decision contexts.

Additionally, the paper connects with recent developments in experimental strategy (Gans 2023, Barney et al. 2024, Shelef et al. 2024), illustrating the conditions under which decision makers might use unbiased or biased experiments to choose among alternative theories of value.

The implications of our findings for strategic decision making are profound. Executives should ground their strategic decisions on well-formulated theories and rigorous experiments, especially in scenarios where past data are not available or useful. This approach enables more informed and effective strategic choices, providing firms with a sustained competitive advantage. The ability to experiment with and select better theories leads to better-formulated decision problems and continuous learning, which in turn fosters innovative and resilient business strategies.

We see multiple research avenues stemming from the extension of our framework. Experimenting with and selecting better theories could represent a source of sustained competitive advantage as the choice of actions is based on better-formulated decision problems, and the experimentation to select theories can generate anomalies that continuously breed new theory development. The framework suggests a mechanism by which firms can hyper-scale because of the superadditivity of experimenting with sufficiently uncertain theories providing microfoundations of the role of uncertainty in strategic decision making.

We leave several open questions for future research. We do not provide a theory of how decision makers identify, select, and combine attributes and logical links to formulate and rank theories. Relatedly, it would be also important to investigate the determinants of learning speed, as the rapid exploration of alternative theories can represent a source of competitive advantage. Our analysis assumes identity between decision makers and firms, and future research should investigate how firms develop routines, design organizations, and build management systems that allow this framework to be effectively, consistently, and efficiently deployed across complex organizations.

Appendix A

In this appendix we extend the Moscarini and Smith (2001) framework to the case of two theories. Decision makers experiment in stages, and in each stage they can either commit to one of the two theories or run an experiment on one or the other theory.

Without loss of generality we focus on an experiment on theory Θ. Let $c > 0$ be a fixed cost of running an experiment, $\rho \in (0,1)$ a discount factor accounting for the fact that the experiment delays the final decision, and $H(V')$ the cumulative probability distribution of the update V' of V from the experiment on Θ. We assume that $\mathbb{E}V' = V$ and $\mathbb{E}\omega' = \omega$, where ω' is the postexperiment update of ω, and similarly for the alternative theory.

In each stage of the dynamic program, the condition for running an experiment is

$$V^E \equiv \rho\left[\int_{Q^*}^{1} V' dH(V') + Q^* H(Q^*)\right] - c > \Pi, \tag{A.1}$$

where $Q^* \equiv max(Q, Q^E)$, $\Pi \equiv max(V, Q^*)$, and Q^E is the equivalent value of the experiment on the alternative theory. This condition says that (1) if after the experiment decision makers observe $V' > Q^*$, they will either commit to Θ or run a new experiment on Θ, obtaining V'; (2) otherwise, they either commit to or run an experiment on the alternative theory and obtain Q^*. Decision makers keep experimenting till the present value of the experiment is higher than all the other options V, Q, and Q^E. Assuming decreasing information gains over time from experimentation, V^E and Q^E will get smaller as decision makers keep experimenting till they eventually commit to one of the two theories.

We now establish that, as shown in Figure A.1, there are four zones of exploration depending on the value of V: For V being very small compared with Q, decision makers commit to the alternative theory, and vice versa for V being very high compared with Q; in between, they experiment, and as V increases they experiment with Θ rather than the alternative theory.

Using (A.1), and the fact that $\mathbb{E}V' = V$, rewrite V^E as

$$\begin{aligned} V^E &\equiv \rho\left[V + Q^* H(Q^*) - \int_0^{Q^*} V' dH(V')\right] - c \\ &= \rho\left[V + \int_0^{Q^*} H(V') dV'\right] - c, \end{aligned} \tag{A.2}$$

