This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Ehrig, Timo; Schmidt, Jens

Making biased but better predictions: The tradeoffs strategists face when they learn and use heuristics

Published in:
Strategic Organization

DOI:
[10.1177/1476127019869646](https://doi.org/10.1177/1476127019869646)

Published: 01/05/2021

Document Version
Peer reviewed version

Please cite the original version:

Ehrig, T., & Schmidt, J. (2021). Making biased but better predictions: The tradeoffs strategists face when they learn and use heuristics. *Strategic Organization*, 19(2), 263-284. <https://doi.org/10.1177/1476127019869646>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Making biased but better predictions: The tradeoffs strategists face when they learn and use heuristics*

Timo Ehrig

University of Southern Denmark, Odense, Denmark
Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
ehrig@mis.mpg.de

Jens Schmidt

Aalto University, Helsinki, Finland
jens.schmidt@aalto.fi

June 2019

Forthcoming in *Strategic Organization*

Abstract

The heuristics strategists *use* to make predictions about key decision variables are often *learned* from only a small sample of observations, which leads to a risk of inappropriate generalization when strategists misjudge regularities. Building on the statistical learning literature we show how strategists can mitigate this risk. Strategies to learn heuristics that accept a *bias*, that is, a systematic deviation of predictions from actual outcomes can outperform unbiased strategies because they can reduce the *variance* component of prediction error: The degree to which random fluctuations in observational data are inappropriately generalized. We demonstrate how strategists who are aware of the trade-off between bias and variance can learn heuristics more effectively if they are also aware of the relevant characteristics of their learning environment. We discuss the implications of our results for our understanding of heuristics, (dynamic) capabilities and managerial cognitive capabilities, and we outline opportunities for empirical work.

Keywords: Heuristics; biases; learning; simple rules; strategic decision-making; inappropriate generalization

*The authors would like to thank Pantelis P. Analytis, Eckehard Olbrich, Mikko Rönkkö and Jan Woike as well as our editor Ann Langley and three anonymous reviewers for their highly valuable comments.

1 Introduction

A strategist confronted with a decision of strategic importance –such as whether to allocate resources to a new venture, to go forward with a proposed acquisition, or to hire a candidate for a leadership position– must assess the key parameters that affect the outcome of the decision. For example, the expected performance of the new venture, of the acquired firm or of the candidate hired are all important variables that must be predicted before making a decision based on the information that is available at that point. Rumelt (1979) and Nelson and Winter (1982) early on suggested that strategists should use simple heuristics to make predictions about such decision variables. More recently, strategy scholars have provided empirical evidence that strategists indeed use heuristics as “simple rules” (e.g., Bingham and Eisenhardt, 2011; Maitland and Sammartino, 2015), that these are often learned from and honed with experience (e.g., Bingham and Eisenhardt, 2011; Bingham, Howell and Ott, 2019), and that their use often leads to better decisions (e.g., Astebro and Elhedhli, 2006; Luan, Reb and Gigerenzer, 2019).

While strategy scholars have noted that the heuristics strategists use in a given decision situation have often been developed through a learning process (e.g., Bingham and Eisenhardt, 2011), making the conceptual distinction between learning a heuristic on the one hand and using it on the other hand explicit has consequences that prior work on heuristics in a strategy context has not examined. Because strategists encounter typical decision problems only infrequently they have to learn the relationship between available information and decision variables for a given type of decision from only a small sample of observations (Zollo and Winter, 2002; Zollo, 2009). As a consequence, when learning heuristics that capture this relationship strategists are at risk of making “inappropriate generalizations” (Haleblian and Finkelstein, 1999) –seeing regularities where none exist or misjudging those that do exist (Gilovich, Vallone and Tversky, 1985; Taleb, 2001).

In this article we build on the literature on statistical learning (Hastie, Tibshirani and Friedman, 2013) to formally analyze how accepting *biases* in learning and using heuristics can mitigate the risk of making inappropriate generalizations and thus lead to *more* accurate predictions in typical strategy contexts (Bingham and Eisenhardt, 2014; Loock and Hinnen, 2015; Vuori and Vuori, 2014). In doing so we make use of the work of Gigerenzer and colleagues from the psychology literature and their research on “fast and frugal” heuristics (Gigerenzer and Brighton, 2009; Gigerenzer and

Gaissmaier, 2011), whose core argument is that heuristics will be *biased* when they ignore available information but surprisingly –rather than therefore being inferior to more exhaustive prediction strategies that do not ignore information– the very fact that they ignore information allows for *more* accurate predictions even if the information is freely available and can be processed at no cost. We thus define a heuristic as a prediction strategy that accepts a bias (e.g., by ignoring available information) with the goal of making decisions more accurately than unbiased prediction strategies¹. We then build on a key premise of the fast and frugal program, namely that the accuracy of a prediction from any prediction strategy depends on characteristics of the decision situation or environment (which is a point made by Herbert Simon long ago; e.g., Simon, 1956). That is, a heuristic is “ecologically rational” (Gigerenzer and Selten, 2002; Goldstein and Gigerenzer, 2002) in a given type of environment if it leads to a more accurate prediction than alternative prediction strategies in that environment. In our analysis we make a fundamental distinction between environments that are stable (where the same type of decision problem is repeatedly encountered and the regularities can be learned; e.g., Bingham and Eisenhardt, 2011) and those that are changing (where a new decision problem is different from yet similar to one which has been previously encountered; e.g., Eisenhardt and Martin, 2000). Based on this distinction, we demonstrate that biased heuristics can improve prediction accuracy in the following two ways.

First, consider a strategist in a stable environment, who evaluates acquisition targets and must predict how difficult it will be to integrate an acquired firm’s technology with the acquiring firm’s technology. He has observed three acquisitions of small technology firms. In two of these, the CEO of the acquired firm had an MBA and integrating the acquired firm’s technology was easy. In the third one the CEO had an engineering degree, and technology integration proved very difficult. Based on these observations, should the strategist conclude that CEO educational background is generally predictive of integration success, or would that be an instance of inappropriate generalization?

Ignoring it will systematically *bias* predictions about ease of technology integration if CEO educational background and ease of integration are in fact correlated. Say that on average tech-

¹This definition is a generalization of the idea that “biased minds make better inferences” (Gigerenzer and Brighton, 2008: p.107). Gigerenzer’s definition is more restricted in that it explicitly includes that heuristics ignore information. We use the more general definition because it allows us to also account for heuristics that do not ignore biased experience, which is relevant in changing environments.

nology integration is easier if CEOs have an MBA than if CEOs have an engineering degree. Then ignoring CEO background will lead to too optimistic predictions if the CEO has an engineering degree and to too pessimistic predictions if the CEO has an MBA. But the strategist does not know the true relationship and therefore needs to learn it from her observations. Because easiness of integration will also be determined by a number of other factors, the strategist –given a sample of just three observations– may well (simply by chance) observe two cases of successful integration with a CEO who had an engineering degree and one case of integration failure and a CEO with MBA background. Given such a pattern, concluding that integration will be easier for firms with an engineering CEO would be an instance of inappropriate generalization. The *variance* component of prediction error describes the average error a strategist makes as a result of such inappropriate generalization.

Generally there is a bias-variance tradeoff (Hastie et al., 2013): Determining which observable attribute to include in a prediction strategy requires a strategist to weigh the risk of inappropriate generalization against the bias from ignoring an attribute that actually matters. As we show, the question whether the strategist should generalize about a relationship he observes in a small sample depends on characteristics of the learning environment. For example, the risk of inappropriate generalization is generally the larger the fewer observations are available and the greater the variability between situations of the same type (which is the degree to which the decision variable is determined by factors unique to the specific situation).

Second, assume that the same strategist is exposed to a changing environment: the firm he is working for has recently started to acquire much larger firms. In the meanwhile he has seen dozens of acquisitions of small technology firms and has (correctly) learned that on average technology integration will be easier for small technology firms whose CEO has an MBA background. After having observed three large acquisitions, a similar pattern as above emerges: In the two acquisitions where the CEO had an engineering background technology integration was difficult, whereas in the third one where the CEO had an MBA integrating the acquired firm’s technology was much easier. The strategist could conclude that it may be better to ignore CEOs’ educational background because he has too few observations to reliably generalize the observed relationship for large firms. However, he could also transfer his ample prior experience with small firms. Doing so will, however, lead to a bias if there are systematic differences between technology integration in small and large firms

(Ellis, Reus, Lamont and Ranft, 2011).

