STRATEGY EVALUATION WHEN USING A STRATEGIC PERFORMANCE MEASUREMENT SYSTEM: AN EXAMINATION OF MOTIVATIONAL AND COGNITIVE BIASES

Lan Guo, Theresa Libby, Bernard Wong-On-Wing and Dan Yang

ABSTRACT

The multiple performance measures in strategic performance measurement systems should be selected to represent a set of causally linked strategic drivers and outcomes. The pattern of results thus can provide information concerning the proper execution of the strategy (i.e., the performance evaluation role) and the strength of the cause-and-effect linkages assumed by the strategy (i.e., the strategy evaluation role). Unfortunately, managers’ tendency to re-evaluate the strategy when performance falls short of target is low in practice. Possible explanations include motivational and cognitive biases. We experimentally examine two decision aids, an attribution aid, and a decomposition aid, designed to help managers ease these challenges. Study 1 shows the decision aids, individually and in combination, increase managers’ tendency to re-examine a problematic strategy. Study 2 demonstrates the effectiveness of the two decision aids, when used together, under a different pattern of results and among a sample of more experienced managers.

Keywords: Strategic performance; performance measurement; motivation; cognitive bias; balance scorecard; judgment and decision-making
INTRODUCTION

Strategic performance measurement systems (SPMS) such as the balanced scorecard (BSC) can serve as both a comprehensive performance measurement system and a strategic management tool (Chenhall, 2005; Ittner & Larcker, 2005; Kaplan & Norton, 2001, 1996). Organizations can use an SPMS to clarify and communicate strategy throughout the organization, set performance targets that align unit and individual objectives with the selected strategy, and periodically evaluate the strategy and identify whether a change in strategy is required. Unfortunately, managers’ use of SPMS results to assess the effectiveness of their strategy, particularly when performance falls short of target, is relatively low in practice (Campbell, Datar, Kulp, & Narayanan, 2015; Ittner & Larcker, 2005). In the current chapter, we propose and test two decision aids intended to increase managers’ tendency to use a firm’s SPMS to evaluate the quality of its strategy.1

Much prior research has examined the use of an SPMS for performance evaluation (e.g., Banker, Chang, & Pizzini, 2004; Libby, Salterio, & Webb, 2004; Lipe & Salterio, 2000; Wong-On-Wing, Guo, Li, & Yang, 2007), while research on the strategic management function of the SPMS is relatively scarce. Studies that have examined the use of SPMS to evaluate strategy are of two main types: those that use field data to test the assumed causal relations in the business model underlying the SPMS (e.g., Campbell et al. 2015; Dikolli & Sedatole, 2007; Huelsbeck, Merchant, & Sandino, 2010) and those that examine how individual managers make strategy-related judgments and decisions using SPMS results (Cheng & Humphreys, 2012; Choi, Hecht, & Tayler, 2013; Tayler, 2010). Our study falls within this second stream of research by examining managers’ ability to interpret SPMS results from a strategic perspective and to use this information to evaluate the appropriateness of their division’s strategy.