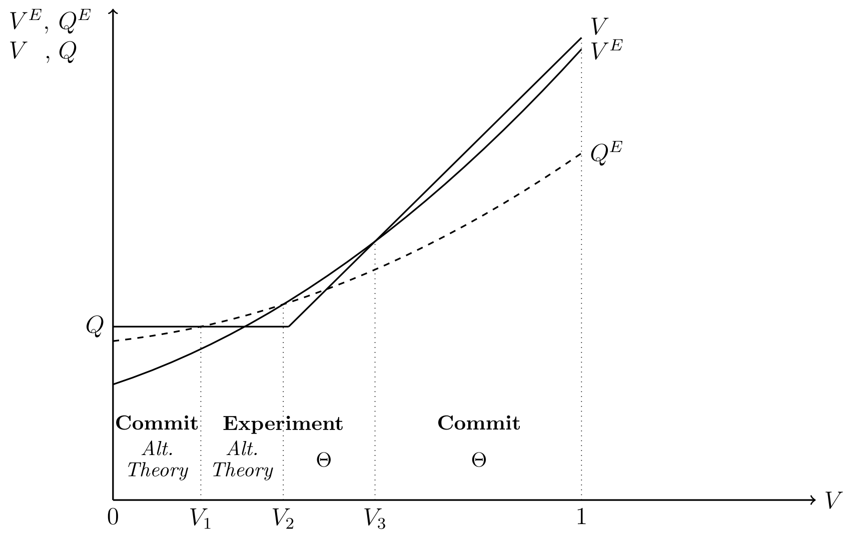
where the second equality stems from integration by parts. We first establish that $\frac{\partial V^E}{\partial V} > \frac{\partial Q^*}{\partial V} \geqslant 0$. Let $D = 1 - \rho^2 H(Q^E) K(V^E)$, where K is the cumulative distribution of the update Q' of Q if decision makers run an experiment on the alternative theory. Take the differentials dV^E, dQ^*, and dV in (A.2) and, if $Q^* = Q^E$, in the equivalent expression for Q^E. Solving the system, or setting $dQ^* = 0$ if $Q^* = Q$, obtain $\frac{\partial V^E}{\partial V} = D^{-1}\rho > \frac{\partial Q^*}{\partial V}$, which is equal to 0 or $D^{-1}\rho^2 K(V^E)$ depending on whether $Q^* = Q$ or $Q^* = Q^E$.

When $V \leqslant Q^*, \Pi = Q^*$, and the fact that V^E increases with V faster than Q^* is a necessary condition to state that there is a threshold (V_2 in Figure A.1) such that V^E switches from smaller to higher than Q^*. This condition is sufficient for appropriate values of ρ or c. Because Q^E grows with V, unlike Q, Q^* could switch from Q to Q^E at a smaller threshold ($V_1 < V_2$ in Figure A.1).

When $V > Q^*, \Pi = V$, and V^E increases at a lower rate than V if $D^{-1}\rho < 1$. Because $max(V, V^E) < 1$, the V^E curve will cut the V curve from above at a threshold V_3 in Figure A.1 such that if $V > V_3$ decision makers commit to theory Θ. The slower growth of Q^E than V^E with respect to V is a necessary condition for $Q^E < V^E$ at $V = V_3$.

Finally, we show that more uncertain theories are superadditive. We capture uncertainty with the notion of second-order stochastic dominance, which increases V^E because it increases the integral in the second expression of (A.2). Solving for the differentials of V^E and Q^* as we did earlier, $dV^E = D^{-1}\rho > dQ^*$, which is equal to 0 or $D^{-1}\rho^2 K(V^E)$ depending on $Q^* = 0$ or Q^E. Because both V^E and Q^E increase, the threshold V_1 in Figure A.1 decreases and the threshold V_2 in Figure A.1 increases. Expected values increase

Figure A.1. Value of Experimentation



Note. Alt., alternative.

because for each V decision makers expected a higher value. The marginal condition (A.2) implies that V^E and Q^E are complementary. Because they are also complementary with an increase in their spread, Milgrom and Roberts (1990) imply that a greater uncertainty of one magnifies the effects of a greater uncertainty of the other.