This example shows that a similar tradeoff exists in changing environments when a strategist has only few observations from the new environment but ample experience from the previous environment. Specifically, transferring earlier experience also has two effects: On the one hand it leads to a systematic bias in predictions due to the dissimilarity between contexts (e.g., March, 1991). On the other hand, it can also reduce the risk of inappropriate generalization because it increases the number of observations from which the relationship between attributes and the decision variable has been learned. As we show, a “transfer heuristic” that makes predictions by combining what the strategist learns from the small number of observations in the changed environment with his experience from the previous environment can address this bias-variance tradeoff. Interestingly, the bias-variance tradeoff in changing environments has the opposite direction as the one in stable environments: making use of biased experience from the previous context can improve prediction accuracy while ignoring it can increase the risk of making inappropriate generalizations.

We build on the statistical learning literature (e.g., Hastie et al., 2013) to formally examine how characteristics of learning environments affect the prediction accuracy of heuristics that strategists learn from small samples in both stable and changing environments. Formally, we treat the learning process as the process of determining weights for different observable attributes that strategists will use in making predictions (in our above example, the educational background of the CEO is one such attribute). Ignoring an attribute means giving zero weight to it and transferring experience means calculating a weighted average of the attribute weights determined from observations in the previous environment and the weights determined from observations in the changed environment. In addition to distinguishing between stable and changing environments, we also examine several characteristics of learning environments that have been analyzed in the fast and frugal heuristics literature (Gigerenzer and Gaissmaier, 2011).

A key insight of our article is that there is a conceptual difference between rules for learning heuristics on the one hand and heuristics as prediction strategies on the other hand. Learning rules include, for example, rules to decide whether to generalize (e.g., *do not generalize after only three observations*) or rules to determine how much weight to give to prior experience that is transferred to a new context (e.g., *weigh prior experience highly if the new environment is very similar to the previous one*), whereas heuristics are the resulting prediction strategies (e.g., *if the CEO has an*

MBA technology integration will be easier).

The key contribution our study makes is to show how such learning rules that capture the key characteristics of the learning environment affect the prediction accuracy of the heuristics strategists learn from experience. An implication of our findings is that to learn and use heuristics effectively strategists must correctly assess the relevant environmental characteristics. Therefore strategists who are aware of these characteristics will learn heuristics that will allow them to make more accurate predictions. This, in turn, implies that knowledge about environmental characteristics is an important but understudied aspect of managerial and dynamic capabilities (Helfat and Peteraf, 2015). In addition, our findings also raise the important question how fungible and thus transferable heuristics are across different types of situations (Levinthal and Wu, 2010).

In the following, we formalize our argument. We then use this formal approach to explain the bias-variance tradeoff in both stable and changing environments and show how characteristics of learning environments affect the prediction accuracy of the heuristics that are learned from small samples. We then discuss the implications of our arguments and provide guidance for further work on heuristics in a strategy context.

2 Heuristics, information and learning environments

2.1 The strategist’s prediction task and available information

Our starting point is a strategist who must evaluate a decision in terms of a key decision variable (which we call the “outcome”). In our introductory example ease of technology integration in an acquisition is such a key variable. In the context of internationalization (Bingham, Eisenhardt and Furr, 2007) a key variable could be some measure of success of a market entry decision under consideration (e.g., market share or total sales). For predicting the outcome the strategist has available a number of observable “attributes”. For example, for a market entry decision the observable attributes include, among many others, market language or country GDP. In addition to the specific decision problem at hand, the strategist has a number of prior observations of the same “type” of decision problem from which he can learn the relationship between attributes and outcome. Being of the same type not only means that they are about the same task or process (e.g., acquisition, new product development, new market entry) but we assume that they also have the same observable

attributes and that the outcome to be predicted is also observable for all instances.

We can formalize this situation as a discrete stochastic process. Formally, at each time t nature draws a decision problem, which is characterised by p observable attributes $x_t = (x_{t,1}, x_{t,2} \dots x_{t,p})$ and an outcome y_t (all of which have distributions with finite variance). The attributes may be correlated (but we rule out perfect correlations). For decision type i the true relationship f_i between attributes and outcomes is given as a stochastic process:

$$y_t = f_i(x_t) = \beta_i x_t + \epsilon \quad (1)$$

where both x_t and ϵ are i.i.d. distributed for all $t \in \{1, \dots, n\}$ and ϵ is Gaussian with a mean of zero, an expected value of zero given any values of the p attributes and constant variance of σ^2 (with $\sigma^2 \geq 0$). The vector β_i encodes the linear weights of the p attribute values. In particular, β_i specifies the regularities for decision type i and thus the true relationship between attributes and the decision variable the strategist is interested in predicting. In addition to these regularities there is also variability across the same type of situation (the degree of variability is given by σ^2), which may be due to factors that are either unobservable or idiosyncratic to the specific instance of the decision situation.

The strategist's task is to predict the outcome of the decision variable y_0 for a new decision problem of the same type as accurately as possible given the available information. Throughout the paper, we assume that at time t the strategist can observe the attributes x_t but not the outcome y_t , which he only observes at time $t + 1$. We assume that all attributes are perfectly observable. The information available to the strategist is thus given by the set of n prior observations of the type of decision problem, with p observable attributes and the outcome y_t for each, as well as x_0 as the values of the p attributes for the decision problem at hand.

2.2 Heuristics and learning rules

As defined above, a heuristic is a prediction strategy that accepts a bias (e.g., by ignoring information) with the goal of making decisions more accurately than unbiased prediction strategies. If the true relationship f_i were known to the strategist, he would not need to learn it and thus there would be no risk of inappropriate generalization. He could simply use the observable attributes

x_0 to make his prediction by using β_i in equation (1), which would tell him how each of the p attributes contributes to the outcome. Denoting his prediction of the outcome by $\hat{y}_0$ and the observable attribute values of the new decision problem by x_0 , his prediction would then be given by $\hat{y}_0 = \beta_i x_0$. Given our assumptions, this prediction would be unbiased but (due to variability across situations of the same type) it would be subject to a prediction error ϵ .

However, we assume that the strategist only knows the functional form of f_i but not the true parameters β_i . So he must learn this relationship from the available information, which introduces an additional source of error in making predictions. We denote the estimate of the relationship that he uses in making his prediction by $\hat{\beta}_i$ (note the $\hat{\cdot}$ sign over β : β_i is the true relationship between attributes and outcome while $\hat{\beta}_i$ is the strategist's estimate of this relationship). The strategist's prediction strategy is thus given by $\hat{y}_0 := \hat{\beta}_i x_0$.

A prediction strategy is the result of applying a rule to learn $\hat{\beta}_i$. For example, the strategist could decide to ignore a subset of the available attributes (which means that the respective components in $\hat{\beta}_i$ are set to zero; Davis-Stober, Dana and Budescu, 2010). Given our assumptions, a heuristic that results from applying a learning rule that constrains the estimate $\hat{\beta}_i$ in such a way will lead to a systematic bias when it is used for making predictions.² However, the fact that the strategist has to learn $\hat{\beta}_i$ from available observations constitutes a source of error in addition to any systematic bias, and it is for this reason that biased heuristics can lead to better predictions, as we will explain below.

2.3 Characteristics of learning environments

Whether and by how much a biased heuristic will improve prediction accuracy depends on characteristics of the learning environment. Gigerenzer and Gaissmaier (2011) list four characteristics of learning environments that have been identified in the fast and frugal program, which influence the potential superiority of biased heuristics: the number of observations of prior instances available to the strategist n , the magnitude of variability across the same type of decision σ^2 , the relative

²Our model fulfills the Gauss-Markov assumptions and thus estimating $\hat{\beta}_i$ using all available information is equivalent to the ordinary least squares (OLS) estimator $\hat{\beta}_{i,OLS}$ in multivariate linear regression (Wooldridge, 2013). Because $\hat{\beta}_{i,OLS}$ is an unbiased estimator of the true β_i our model is well-suited to illustrate the bias-variance tradeoff. Other estimators $\hat{\beta}_i$ (for example, one where only one attribute is used and all others are set to zero) could be biased but could have lower variance. Importantly, our model also provides for a stylized representation of several particular types of heuristics that have been extensively studied in the psychology literature. See Appendix 1 for details.

importance of particular attributes for explaining the outcome (which is contained in the individual components of β_i), and their redundancy in terms of correlations among the attributes.