Prior research indicates that evaluating the validity of strategy based on SPMS performance patterns may be subject to motivational and cognitive biases.2 First, since upper-level managers are often involved in strategy design, self-serving attributional bias (Heider, 1958; Zuckerman, 1979) may prevent them from attributing poor outcome performance to inappropriate strategies that are (partly) designed by them. This is because external (vs internal) attribution can protect one’s self-esteem in times of failure (Pyszczynski & Greenberg, 1987; Zuckerman, 1979). Consistent with this view, Tayler (2010) finds that managers who are involved in choosing strategic initiatives view those initiatives to be more successful than those who are not involved in the selection process. Like the “attribution therapy” used by psychologists to reduce individuals’ self-serving attributional bias (e.g., Noel, Forsyth, & Kelley, 1987), we introduce a simple decision aid that directs managers’ re-attribution of poor outcome performance. We consider whether this decision aid that is similar to that used by Wong-On-Wing et al. (2007) to improve performance evaluation judgments will also improve manager’s ability to recognize they may be pursuing a failing strategy when faced with negative SPMS results. We label this our “attribution” decision aid.
Second, Kaplan and Norton (2008) argue the pattern of SPMS results should reflect the strength of the cause-and-effect linkages assumed by the firm’s strategy. If the expected correlations between driver (e.g., learning and growth and/or internal business process measures) and outcome measures (e.g., customer and financial performance measures) included in the SPMS are not observed, the existing strategy should be re-evaluated for its effectiveness. Even so, a few factors can make this apparently straightforward evaluation task cognitively challenging: (1) the SPMS usually includes a large set of measures and managers may not attend to all measures and their results simultaneously (which may result in pattern recognition difficulties), (2) the underlying business model being invalid is usually an obscure or unavailable hypothesis for managers (which may result in hypothesis generation difficulties), and (3) many managers are unfamiliar with the strategy evaluation function of the SPMS (which results in an inappropriate match between the strategy evaluation tool and manager’s level of expertise). To facilitate managers’ information processing in strategy evaluation, we design a second decision aid that decomposes this complex evaluation task into three smaller judgment components. We label this our “decomposition” decision aid.

To examine the potential effectiveness of our decision aids, we report the results of two case-based experiments where participants act as strategic business unit (SBU) managers of a chain of specialty clothing stores targeting professional women. Participants learn that the SBU manager together with corporate management designed and adopted the new growth strategy three years ago, under the assumption that most target customers are not price-sensitive. Within three years, this strategy should yield increased profits, but SPMS results indicate performance on driver measures is greater than target while performance on outcome measures is lower than target (labeled the “good driver-poor outcome pattern”), a pattern that Kaplan and Norton (2008) argue is highly indicative of a need to reconsider firm strategy due to a possible invalid strategic assumption (i.e., customers may be more price sensitive than assumed).

In Study 1, we utilize a 2 (attribution aid present/absent) × 2 (decomposition aid present/absent) experimental design. Participants are 78 MBA students with an average of 5.67 years of full-time work experience. Our dependent variable is the participants’ perceived need to re-examine the current strategy after considering the SPMS results. Although the case material clearly suggests potential problems with the strategy, results of Study 1 indicate that without the help of the decision aids, participants do not always consider the need to re-examine the strategy. We find that both the attribution and decomposition decision aids significantly raise the participants’ tendency to re-examine the strategy as predicted. This implies that both the self-serving attributional bias and information processing difficulties to some degree affect managers’ strategy evaluation.

In Study 2, we use a 2 (joint decision aids present/absent) × 2 (poor driver, poor outcome pattern/good driver, and poor outcome pattern) experimental design to further explore the most effective condition from Study 1 (i.e., when both decision aids are present) while adding a “poor driver-poor outcome performance pattern.” This pattern suggests a positive rather than a negative or no
relationship between driver and outcome performance. If the decision aids work as theory suggests, we expect that participants using the joint decision aids will recognize the need to place less emphasis on re-examining the strategy under this poor driver-poor outcome performance pattern than under the good driver-poor outcome pattern (the pattern presented in Study 1).

Participants are 57 middle managers from 2 public companies who were participating in an executive education session. Results indicate that, similar to findings of Study 1, the use of the joint decision aids increases these more experienced managers’ tendencies to re-examine the strategy under the good driver-poor outcome performance. We also find that only with the use of the joint decision aids are managers able to place more emphasis on re-visiting the strategy under the good driver-poor outcome pattern than under the poor driver-poor outcome pattern, indicating further evidence of the effectiveness of the decision aids.