Appendix B

In this appendix we discuss the case of PayPal. Our goal is twofold. First, we show that the framework is versatile and can be applied to different cases, and we extend it from dichotomous to continuous realizations of attributes.

Max Levchin, Luke Nosek, and Peter Thiel cofounded Fieldlink, believing they could grow a digital security company providing services for personal digital assistant (PDA) devices. They focused on four attributes. $X_d = \{x_d\}$ is the end state of interest, where x_d is a continuous measure of the demand for handheld device encryption software, depending on two attributes: $X_e = \{x_e\}$, a continuous index of the efficiency of the available encryption technology, and $X_s = \{x_s\}$, a continuous index of the perceived need for data security by handheld device users. The attribute $X_s = \{x_s\}$ is determined by $X_h = \{x_h\}$, which captures the spread of handheld devices. We represent their theory with the causal structure shown in Figure B.1, which rests on the following chain of subjective probabilities:

$$p(x_d, x_e, x_s, x_h \mid \theta) = p(x_d \mid \theta_{des}, x_e, x_s)\, p(x_e \mid \theta_e)\, p(x_s \mid \theta_{sh}, x_h)\, p(x_h \mid \theta_h), \tag{B.1}$$

Figure B.1. Theory for Foundation of Fieldlink

$$
\begin{tikzcd}
 & X_e \arrow[dr] & \\
X_h \arrow[dr] & & X_d \\
 & X_s \arrow[ur] &
\end{tikzcd}
$$

where $\theta = \{\theta_{des}, \theta_e, \theta_{sh}, \theta_h\}$ is the parameter set of the distributions.

A causal structure such as (B.1) generates a sequence of expected values that we represent as follows:

$$
\begin{aligned}
\mathbb{E}(x_d \mid \theta_{des}, x_e, x_s) &\equiv v_d(\theta_{des}, x_e, x_s) \\
&= \int_{X_d} x_d\, p(x_d \mid \theta_{des}, x_e, x_s)\, dx_d \\
\mathbb{E}[v_d((\theta_{des}, x_e, x_s) \mid \theta_e, \theta_{sh}, x_h)] &\equiv v_{des}(\theta_{des}, \theta_e, \theta_{sh}, x_h) \\
&= \int_{X_s}\int_{X_e} v_d(\theta_{des}, x_e, x_s)\, p(x_e \mid \theta_e) \\
&\qquad p(x_s \mid \theta_{sh}, x_h)\, dx_e\, dx_s \\
\mathbb{E}[v_{des}(\theta_{des}, \theta_e, \theta_{sh}, x_h) \mid \theta_h] &\equiv v(\theta) \\
&= \int_{X_h} v_{des}(\theta_{des}, \theta_e, \theta_{sh}, x_h) \\
&\qquad p(x_h \mid \theta_h)\, dx_a.
\end{aligned}
$$

Consider for simplicity the following linear approximation

$$v_d(\theta_{des}, x_e, x_s) = \theta_{de} x_e + \theta_{ds} x_s, \qquad \text{with}$$

$$\mathbb{E}(x_e \mid \theta_e) = \theta_e, \quad \mathbb{E}(x_s \mid \theta_{sh}, x_h) = \theta_{sh} x_h,$$

where we distinguish between the two elements θ_{de} and θ_{ds} of the vector of parameters θ_{des} that represent, respectively, the correlations between x_e and x_d and x_s and x_d. By replacing the two expected values in $v(\cdot)$, we obtain

$$v_{des}(\theta_{des}, \theta_e, \theta_{sh}, x_h) = \theta_{de}\theta_e + \theta_{ds}\theta_{sh} x_h$$

and then

$$v(\theta) = \theta_{de}\theta_e + \theta_{ds}\theta_{sh}\theta_h,$$

which is a linear approximation of the expected value $v(\theta)$ of the state (x_d, x_e, x_s, x_h) conditional on the parameter set θ of the underlying probability distribution. In this expression, the subscripts ij of each parameter denote that the parameter accounts for the strength of the causal link from j to i, whereas θ_e and θ_h denote the beliefs, at the top of the causal chain, regarding the quality of the encryption technology and the spread of handheld devices.