However, there is also one aspect in which a typical strategy context can differ from the types environments studied by Gigerenzer and colleagues. Specifically, the basic argument in the fast and frugal research program why heuristics work is that they are effective at exploiting regularities across instances of the same type of decision problem (Goldstein and Gigerenzer, 1999; Gigerenzer, 2008). To take an example for a regularity from Bingham and Eisenhardt (2011), in the context of internationalization using language spoken in a country as a cue to determine its attractiveness can only work if the relationship between language and attractiveness is the same across countries. On the other hand, however, it has been argued that a key characteristic of the environments faced by strategists is that they are often changing (Vuori and Vuori, 2014). Thus, one needs to account for whether and to what extent regularities persist across instances of a decision problem encountered by a strategist.

A stylized distinction to account for differences in the persistence of regularities is to distinguish between two basic types of environments: those in which regularities persist across instances of the same type of decision problem (stable environments) and those in which the regularities are subject to a one-time change (changing environments). Clearly both will be relevant in strategic decision contexts. For example, there are likely regularities across decision situations that require evaluating proposals for new ventures or for evaluating acquisition targets. On the other hand, the entry of a competitor may change the regularities for new product development (as has happened when Apple launched the iPhone into the mobile handset market), or a restaurant chain that needs to decide about opening restaurants in a new country may be confronted with regularities that are different from yet similar to those in its home country.

We formalize this distinction as follows. In a stable environment, the relationship f_i between attributes and outcomes (given by equation (1)) is the same for all observations as well as the new decision problem. In a changing environment, there is a one-time change in the regularities (but not in the distribution of the attributes and their correlations) after the first m decision problems have been encountered. Specifically, the first $t = 1, \dots, m$ decision problem are drawn from f_1 and the decision problems at $t = m + 1, \dots, m + n$ decision problems are drawn from f_2 (with $\beta_1 \neq \beta_2$). The magnitude of change is given by the Euclidean distance $\|\beta_2 - \beta_1\| > 0$. We also assume that

σ^2 is constant for $t = 1, \dots, m+n$. Finally, the decision problem for which the strategist must make a prediction is drawn from f_2 .

*** INSERT TABLE 1 ABOUT HERE ***

In the following two sections we explain –for both stable and changing environments– the respective bias-variance tradeoff and how the relevant environmental characteristics enable strategists to make more accurate predictions. For stable environments we examine the four characteristics from Gigerenzer and Gaissmaier (2011) listed above and for changing environments we examine n and m , respectively, σ^2 as well as the magnitude of change $\|\beta_2 - \beta_1\|$. Table 1 provides an overview of these characteristics and their effect on learning heuristics. In explaining the results, we open up the underlying mechanisms and provide intuition for the results. The formal results can be found in Appendix 1 (for stable environment) and Appendix 2 (for changing environments).

3 Heuristics and learning rules in stable environments

Typical decision problems faced by strategists in which regularities are stable and thus can be reliably learned by using appropriate learning rules concern, among others, evaluating proposed projects and new ventures as well as allocating resources to them (Bower, 1970; Burgelman, 1983), selecting markets for entry as well as mode of entry (Barkema and Vermeulen, 1998; Bingham and Eisenhardt, 2011), acquisition target evaluation and selection (Capron and Shen, 2007), or evaluating candidates for leadership positions (Barney and Wright, 1998). The “flow of opportunities” in the work of Eisenhardt and colleagues (e.g., Eisenhardt, Bingham and Furr, 2010) can also be seen as repeated exposure to the same type of decision problems.

3.1 The bias-variance trade-off in stable environments

Consider the following stylized example to illustrate the challenge strategists face in learning about the relationship between attributes and outcome in a stable environment due to the bias-variance tradeoff. Suppose that a company’s strategist must decide whether or not to offer the position of head of a new venture to a manager who has recently been hired from a competitor to lead another project for the company. The outcome variable to be predicted is the performance of the manager

as head of the venture. To keep it simple, suppose the strategist has only two pieces of information available to predict the outcome: the manager’s performance on the project that he had been hired to lead and his performance on his last project working for the competitor. Assume that both can be perfectly observed and that both attributes are informative about the performance of the employee: on average, higher performance on in-house projects as well as higher performance on projects at other firms are both associated with higher performance as head of a new venture. Also assume that the former is a stronger predictor of new venture performance than the latter.

The strategist must now consider which of these attributes to use and which to ignore. Say he decides to only use performance on the in-house project in predicting the outcome. This is an instance of a “one good reason” heuristic, which uses the one attribute that best predicts the outcome of the target variable while ignoring all other attributes (Gigerenzer and Gaissmaier, 2011). It has been shown that this type of heuristic can be superior in correctly predicting outcomes compared to multiple linear regression models that use all available information, and it can even outperform complex non-linear prediction strategies (Czerlinski, Gigerenzer, and Goldstein, 1999). Clearly, this heuristic leads to a systematic bias in predicting the manager’s performance compared to using both pieces of information. For example, if the candidate’s performance at the competitor would be very high, ignoring it would lead to underestimating expected performance on the new venture.

On the other hand, the fact that the strategist doesn’t know whether and how much specific attributes explain an outcome of interest introduces a second component of prediction error. Say the strategist has observed three prior instances of this type of hiring decision and must learn the relationship between manager attributes and outcome from these three instances. Clearly, any generalization from just three observations about the relationship between manager attributes and outcome will be subject to variation (i.e., for any set of three different prior observations the strategist will likely arrive at a different estimate for the relationship). For example, by sheer coincidence among the three instances there may have been a negative association between performance at the competitor and performance as head of new venture even though the true relationship is positive. Ignoring an attribute in making a prediction will lower the risk of inappropriate generalization.

In deciding which attributes to pay attention to and which to ignore the strategist thus faces a tradeoff. To formalize this tradeoff, the prediction accuracy of any prediction strategy is given

by the expected squared prediction error (ESPE), which describes the expected difference between predictions and actual outcomes (see, e.g., Hastie et al., 2013, p.223). It is given as follows (in addition to bias and variance, there is a third source of prediction error because the new decision situation for which the outcome is predicted has idiosyncratic characteristics that are not captured by the observable attributes):

$$ESPE = Bias^2 + Variance + Irreducible\ error \quad (2)$$

In our example, ignoring the attribute of performance on projects at other firms can be justified if the systematic bias from ignoring it is smaller than the prediction error due to inappropriate generalization. In addition, the decision to include performance on the in-house project but excluding the other attribute may be justified by the former being a stronger predictor than the latter. More generally, to decide which attributes to include and which to ignore the strategist must assess how reliably he can learn about whether and how much each of the attributes matters to the outcome. This, in turn, depends on characteristics of the learning environment.

3.2 Learning rules and characteristics of stable environments

To show how characteristics of the learning environment affect how strategists can effectively learn heuristics, we examine the four characteristics of stable environments listed by Gigerenzer and Gaissmaier (2011). Appendix 1 provides an overview of the literature that contains the formal results.

3.2.1 Number of available prior observations

Strategic decision situations are often relatively novel or strategists encounter specific decision situations (like acquisitions) only infrequently (Zollo and Winter, 2002; Zollo, 2009). Therefore, the number of observations of prior instances of the specific type of situation is often relatively small. As any student of statistics knows, the smaller the sample size from which predictions are made, the larger the chance of over- or underestimating the true relationship from the specific observed values (e.g., Wooldridge, 2013). This effect also applies to strategists: attempts to identify which attributes matter and how they relate to an outcome from a small number of available observations are fraught with a risk of making inappropriate generalizations.

In such situations, learning rules that ignore some or most of the available attributes lead to simple heuristics that have higher prediction accuracy than more comprehensive prediction strategies, because they reduce the risk of making inappropriate generalizations (Gigerenzer and Goldstein, 1996; Hogarth and Karelaia, 2005). As an example, consider a company that seeks to select markets for entry and that has already entered a few markets. In selecting additional markets, the company may do better by applying a learning rule that identifies the one or two attributes that are most strongly associated with success in prior entries than by performing in-depth studies of the markets in question and, for example, using a scoring model with a large number of attributes (e.g., Calantone, Di Benedetto and Schmidt, 1999).

Bingham and Halebian (2012) provide evidence for a specific learning rule that strategists seem to actually use. They show that whether an attribute is included in a heuristic is affected by whether managers agree or disagree about its relevance. In the light of the bias-variance tradeoff, using mutual agreement to infer that an attribute matters in a specific type of situation can be seen as a learning rule that reduces the risk of inappropriate generalization by increasing the sample size from which strategists learn about what is important and what is not.

3.2.2 Variability

Strategic decision situations differ in their degree of variability across instances (Davis, Eisenhardt and Bingham, 2009), where high variability means that the outcome is to a large extent determined by factors unique to the specific situation that do not generalize to other situations of the same type. For example, new product launches in the fashion industry may be high in variability whereas new product launches in the chemicals industry may exhibit relatively low variability. A high degree of variability makes it difficult to discern the effect of the observable attributes that are stable over time from the effect of the idiosyncratic characteristics of the situation (Halebian and Finkelstein, 1999). Therefore, variability exacerbates the problem of learning from small samples noted above and thus increases the risk of inappropriate generalization. As a consequence a strategist working in the fashion industry will have a higher risk of inappropriate generalizations after having observed ten new product launches than a strategist working in the chemicals industry, who has also observed ten new product launches. Learning rules that account for the degree of variability can lead to heuristics that have higher prediction accuracy.

3.2.3 Relative importance of attributes and redundancy among attributes

Among the observable attributes, some are more important and thus more informative about the outcome than others. The less important an attribute, the smaller the increase in bias from ignoring it relative to the reduction in the variance component of prediction error. This means that the distribution of the importance of attributes affects which combination of attributes minimizes expected prediction error. An important aspect of learning rules is therefore to identify the relative importance of attributes and, in particular, to decide which of them can be safely ignored.

If the distribution of attributes is highly skewed, heuristics that ignore less important attributes will be ecologically rational (Goldstein and Gigerenzer, 1996; Hogarth and Karelaia, 2005). In particular when attributes are non-compensatory, which means that one attribute is more informative about the outcome than all other attributes combined, using only this one attribute can lead to lower expected prediction error than using multiple attributes (Goldstein and Gigerenzer, 1996). In addition, the more skewed the distribution of the importance of attributes, the easier it is to rank or prioritize them (Gigerenzer and Gaissmaier, 2011) and, as a consequence, the more easily the most important attributes can be reliably identified in small samples.

If two or more attributes are highly correlated they largely contain the same information about the outcome. They are thus (imperfect) substitutes in predicting the outcome and at least partly redundant (Bingham and Eisenhardt, 2014; Hogarth and Karelaia, 2005). High correlations among attributes makes it more difficult to reliably learn the relationship between these attributes and the outcome because it makes it difficult to discern their individual effects on the outcome. Therefore ignoring one of two correlated attributes can lower the risk of inappropriate generalization. As a consequence an important aspect of learning rules is to assess whether any two attributes are highly correlated so that one of them can be safely ignored.

4 Heuristics and learning rules in changing environments

Strategists must also often predict decision-critical variables in an environment in which the regularities are undergoing change. For example, a stable environment may be exposed to a discrete “shock” that leads to a change in regularities for a given type of decision problem. Take the example of Apple’s launch of the iPhone, which changed the characteristics that make a mobile phone

successful. Similarly, strategists may encounter a new type of decision situation that is similar to one they have already frequently encountered in the past. As an example, consider Cisco, which in the 1990s grew through making a large number of acquisitions of small companies, but later switched to also acquiring large companies. In both cases, the strategist is faced with the question whether and to what extent what he has learned about the relationship between available information and the decision variable in the original context is transferable to the new situation, whether adjustments are necessary, or whether it is better to discard prior experience altogether (Finkelstein and Halebian, 2002).

4.1 The bias-variance tradeoff in changing environments

We examine the simplest possible setup of a changing environment, namely one in which regularities are stable over time except for a one-time change after which the regularities are different from the situation before the change. While this is a highly stylized type of change, it allows to nicely illustrate the bias-variance tradeoff in changing environments.

As above, the strategist wants to make an as accurate prediction as possible. Assume he knows that the environment has changed and that he has a small number of observations from the environment after the change has occurred. Intuitively, to make accurate predictions he should only use the observations from the environment after the change to learn about the relationship between attributes and outcome and ignore his experience from before the change, because the latter is biased with respect to the true relationship after the change (March, 1991). For example, Bettis and Prahalad (1995) have argued that firms and strategists need to unlearn or forget outdated lessons once the environment has changed. Of course the magnitude of the bias depends on the degree of similarity between the relationship between attributes and outcomes before and after the change (what we call the magnitude of change). Thus, outdated experience or experience from a different but similar decision situation may nevertheless contain useful lessons for the situation at hand (Gentner, 1983), and the more so the smaller the magnitude of change or the larger the degree of similarity (Finkelstein and Halebian, 2002).

This does not, however, explain why using outdated experience can lead to *more* accurate predictions but only that the bias induced by doing so may sometimes be relatively small (but a bias nonetheless). What is missing –as in stable environments– is to account for the variance

component of prediction error when strategists have to learn the relationship between attributes and outcome. Specifically, ignoring experience from before the change will eliminate systematic bias in learning this relationship but at the same time reduce the total number of observations used in learning. As a consequence of ignoring this experience the variance component of prediction error may be large. That is, ignoring experience from before the change will increase the risk of inappropriate generalization.

Taken together, this means that transferring outdated experience to a changed or different environment has two effects: It leads to a systematic bias in making predictions but can reduce the variance component of prediction error. Thus there is also a bias-variance tradeoff in changing environments: Strategists have to weigh the risk of making wrong predictions because their prior experience doesn't apply to the changed or new environment against the risk of making wrong predictions from inappropriately generalizing from a small sample of observations in the changed or new environment. This bias-variance tradeoff works in the opposite direction to the one in stable environments: Ignoring prior (outdated) experience will lower the bias but increase the variance component of prediction error. As a consequence, *not* ignoring biased experience can lower the risk of inappropriate generalization. The implication is that, counter-intuitively, it may be better to use experience that is known to be biased if doing so will reduce the variance component of prediction error by more than it increases the bias.

4.2 Learning rules, experience transfer and characteristics of changing environments

4.2.1 A transfer heuristic

As in stable environments, the bias-variance tradeoff in changing environments provides a general explanation why it may be beneficial to use a biased heuristic to predict an outcome. Likewise, whether this is the case depends on characteristics of the learning environment. We thus examine this question with respect to three specific characteristics of changing environments: the number of observations m (before the change) and n (after the change), the magnitude of variability across the same types of decision σ^2 , and the magnitude of change $\|\beta_2 - \beta_1\|$. To do so, we define and analyze a “transfer heuristic” as a prediction strategy that uses the lessons from the situation before

the change to adjust predictions that would otherwise be based on generalizing only from a small number of observations made after the change. While this allows us to systematically examine how environmental characteristics affect the bias-variance tradeoff in changing environments, we also note that the bias-variance tradeoff is generic and applies to all prediction strategies that involve deciding about whether and to what extent to transfer what has been learned in one context to another one.