The theory is the set Θ with θ_{de}, θ_{ds}, $\theta_{sh} > 0$, and relatively high values of θ_e and θ_s. The expected value of this theory is V_Θ, defined by (4), with a probability distribution $\mu(\theta|\Theta)$ of these parameters. Given a null hypothesis on the parameters and a prior that the theory is true, (6) defines the unconditional expected value of this theory.

Almost in parallel, Levchin, Thiel, and friends thought about alternative theories. In particular, together with Luke Nosek and Ken Howery, and later on Elon Musk, they believed that the rise of the internet, email messaging, and e-commerce would transform financial transactions, requiring the use of encryption to make fast and secure online payments and help internet users transfer money easily while making online transactions. Therefore, in 2000, they founded PayPal, which they envisioned as a growing platform that provided efficient and secure digital payment services in the rising internet economy.

Their theory and causal structure could be represented as shown in Figure B.2.

The future state of interest is attribute $X_p = \{x_p\}$, which denotes the extent to which people will use digital payments for their online purchases. This depends on two attributes: $X_d = \{x_d\}$, which measures the extent to which a secure and efficient digital payment platform will be available, and $X_s = \{x_s\}$, which denotes whether consumers and merchants will need fast and secure digital payments to support online commerce and P2P transactions. In turn, $X_s = \{x_s\}$ is determined by $X_i = \{x_i\}$, which captures the extent to which e-commerce and the internet economy will rise. PayPal's theory can be formalized as

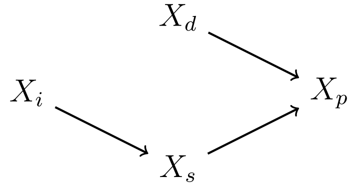
$$p(x_p, x_d, x_s, x_i|\theta) = p(x_p|\theta_{pds}, x_d, x_s)\, p(x_i|\theta_i)\, p(x_s|\theta_{si}, x_i)\, p(x_d|\theta_d).$$

As in the previous theory, here we also consider for simplicity the linear approximation $v(\theta) = \theta_{pd}\theta_d + \theta_{ps}\theta_{si}\theta_i$, where again the subscripts ij of each parameter denote that the parameter accounts for the strength of the causal link from j to i, whereas the single subscripts account for the beliefs at the top of the causal chain.

The theory is the set Θ with θ_{pd}, θ_{ps}, $\theta_{si} > 0$, and relatively high values of θ_d and θ_i. The conditional expected value of this theory is again analogous to V_Θ, defined by (4), with a probability distribution $\mu(\theta|\Theta)$ of these parameters, whereas V, defined by (6), represents the unconditional expected value of the theory given a null hypothesis and a prior that the theory is true.

Compared with the Fieldlink theory, the PayPal theory was initially less plausible and potentially valuable (lower unconditional expected value). The idea of a digital payment platform was totally novel and, absent the attributes and causal links, implausible. But the attributes and causal links Levchin and Thiel were using had the potential to greatly increase its likelihood of occurrence.

Figure B.2. Theory for Foundation of PayPal



This was a "high-variance" theory. Because there was little knowledge about a market for digital payments associated with e-commerce, the subjective probability distributions of the attributes and causal links (parameters θ) were dispersed and their potential contribution to the expected value of the theory was large. Hence, the potential update of V from experiments was large.