We formalize the transfer heuristic as follows. Assume the environment has changed from f_1 to f_2 . The strategist has m observations from the environment before the change (which are from f_1), which he uses to estimate the relationship between attributes and outcomes β_1 , and n observations from the environment after change (which are from f_2), which he uses to estimate β_2 . Also assume that $m > n$, so on average the strategist will be able to estimate β_1 more accurately than β_2 . The transfer heuristic uses a “transfer rate” b (with $0 \leq b \leq 1$) for deciding how much weight the strategist should give to his outdated (and thus biased) experience in predicting an outcome in a situation after a change has occurred. It is thus a simple heuristics that uses only one parameter b . Formally, the transfer heuristic uses a weighted average of the estimate of the relationship between attributes and outcomes before the change (relative weight $1 - b$) and after it has occurred (relative weight b). If $b = 1$, the strategist gives full weight to new information and ignores all prior experience. If $b = 0$, the strategist only uses his prior experience and ignores his observations in the new environment. Formally, the transfer heuristic predicts the outcome variable as follows:

$$\hat{y}_l(x_0) = (1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0 \quad (3)$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the strategist’s estimates of the relationship between attributes and outcomes before and after the change, respectively, and x_0 are the values of the observable attributes in the decision situation for which he must predict the outcome. That is, $\hat{\beta}_1$ is the prediction strategy that has been learned in the previous environment and $\hat{\beta}_2$ is the prediction strategy learned in the environment after the change, which the strategist would use if he would choose to ignore his prior experience. Given our assumption that f_1 and f_2 have the same fundamental structure, $\|\beta_1 - \beta_2\|$ is a measure for their (dis)similarity (Gentner, 1983; Finkelstein and Halebian, 2002).

4.2.2 Learning rules in changing environments

In our stylized model, after a change has occurred, the strategist essentially finds himself in a stable environment except that he can now draw on experience from the situation before the change has occurred. Therefore a key aspect of learning in a changed environment or in a new environment that is similar to one that one is familiar with is to decide to what extent to start learning afresh based on using observations from the new or changed environment and to what extent to rely on and transfer prior experience.

In the context of our model, this means that a strategist who is confronted with a new decision problem that he knows is drawn from f_2 must set b so to minimize expected prediction error. Using the transfer heuristic leads to a biased prediction of the outcome if the strategist chooses $b < 1$ (but is unbiased for $b = 1$, which means that prior experience is ignored). However, choosing $b < 1$ can lead to more accurate predictions than $b = 1$ if the number of observations in a new environment n is low and/or variability σ^2 is high (for a given m). In this case using prior experience will lead to a particularly large reduction in the variance component of prediction error. Likewise, the higher the degree of similarity (the smaller $\|\beta_1 - \beta_2\|$) the lower the bias from using outdated experience. In Appendix 2, we formally derive results about the upper bound of the optimal transfer rate b given the three characteristics of changing decision environments as parameters of the model.

For learning rules that help decide about experience transfer it is important to identify situations in which strategists will want to give strong weight to prior lessons learned (and thus discount what they have learned in the environment after change). For example, in the context of a U.S. firm's entry into Israel Szulanski and Jensen (2006) found that initially selecting a higher transfer rate was associated with higher performance. Intel's famous "copy exactly" rule –which rules out that any learning about the new environment will be used in setting up new facilities– can be seen as a learning rule that is effective because it avoids the risk of misjudging what matters in the new environment. As observations in the changed or new environment accumulate, the strategist's prediction error can be reduced by gradually reducing the transfer rate and increasingly relying on the learning from the new environment. Similarly, a higher transfer rate can increase the accuracy of predictions in the new environment if variability is high, because higher variability increases the risk of inappropriate generalization.

4.2.3 Variability and change as distinct environmental characteristics

In our model, variability and change are two distinct parameters. The strategy literature often confounds them in verbal treatments (e.g., Brown and Eisenhardt, 1997; Eisenhardt and Martin, 2000), whereas formal treatments typically assume a stable environment. For example, the environment in the simulation study by Davis et al. (2009) is stable, because the underlying distribution from which the opportunities are drawn does not change over time, but exhibits variability (thus, in principle, the parameters of the process that generates the opportunities could be learned by the agents, even though Davis et al. (2009) do not model this).

Our analysis shows that distinguishing variability from change is important because they impose two different types of constraints on how strategists may be able to deal with situations in which they have to learn the relationship between available information and the decision variable they must predict. While variability makes learning difficult because it increases the risk of making inappropriate generalizations, change means that what has been learned earlier will not necessarily be applicable anymore. This has consequences for the types of learning rules strategists may best use to reduce prediction errors: In stable environments, higher variability increases the benefits of simple heuristics that ignore observable attributes. In changing environments, higher variability increases the optimal rate at which the lessons learned in prior environments is transferred to a similar, new context.

5 Discussion

In the following we discuss implications of our findings for research on heuristics, managerial cognition and (dynamic) capabilities, which we discuss in the following. We conclude by highlighting limitations of our approach.

5.1 Heuristics, simple rules and learning rules

The distinction between learning and using heuristics allows for conceptual clarity and thus aids in interpreting results of empirical research on heuristics in a strategy context (Vuori and Vuori, 2014). For example, while Bingham and Eisenhardt (2011) argue that “firms learn heuristics” (p.37) they only describe simple rules as the outcome of a learning process. More recent research

by Bingham and colleagues (Bingham and Halebian, 2012; Bingham et al., 2019), on the other hand, sheds light on the underlying learning process. Prior research on heuristics also provides hints that strategists indeed use learning rules to reduce the risk of inappropriate generalization. As stated earlier, an example can be found in the work of Bingham and Halebian (2012), who show that managers decide whether an attribute is included in a heuristic based on whether managers agree or disagree about its relevance. Other learning rules (for example concerning under which conditions to generalize or to transfer prior experience, or to select one of multiple observable attributes, or how to weigh them) may exist and be frequently used by strategists.

However, our finding that the learning process exposes strategists to a bias-variance tradeoff has yet to be accounted for in empirical research on heuristics. In particular, the question whether and to what extent strategists take into account that they incur a risk of inappropriately generalizing from a small number of observations, and what mechanisms they use to mitigate this risk, remains largely unanswered. The implication for empirical research on heuristics is that –in addition to identifying and documenting the heuristics strategists use (e.g., Bingham and Eisenhardt, 2011)– researchers also need to dig deeper into the underlying learning processes and seek to systematically identify the learning rules which strategists use to learn heuristics in the face of the bias-variance tradeoff. Here our work provides the basis for examining whether and why such learning rules and the resulting heuristics can be considered ecologically rational.

5.2 Heuristics, learning rules and dynamic capabilities

While heuristics as simple rules (as outcome of the learning process) can be considered an aspect of a capability, the underlying learning processes (including learning rules and decisions about transfer of prior experience) can be considered an aspect of dynamic capabilities (Winter, 2003). Bingham et al. (2019) provide initial empirical evidence for this link by showing how strategists develop and adapt heuristics over time and turn them into capabilities. However, the distinction between learning and using heuristics also has broader implications for the link between heuristics and dynamic capabilities at the level of the organization and can thus shed light on how learning rules can serve as microfoundations for dynamic capabilities (Felin and Foss, 2005; Di Stefano, Peteraf and Verona, 2014).

Because their applicability depends on characteristics of the environment, effectively applying

learning rules and transferring prior experience from one context to another, related one requires an understanding and a correct assessment of the relevant environmental characteristics. For example, situations with high variability require having more observations or weighing prior experience more strongly when transferring experience than situations with low variability. Similarly, deciding about transfer of experience requires an assessment of the similarity between contexts (Finkelstein and Halebian, 2002). As a consequence, knowledge about environmental characteristics may be an important aspect of dynamic capabilities. It is a kind of “meta-knowledge” (Foss and Jensen, 2018), and examining its role in learning heuristics and transferring prior experience is an important avenue for further research.

On the other hand, our analysis of the transfer heuristic implies that characteristics of environments may also constrain the extent to which heuristics are fungible across situations (Levinthal and Wu, 2010). In fact, the lessons from one context may simply not be applicable in another context, and such lack of transferability means that strategists may be forced to start learning afresh, which in turn increases the risk of making inappropriate generalizations from small samples. The question of how fungible and thus transferable heuristics are across different types of situations is therefore highly important for our understanding of dynamic capabilities, not the least because it also affects the extent to which heuristics may be idiosyncratic to specific firms (Eisenhardt and Martin, 2000).

5.3 Avoiding inappropriate generalization and managerial cognitive capabilities

Our results also have implications for understanding the sources of heterogeneity among individual strategists in their ability to make good decisions in novel and changing environments. Such heterogeneity is typically explained through differences in managerial cognitive capabilities (Adner and Helfat, 2003; Helfat and Peteraf, 2015), which are, in turn, considered to stem from either differences in relevant prior experience (e.g., Hambrick and Mason, 1984; Castanias and Helfat, 1991) or differences in specific mental faculties (e.g., Helfat and Peteraf, 2015). Our findings imply that understanding the bias-variance tradeoff and being aware of the risk of inappropriate generalization in combination with knowledge of the relevant environmental characteristics are important aspects of managerial cognitive capabilities because they lead to better learning rules and, by consequence, to heuristics that with higher prediction accuracy.

This leads to empirically testable predictions concerning the relationship between managers' characteristics and the learning rules they use or the accuracy of prediction from learned heuristics. For example, managers trained in statistics (e.g., those who have an engineering degree) may be more aware of the risk of inappropriate generalization and thus use more appropriate learning rules. Similarly, managers with certain experience may be better at assessing the similarity between contexts and thus be better at deciding about whether and to what extent to transfer prior experience. An intriguing question for further research is whether differences between strategists can be reduced by training them about the bias-variance tradeoff. Laboratory experiments can be used to answer that question. Finally, empirical research may also examine whether individual strategists deliberately and consciously avoid inappropriate generalization or whether they are intuitively aware of the bias-variance tradeoff and what explains differences among strategists in that respect. Using think-aloud protocols may help to answer that question.

5.4 Limitations and conclusions

The bias-variance tradeoffs that we explained in this article provide an explanation for how characteristics of decision environments affect the accuracy of predictions strategists make when they have to learn the relationship between available information and the decision variable they must predict from available observations. While this is an important explanation for why simple and biased heuristics are associated with more accurate predictions, it is not the only one. For example, Bettis (2017) has argued that a key benefit of heuristics is that they are effective in dealing with complexity in strategic decision-making. In addition, Vuori and Vuori (2014) have argued that a key benefit of heuristics is that they facilitate coordination among individuals within an organization.

We have also made a number of simplifying assumptions in our formal approach. For example, we have assumed a linear relationship between attributes and outcomes and a continuous outcome variable. However, the fundamental explanation why simple heuristics allow strategists to make more accurate predictions (the bias-variance tradeoff and the risk of inappropriate generalization) applies more generally and thus the results we have presented above are valid beyond the heuristics that can be modeled with the functional form we use. In addition, in studying changing environments we have exclusively focused on the most simple instance of change, namely a one-time shift. However, in reality often more complex patterns of change exist. For example, while some regular-

ities may be stable across instances of the same decision type others will change more frequently, In particular, there may be higher-order regularities which give patterns to changing environments (e.g., industry lifecycles; Klepper, 1997). These, in turn, will affect how strategists may learn heuristics effectively. Further insights may thus be gained through formal work that relaxes some of the simplifying assumptions we have made.

We have highlighted some areas for further research above. The most important implication of our findings for future research is that both theoretical and empirical research must account for how characteristics of strategic environments affect how strategists learn and use heuristics.

References

- Adner R and Helfat CE. (2003) Corporate effects and dynamic managerial capabilities. *Strategic Management Journal* 24(10): 1011-1025.
- Astebro T and Elhedhli S. (2006) The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science* 52(3): 395-409.
- Bagos PG and Adam M. (2015) On the covariance of regression coefficients. *Open Journal of Statistics* 5(07): 680.
- Barkema H, G. and Vermeulen F. (1998) International expansion through start-up or acquisition: A learning perspective. *Academy of Management Journal* 41(1): 7-26.
- Barney JB and Wright PM. (1998) On becoming a strategic partner: The role of human resources in gaining competitive advantage. *Human Resource Management* 37(1): 31-46.
- Bettis RA. (2017) Organizationally intractable decision problems and the intellectual virtues of heuristics. *Journal of Management* 43(8): 2620-2637.
- Bettis RA and Prahalad CK. (1995) The dominant logic: Retrospective and extension. *Strategic Management Journal* 16(1): 5-14.
- Bingham CB and Eisenhardt KM. (2011) Rational heuristics: The “simple rules” strategists learn from their process experiences. *Strategic Management Journal* 32(13): 1437-1464.
- Bingham CB and Eisenhardt KM. (2014) Heuristics in strategy and organizations: Response to Vuori and Vuori. *Strategic Management Journal* 35(11): 1698-1702.
- Bingham CB, Eisenhardt KM and Furr NR. (2007) What makes a process a capability?

- Heuristics, Strategy, and effective capture of opportunities. *Strategic Entrepreneurship Journal* 1(1-2): 27-48.
- Bingham CB and Halebian J. (2012) How firms learn heuristics: Uncovering missing components of organizational learning. *Strategic Entrepreneurship Journal* 6(2): 152-177.
- Bingham CB, Howell T and Ott T. (2019) Capability creation: Heuristics as microfoundations. *Strategic Entrepreneurship Journal* forthcoming.
- Bower JL. (1970) *Managing the resource allocation process*, Boston, Mass.: Harvard Business School Press.
- Brighton H and Gigerenzer G. (2015) The bias bias. *Journal of Business Research* 68(8): 1772-1784.
- Brown SL and Eisenhardt KM. (1997) The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly* 42(1): 1-34.
- Burgelman RA. (1983) A model of the interaction of strategic behavior, corporate context, and the concept of strategy. *Academy of Management Review* 8(1): 61-70.
- Calantone RJ, Di Benedetto CA and Schmidt JB. (1999) Using the analytic hierarchy process in new product screening. *Journal of Product Innovation Management* 16(1): 65-76.
- Capron L and Shen J-C. (2007) Acquisitions of private vs. public firms: Private information, target selection, and acquirer returns. *Strategic Management Journal* 28(9): 891-911.
- Castanias RP and Helfat CE. (1991) Managerial Resources and Rents. *Journal of Management* 17(1): 155-171.
- Czerlinski J, Gigerenzer G and Goldstein DG. (1999) How good are simple heuristics? In: G G, P T and Group AR (eds) *Simple Heuristics that Make Us Smart*. New York: Oxford University Press, 97-118.
- Davis JP, Eisenhardt KM and Bingham CB. (2009) Optimal Structure, Market Dynamism, and the Strategy of Simple Rules. *Administrative Science Quarterly* 54(3): 413-452.
- Davis-Stober CP, Dana J and Budescu D. (2010) A constrained linear estimator for multiple regression. *Psychometrika* 75(3): 521-541.
- Dawes RM. (1979) The robust beauty of improper linear models in decision making. *American Psychologist* 34(7): 571.

- Di Stefano G, Peteraf M and Verona G. (2014) The organizational drivetrain: A road to integration of dynamic capabilities research. *The Academy of Management Perspectives* 28(4): 307-327.
- Eisenhardt KM, Furr NR and Bingham CB. (2010) Microfoundations of Performance: Balancing Efficiency and Flexibility in Dynamic Environments. *Organization Science* 21(6): 1263-1273.
- Eisenhardt KM and Martin JA. (2000) Dynamic capabilities: What are they? *Strategic Management Journal* 21(10-11): 1105-1121.
- Ellis KM, Reus TH, Lamont BT, et al. (2011) Transfer effects in large acquisitions: How size-specific experience matters. *Academy of Management Journal* 54(6): 1261-1276.
- Felin T and Foss NJ. (2005) Strategic organization: a field in search of micro-foundations. *Strategic Organization* 3(4): 441-455.
- Finkelstein S and Halebian J. (2002) Understanding acquisition performance: The role of transfer effects. *Organization Science* 13(1): 36-47.
- Gentner D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2): 155-170.
- Gigerenzer G. (2008) Why heuristics work. *Perspectives on Psychological Science* 3(1): 20-29.
- Gigerenzer G and Brighton H. (2009) Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science* 1(1): 107-143.
- Gigerenzer G and Gaissmaier W. (2011) Heuristic Decision Making. In: Fiske ST, Schacter DL and Taylor SE (eds) *Annual Review of Psychology*, Vol 62. 451-482.
- Gigerenzer G and Goldstein DG. (1996) Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review* 103(4): 650.
- Gigerenzer G and Selten R. (2002) *Bounded rationality: The adaptive toolbox*, Boston, MA: MIT Press.
- Gilovich T, Vallone R and Tversky A. (1985) The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17(3): 295-314.
- Goldstein DG and Gigerenzer G. (1999) The recognition heuristic: How ignorance makes us smart. In: G G, P T and Group AR (eds) *Simple Heuristics that Make Us Smart*. New York: Oxford University Press, 37-58.
- Goldstein DG and Gigerenzer G. (2002) Models of ecological rationality: The recognition