As per our framework, Levchin, Thiel, Nosek, and Musk explored this theory and ran experiments to test it. For example, one experiment, conducted in 1999, consisted in the launch of their digital payment system on a specific platform, Power Seller on eBay, comprising 20,000 customers, known for being conservative, difficult, and demanding customers. In our framework, this corresponds to testing, through a biased experiment, the parameters θ_{pd} and θ_{ps}. The experiment showed that the digital payment system worked and, hence, positively updated $\mu(\theta_{pd})$ and $\mu(\theta_{ps})$. The new values $\mu'(\theta_{pd})$ and $\mu'(\theta_{ps})$ implied an updated, larger value V'_Θ, defined by (4), and a larger V', defined by (6). It was under this theory that PayPal became a worldwide diffused payment system.

Appendix C

In this appendix we provide a glossary of key definitions of relevant concepts for a theory-driven approach to low-frequency, high-impact strategic management decisions.

Definition C.1 (Classical Decision Problem). A decision problem is a quartet (A, S, C, ξ) in which

- A is a collection of available actions
- S is the space of all payoff-relevant contingencies called states of nature
- C is a collection of consequences
- $\xi : A \times S \to C$ is a consequence function that details the consequence $c = \xi(a, s)$ of action a when state $s \in S$ obtains.

Definition C.2 (Attributes). An attribute X_j is the set of all the alternative outcomes of one element of a problem. A space of attributes $X = \prod_{j\in J} X_j$ is the space of all the attributes $j \in J$ that DMs believe are relevant for their problem. An attribute X_j and a space of attributes X are maps $D \to X$ that reduce the dimensionality of the exploration problem within the decision makers' domain D.

Definition C.3 (Domain). A domain is the set of all attributes $\mathcal{D} = \prod_{j\in J} X_j$, $J = \{1, 2, \ldots N\}$ that decision makers know or could be aware of at any given moment in time given their knowledge, experience, or reference points. It does not necessarily contain S. This depends on whether decision makers are aware of all the relevant contingencies for the consequences of their problem. Domains change as decision makers learn.

Definition C.4 (Model). A model θ is the realization of a parameter (or vector of parameters) that identifies a specific probability distribution p_θ within a family of probability distributions $P = \{p_\theta\}$ on the space of attributes.

Definition C.5 (Theory). A theory is a family of probability distributions $P_\Theta = \{p_\theta\}_{\theta\in\Theta}$ such that the parameter set is a subset Θ of the set all possible realizations of the parameters θ. We call theory P_Θ or Θ interchangeably.

Definition C.6 (Experiment). An experiment is a map $f:\Theta \to \Delta(Y)$ from models θ to probability distributions $f(y|\theta)$, $y \in Y$, where $f(y|\theta)$ is the probability of receiving signal y under model θ.

Endnotes

[1] A formal definition of decision problem is provided in Appendix C. There is a strong link between the theory-based and problem-focused perspectives on strategic decision making. Theories provide a framework for viewing decision problems as means to address critical issues or anticipated future conditions that need to be tackled (Baer et al. 2013, p. 199).

[2] Throughout the paper we express beliefs as subjective probabilities.

[3] We use the term "attribute" to indicate the elements of a state space, and we represent them as nodes of a Bayesian network. Below, we articulate spaces of attributes as conceptual causal structures distinguishing between attributes that are future states of interest (effects) and attributes that are assumptions (causes). Although logically different, they together constitute the future state space strategists envision and represent the foundation of their theories.

[4] Different approaches have been used to model the subjective process of building distributions for novel state spaces out of sparse information. For example, Gilboa et al. (2020) propose a model that matches case-based reasoning (i.e. sparse past data) and theories to form prior beliefs. Our approach is similar, although we do not limit the role of theories to analogical reasoning on past cases.

[5] Alternatively, decision makers could reason counterfactually (Pearl and Mackenzie 2018) and build a countertheory that increases $P(L)$, thus adopting a falsificationist approach.

[6] The subjective nature of the probability distributions is reflected in the original denomination of Bayesian networks as "belief networks" (Pearl 1986).