- heuristic. *Psychological Review* 109(1): 75-90.
- Haleblian J and Finkelstein S. (1999) The influence of organizational acquisition experience on acquisition performance: A behavioral learning perspective. *Administrative Science Quarterly* 44(1): 29-56.
- Hambrick DC and Mason PA. (1984) Upper Echelons: The Organization as a Reflection of its Top Managers. *Academy of Management Review* 9(2): 193-206.
- Hastie T, Tibshirani R and Friedman J. (2008) *The elements of statistical learning*, New York: Springer Series in Statistics.
- Helfat CE and Peteraf MA. (2015) Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal* 36(6): 831-850.
- Hogarth RM and Karelaia N. (2005) Ignoring information in binary choice with continuous variables: When is less ?more?? *Journal of Mathematical Psychology* 49(2): 115-124.
- James W and Stein C. (1961) Estimation with quadratic loss. *Proceedings of the 4th Berkeley symposium on mathematical statistics and probability (vol.1)*. Berkeley: University of California Press, 361-379.
- Klepper S. (1997) Industry lifecycles. *Industrial and Corporate Change* 6(1): 145-181.
- Levinthal DA and Wu B. (2010) Opportunity costs and non-scale free capabilities: Profit maximization, corporate scope, and profit margins. *Strategic Management Journal* 31(7): 780-801.
- Loock M and Hinnen G. (2015) Heuristics in organizations: A review and a research agenda. *Journal of Business Research* 68(9): 2027-2036.
- Luan S, Reb J and Gigerenzer G. (2019) Ecological Rationality: Fast-and-Frugal Heuristics for Managerial Decision Making under Uncertainty. *Academy of Management Journal* forthcoming.
- Maitland E and Sammartino A. (2015) Decision making and uncertainty: The role of heuristics and experience in assessing a politically hazardous environment. *Strategic Management Journal* 36(10): 1554-1571.
- March J. (1991) Exploration and exploitation in organizational learning. *Organization Science* 2(1): 71-87.
- Nelson RR and Winter SG. (1982) *An Evolutionary Theory of Economic Change*, Cambridge,

MA: Belknap Press of Harvard University Press.

Rumelt RP. (1979) Evaluation of strategy: Theory and models. In: Schendel DE and Hofer CW (eds) Strategic management: A new view of business policy and planning. Boston: Little, Brown and Company, 196-212.

Szulanski G and Jensen RJ. (2006) Presumptive adaptation and the effectiveness of knowledge transfer. Strategic Management Journal 27(10): 937-957.

Taleb NN. (2001) Fooled by randomness: The hidden role of chance in life and in the markets, London, UK: Penguin.

Wooldridge JM. (2013) Introductory econometrics: A modern approach, Fifth edition, Mason, OH: South-Western.

Zollo M. (2009) Superstitious Learning with Rare Strategic Decisions: Theory and Evidence from Corporate Acquisitions. Organization Science 20(5): 894-908.

Zollo M and Winter SG. (2002) Deliberate learning and the evolution of dynamic capabilities. Organization Science 13(3): 339-351.

Appendix 1: Formal results for stable environments

Because the true relationship β_i between attributes x_t and outcome y_t for instances of decision problems of type i is unknown to the strategist, he must assess it by estimating $\hat{\beta}_i$ from the n available observations. Our assumptions are equivalent to the Gauss-Markov assumptions (Wooldridge, 2013). In stable environments, given our assumptions the prediction strategy that uses all p attributes to estimate the $\hat{\beta}_i$ that best explains the relationship between attributes and outcome in the set of n prior observations is equivalent to determining the ordinary least squares (OLS) estimator $\hat{\beta}_{i,OLS}$ in multivariate linear regression. Here best explanation refers to best fit: minimizing the squared differences between the n observed values and the values predicted by using $\hat{\beta}_{i,OLS}$ (Brighton and Gigerenzer, 2015). As is well-known, the OLS estimator $\hat{\beta}_{i,OLS}$ is an unbiased estimator for β_i . Thus, given our assumptions, when the strategist uses all information to estimate $\hat{\beta}_{i,OLS}$, this estimator will be unbiased.

Estimating $\hat{\beta}_i$ from available information provides for a stylized representation of a wide class of prediction strategies, including the determination of linear weights in multivariate linear OLS

regression, but also covers particular types of heuristics that have been extensively studied in the psychology literature, such as take-the-best as an instance of one good reason decision-making (by only using the one attribute that best explains the outcome; Gigerenzer and Goldstein, 1996) or tallying (which uses unit weights for all attributes; Dawes, 1979). Formally, such heuristics can be depicted as “improper linear models” (Dawes, 1979, p.571) with $\hat{\beta}_{i,C}$ as a “constrained linear estimator” (Davis-Stober, Dana and Budescu, 2010, p.521). For example, one good reason decision-making can be formalized by constraining the estimation of $\hat{\beta}_i$ so that all but one of the p attributes in $\hat{\beta}_{i,C}$ will get zero weight. As is well-known, estimating $\hat{\beta}_i$ from available observations is subject to a bias-variance trade-off: the mean squared error of $\hat{\beta}_i$ can be decomposed into a bias and a variance component. Whereas the OLS estimator $\hat{\beta}_{i,OLS}$ is not only an unbiased estimator for the true β_i but also the one among all unbiased estimators that minimizes mean squared error, it has been shown that there is always at least one biased estimator that dominates $\hat{\beta}_{i,OLS}$ in terms of mean squared error (i.e., the increase in bias is offset by lower variance; James and Stein, 1961; Vinod, 1978). Davis-Stober et al. (2010) provide a precise formal analysis of the conditions under which improper linear models that use a constrained estimator $\hat{\beta}_{i,C}$ can allow for more accurate estimation of the true β_i than OLS regression due to this bias-variance tradeoff.

The variance component of the mean squared error of $\hat{\beta}_i$ is due to the fact that $\hat{\beta}_i$ is drawn from a finite sample. All else equal, the variance component will be the larger the smaller the number of available observations n , the larger variability σ^2 , the and the more highly correlated the attributes are among each other (the latter is known as “multicollinearity”; Wooldridge, 2013, p.94). A constrained estimator $\hat{\beta}_{i,C}$ will have lower variance compared to the OLS estimator $\hat{\beta}_{i,OLS}$ for given n , σ^2 and correlations among attributes, and can thus lead to lower expected prediction error than using the OLS estimator under these conditions (see the Appendix in Shmueli, 2010). The expected prediction error (and its respective bias and variance component) from using an estimator $\hat{\beta}_i$ that has been estimated from n observations in predicting the outcome in a new instance of the same type i will depend on the particular realization of attribute values x_0 in that new instance.

Appendix 2: A bound for the optimal transfer rate in a changing environment

The strategist uses a transfer heuristic of the following form (with the weighing parameter $0 \leq b \leq 1$.) to make a prediction about outcome y_0 in a new situation that is known to be drawn from f_2 with the realization of attribute values x_0 :

$$\hat{y}_0(x_0) = (1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0 \quad (4)$$

where $\hat{\beta}_1$ has been estimated using ordinary least squares (OLS) on all observations of process f_1 and $\hat{\beta}_2$ has been estimated using OLS on all observations of process f_2 . Given our assumptions, $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased estimators for the true β_1 and β_2 , respectively (see Appendix 1).