[7] Attributes at the top of the Bayesian network represent the theory assumptions (uncaused causes) which can be exogenous events or hypothetical interventions of the decision maker (Pearl and Mackenzie 2018). In the Luxottica case, θ_Y is a "hypothetical intervention" (whether Luxottica will be able to design, produce, and market fashion eyeglasses) for which, as illustrated above, the decision makers *set* the parameters. Differently from the other attributes, in this case conditional probabilities unambiguously reflect causality and not correlations.

[8] We will elaborate on these "priors on priors" in the next section.

[9] Decision makers can elicit different $\mu_\Theta(\theta)$ compatible with their partial knowledge and information. Although Bayesian statistics offer a variety of methods (see Kass and Wasserman 1996 for a review), the principle of maximum entropy—that is, a prior probability assignment that incorporates the fewest assumptions on the data (Jaynes 1968)—is particularly appealing to decision makers who adopt a theory-based approach as it ensures the highest theory variance, which—as we demonstrate in Section 4.3—is an important property of theories when decision makers conduct experiments to choose among them.

[10] More generally, theories might be different because they (1) deal with different problems, that is, they refer to completely different state spaces and states of interest (this is the case of *alternative* theories in our framework), (2) comprise different sets of attributes or use different causal links for a given end state of interest (this is the case of the null hypothesis in our framework), and (3) apply the same set of attributes and causal links to different end states. However, because decision makers choose at which level of granularity they want to develop their theories, whether theories are different or not is subjectively defined.

[11] We compare two probabilities V and Q rather than values, implying that decision makers are learning about the plausibility of theories rather than making a full comparison of expected utilities.

References

Adner R, Levinthal D (2024) Strategic experiments in non-experimental settings. *Strategy Sci*. 9(4):311–321.

Arrow K, Debreu G (1954) Existence of an equilibrium for a competitive economy. *Econometrica* 22(3):265–290.

Baer M, Dirks KT, Nickerson JA (2013) Microfoundations of strategic problem formulation. *Strategic Management J.* 34(2):197–214.

Barney JB, Bigelow LS, Shelef O, Wuebker R (2024) Experimental strategy with entangled ideas. Preprint, submitted May 20, http://dx.doi.org/10.2139/ssrn.4834613.

Camuffo A (2003) Transforming industrial districts: Large firms and small business networks in the Italian eyewear industry. *Indust. Innovation* 10(4):377–401.

Camuffo A, Cordova A, Gambardella A, Spina C (2020) A scientific approach to entrepreneurial decision-making: Evidence from a randomized control trial. *Management Sci.* 66(2):564–586.

Camuffo A, Gambardella A, Messinese D, Novelli E, Paolucci E, Spina C (2024) A scientific approach to entrepreneurial decision making: Large scale replication and extension. *Strategic Management J.* 45(6):1209–1237.

Carroll GR, Sorensen JB (2021) *Making Great Strategy: Arguing for Organizational Advantage* (Columbia University Press, New York).

Cerreia-Vioglio S, Hansen LP, Maccheroni F, Marinacci M (2022) Making decision under model misspecification. Working Paper No. 2020-103, University of Chicago, Chicago.

Cerreia-Vioglio S, Maccheroni F, Marinacci M, Montrucchio L (2013) Classical subjective expected utility. *Proc. Natl. Acad. Sci. USA* 110(17):6754–6759.

Choi J, Levinthal D (2023) Wisdom in the wild: Generalization and adaptive dynamics. *Organ. Sci.* 34(3):1073–1089.

Denti T, Pomatto L (2022) Model and predictive uncertainty: A foundation for smooth ambiguity preferences. *Econometrica* 90(2):551–584.

Ehrig T, Schmidt J (2022) Theory-based learning and experimentation: How strategists can systematically generate knowledge at the edge between the known and the unknown. *Strategic Management J.* 43(7):1287–1318.

Felin T, Zenger TR (2017) The theory-based view: Economic actors as theorists. *Strategy Sci.* 2(4):258–271.

Gans JS (2023) Experimental choice and disruptive technologies. *Management Sci.* 69(11):7044–7058.