In using the transfer heuristic, the strategist must choose a value for the weighing parameter b . We can interpret b as a transfer rate: the extent to which the lessons learned from f_1 are used in making predictions for outcomes in f_2 . A prediction will be biased when $b < 1$ is chosen. The choice of b is subject to a bias-variance tradeoff, and its optimal value depends on the three characteristics of the decision environment: the number of observations m and n , the magnitude of variability σ^2 and the magnitude of the change $\|\beta_1 - \beta_2\|$. We derive the following result:

Proposition: *There is an upper bound for the optimal choice of $b \in (0, 1]$, denoted by b^O , that minimizes the expected prediction error, given by*

$$b^O \leq 1 - \frac{\text{Var}[\hat{\beta}_2 x_0]}{\|\beta_2 - \beta_1\|^2 \|x_0\|^2 + \text{Var}[\hat{\beta}_1 x_0] + \text{Var}[\hat{\beta}_2 x_0]}$$

Moreover, the following conditions hold (each conditional on “other things being equal”): (1) the upper bound on b^O decreases as n decreases, (2) the upper bound on b^O decreases as σ^2 increases, and (3) the upper bound on b^O decreases as $\|\beta_1 - \beta_2\|$ decreases.

To prove the proposition, we need to show that there is a unique $b^O \in (0, 1]$ that minimizes the expected prediction error of the weighing strategy $\hat{y}_0(x_0) = (1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0$ for a given new observation x_0 , and that the upper bound on b^O is monotone (increasing or decreasing) in n , σ^2 and $\|\beta_2 - \beta_1\|$.

The expected squared prediction error of $\hat{y}_0(x_0)$ is given by (see Hastie et al., 2008, for the bias-variance decomposition)

$$\begin{aligned}
 ESPE(x_0) &= E[y_0 - (1 - b)\hat{\beta}_1 x_0 - b\hat{\beta}_2 x_0]^2 \\
 &= \sigma^2 + [E[(1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0] - \beta_2 x_0]^2 + Var[(1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0] \\
 &= \sigma^2 + Bias_{ESPE(x_0)} + Variance_{ESPE(x_0)}
 \end{aligned} \tag{5}$$

With $E[\hat{\beta}_1] = \beta_1$ and $E[\hat{\beta}_2] = \beta_2$ the bias component of prediction error is

$$\begin{aligned}
 Bias_{ESPE(x_0)} &= [E[(1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0] - \beta_2 x_0]^2 \\
 &= [(b - 1)(\beta_2 - \beta_1)x_0]^2 \\
 &= (b - 1)^2[(\beta_2 - \beta_1)x_0]^2
 \end{aligned} \tag{6}$$

The derivative of $Bias_{ESPE(x_0)}$ with respect to b is $(2b - 2)[(\beta_2 - \beta_1)x_0]^2$, which is negative for $b < 1$ and 0 for $b = 1$. Thus $Bias_{ESPE(x_0)}$ is monotonously decreasing in b , and it is zero at $b = 1$ where has its minimum.

Because $\hat{\beta}_1$ and $\hat{\beta}_2$ are each estimated independently, $Cov[\hat{\beta}_1, \hat{\beta}_2] = 0$. Thus, given that x_0 is fixed also $Cov[(1 - b)\hat{\beta}_1 x_0, b\hat{\beta}_2 x_0] = 0$. Therefore the variance component of prediction error is

$$\begin{aligned}
 Variance_{ESPE(x_0)} &= Var[(1 - b)\hat{\beta}_1 x_0 + b\hat{\beta}_2 x_0] \\
 &= Var[(1 - b)\hat{\beta}_1 x_0] + Var[b\hat{\beta}_2 x_0] + 2Cov[(1 - b)\hat{\beta}_1 x_0, b\hat{\beta}_2 x_0] \\
 &= (1 - b)^2 Var[\hat{\beta}_1 x_0] + b^2 Var[\hat{\beta}_2 x_0]
 \end{aligned} \tag{7}$$

The derivative of $Variance_{ESPE(x_0)}$ is $(2b - 2)Var[\hat{\beta}_1 x_0] + 2bVar[\hat{\beta}_2 x_0]$, which has a minimum at $b^M = \frac{Var[\hat{\beta}_1 x_0]}{Var[\hat{\beta}_1 x_0] + Var[\hat{\beta}_2 x_0]}$ for $\sigma^2 > 0$ as the second derivative is positive (if $\sigma^2 > 0$ then $Var[\hat{\beta}_1 x_0] > 0$ and $Var[\hat{\beta}_2 x_0] > 0$; if $\sigma^2 = 0$ then $ESPE(x_0) = Bias_{ESPE(x_0)}$ and thus $b^M = 1$).

We now derive the b^O that minimizes $ESPE(x_0)$ by setting the following first-order condition to zero and solving for b :

$$\begin{aligned}
\frac{dESPE(x_0)}{db} &= \frac{d\sigma^2}{db} + \frac{dBias_{ESPE(x_0)}}{db} + \frac{dVariance_{ESPE(x_0)}}{db} \\
&= (2b - 2)[(\beta_2 - \beta_1)x_0]^2 + (2b - 2)Var[\hat{\beta}_1x_0] + 2bVar[\hat{\beta}_2x_0]
\end{aligned} \tag{8}$$

The b^O that minimizes $ESPE(x_0)$ is thus

$$b^O = \frac{[(\beta_2 - \beta_1)x_0]^2 + Var[\hat{\beta}_1x_0]}{[(\beta_2 - \beta_1)x_0]^2 + Var[\hat{\beta}_1x_0] + Var[\hat{\beta}_2x_0]} \tag{9}$$

The second derivative is positive (see above), so $ESPE(x_0)$ has a unique interior minimum $b^O \in (0, 1]$ for $\sigma^2 > 0$ (and $b^O = 1$ for $\sigma^2 = 0$). By applying the Cauchy Schwartz inequality we can derive the following upper bound for b^O :

$$b^O \leq 1 - \frac{Var[\hat{\beta}_2x_0]}{\|\beta_2 - \beta_1\|^2\|x_0\|^2 + Var[\hat{\beta}_1x_0] + Var[\hat{\beta}_2x_0]} \tag{10}$$

A standard results in statistics is that the variance of an OLS estimator $\hat{\beta}_j$ increases in error variance σ^2 and decreases in sample size n (see e.g., Wooldridge, 2013, p.93; a closed-form derivation of the elements in the covariation matrix $Var[\beta_j]$ can be found in Bagos and Adam, 2015). Using this result we can conclude that the upper bound of b^O from equation (10) monotonously increases in $\|\beta_2 - \beta_1\|$, monotonously increases in n (holding m fixed), and monotonously decreases in σ^2 , respectively (for each holding everything else constant).

QED.

Table 1: Characteristics of strategic environments and their effect on learning heuristics

Characteristics of strategic environments	Effect on learning heuristics
The number of prior observations (n and m) is the number of instances of the same type of decision situation i from which the strategist can learn the relationship β_i between attributes and outcome in that type of decision situation.	The smaller the number of prior observations, the higher the risk of making inappropriate generalizations.
Variability (σ^2) is the degree to which each outcome is determined by factors unique to the specific situation that do not generalize to other situations of the same type.	The higher variability, the more difficult it is to discern the true relationship between attributes and outcome for a given number of observations and thus the higher the risk of making inappropriate generalizations.
The relative importance of attributes $x_t = (x_{t,1}, x_{t,2} \dots x_{t,p})$ is the relative contribution of each of the individual attributes $x_{t,1}, x_{t,2} \dots x_{t,p}$ to the outcome, which is given by the components of $\beta_i = (x_{i,1}, x_{i,2} \dots x_{i,p})$ for decision of type i .	The more skewed the distribution of attribute importance, the easier it is to learn which attributes are the most important for a given number of observations.
The correlation among attributes is the degree to which attributes contain the same information about the outcome and are thus proxies for each other.	The larger the correlation among attributes, the more difficult it is to discern their relative importance for a given number of observations.
The magnitude of change ($\ \beta_j - \beta_i\ $) is a measure of the (dis)similarity between two types of decision situations or the difference between a decision situation before and after it has changed.	The larger the magnitude of change, the less transferable are the learnings about the relative importance of attributes from one type of decision situation to another one.