Garivier A, Lattimore T, Kaufmann E (2016) On explore-then-commit strategies. Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, eds. *30th Annual Conf. Neural Inform. Processing Systems 2016* (Neural Information Processing Systems, La Jolla, CA), 784–792.

Gilboa I, Minardi S, Samuelson L (2020) Theories and cases in decisions under uncertainty. *Games Econom. Behav.* 123:22–40.

Hanna R, Mullainathan S, Schwartzstein J (2014) Learning through noticing: Theory and evidence from a field experiment. *Quart. J. Econom.* 129(3):1311–1353.

Hansen LP, Marinacci M (2016) Ambiguity aversion and model misspecification: An economic perspective. *Statist. Sci.* 31(4): 511–515.

Hansen LP, Sargent TJ (2022) Structured ambiguity and model misspecification. *J. Econom. Theory* 199:105165.

Jaynes ET (1968) Prior probabilities. *IEEE Trans. Systems Sci. Cybernetics* 4(3):227–241.

Kamenica E, Gentzkow M (2011) Bayesian persuasion. *Amer. Econom. Rev.* 101(6):2590–2615.

Karni E (2022) A theory-based decision model. *J. Econom. Theory* 201:105444.

Karni E, Vierø ML (2013) "Reverse Bayesianism": A choice-based theory of growing awareness. *Amer. Econom. Rev.* 103(7):2790–2810.

Karni E, Vierø ML (2017) Awareness of unawareness: A theory of decision making in the face of ignorance. *J. Econom. Theory* 168(C):301–328.

Kass RE, Wasserman L (1996) The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* 91(435):1343–1370.

Klibanoff P, Marinacci M, Mukerji S (2005) A smooth model of decision making under ambiguity. *Econometrica* 73(6):1849–1892.

Marinacci M (2015) Model uncertainty. *J. Eur. Econom. Assoc.* 13(6): 1022–1100.

Milgrom P, Roberts P (1990) Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* 58(6):1255–1277.

Moscarini G, Smith L (2001) The optimal level of experimentation. *Econometrica* 69(6):1629–1644.

Nickerson J, Argyres N (2018) Strategizing before strategic decision making. *Strategy Sci.* 3(4):592–605.

Nickerson J, Zenger TR (2004) A knowledge-based theory of the firm: The problem-solving perspective. *Organ. Sci.* 15(6):617–632.

Ortoleva P (2012) Modeling the change of paradigm: Non-Bayesian reactions to unexpected news. *Amer. Econom. Rev.* 102(6):2410–2436.

Pearl J (1986) Fusion, propagation, and structuring in belief networks. *Artificial Intelligence* 29(2):241–288.

Pearl J (2009) *Causality* (Cambridge University Press, Cambridge, UK).

Pearl J, Mackenzie D (2018) *The Book of Why: The New Science of Cause and Effect* (Basic Books, New York).

Shelef O, Wuebker R, Barney JB (2024) Heisenberg effects in experiments on business ideas. *Acad. Management Rev.*, ePub ahead of print February 13, https://doi.org/10.5465/amr.2022.0051.

Thiel P, Masters B (2014) *Zero to One: Notes on Startups, or How to Build the Future* (Crown Business, New York).

Zellweger TM, Zenger TR (2023) Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. *Acad. Management Rev.* 48(3):379–408.

Arnaldo Camuffo, SM, MIT, PhD, University of Venice, is a professor of management and research fellow at ION Management Science Laboratory at Bocconi and Utah University. His website is https://faculty.unibocconi.eu/arnaldocamuffo/.

Alfonso Gambardella, PhD Economics, Stanford University, is a professor of corporate management and research fellow at ION Management Science at Bocconi and Utah University. His website is http://faculty.unibocconi.eu/alfonsogambardella/.

Andrea Pignataro is the founder and CEO of ION Group and a research fellow at ION Management Science Laboratory at Bocconi and Utah University.