The results of the two studies reported here contribute to both research and practice. First, our results suggest that managers may suffer from both motivational biases and information processing difficulties when using an SPMS for strategy evaluation purposes. Our experiments document that even when SPMS results clearly indicate a disconnect between driver and outcome performance, managers do not always recognize the need to reassess the strategy. We also find that with the help of the attribution and/or decomposition decision aids, managers perceive a significantly higher need to re-evaluate the invalid strategy. Based on these results, we can infer that while using the SPMS to make strategy evaluation decisions, managers (both novice and more experienced ones) may encounter biases that the two decision aids were designed to overcome, that is, self-serving attributional bias, and information-processing difficulties. Although many studies (e.g., Banker et al. 2004, Libby et al. 2004, Lipe & Salterio, 2000) have examined judgment challenges in using the SPMS for performance evaluation, few studies have been conducted to understand challenges managers face when using an SPMS to evaluate strategy.

Second, from a practical standpoint, our results are important because features specific to an SPMS are expected to facilitate strategy evaluation and strategic learning (Kaplan & Norton, 2000, 2001, 2008). To obtain this intended benefit, it is important to investigate ways in which managers’ judgment biases can be addressed. Our results suggest that the attribution decision aid can reduce upper-level managers’ self-serving attributional bias while the decomposition decision aid can reduce the complexity of the strategy evaluation task, both of which help (independently and when used in tandem) to improve the quality of the manager’s strategy evaluation judgment.

The remainder of this chapter is organized as follows. The next section provides the theoretical background, which leads to the development of the hypotheses. We subsequently describe the research method and results of Study 1, followed by those of Study 2. In the last section, we discuss the main findings and limitations of both studies.
LITERATURE REVIEW AND HYPOTHESES

Strategic Performance Measurement Systems and Strategy Evaluation

An SPMS can be used as a strategic management tool that provides an integrated approach to relate operating performance to the firm’s strategic vision (Chenhall, 2005). The SPMS provides managers with financial and non-financial information that, taken together, is meant to communicate the causal linkages that must exist throughout the organization to allow for the achievement of strategic goals (Kaplan & Norton, 2001). In a BSC, the specific type of SPMS examined in the current study, there are typically four main categories of performance measures. These categories can be represented as a causal chain. Investments in employee learning and growth can impact the effectiveness of internal business processes which in turn impact customers’ positive experiences with the firm thus improving financial results (Kaplan & Norton, 2001). The measures appearing earlier in the sequence of causal events have been dubbed “driver” measures of strategy while the measures appearing in the later categories have been dubbed “outcome” measures (Ittner et al. 2003).

Kaplan and Norton (2008) argue that BSC results indicating high performance relative to target on driver measures (e.g., learning and growth or internal business process measures) and low performance relative to target on outcome measures (e.g., financial measures) may indicate that strategic drivers are disconnected from strategic outcomes. While this pattern of results may result from implementing a new strategy, if this pattern continues over several periods after the new strategy has been implemented, then managers should consider reassessing the validity of the strategic assumptions underlying the BSC (also discussed in Kaplan & Norton, 1996, 2001).4

Surprisingly, managers’ use of SPMS results to assess the effectiveness of their strategy is relatively low in practice (Campbell et al. 2015; Ittner & Larcker, 2003). For example, Ittner and Larcker (2003) find that among firms that create causal business models, only 21 percent make the effort to validate the causal links between driver and outcome measures. Ittner and Larcker (2003, p. 89) comment that “businesses often fail to establish such links partly out of laziness or thoughtlessness.” Below we discuss two potential biases that may contribute to the minimal use of SPMS by managers for evaluating strategy and propose a decision aid for addressing each bias.5

Self-serving Attributional Bias and an Attribution Decision Aid

Self-serving attributional bias is the tendency for individuals to attribute positive events to their own personal characteristics, but attribute negative events to external factors (Arkin, Cooper, & Kolditz, 1980; Heider, 1958). Several reviews in the social psychology literature confirm the robustness of self-serving attributional bias across different populations and cultures (e.g., Anderson, Krull, & Weiner, 1996; Campbell & Sedikides, 1999; Greenberg, Pyszczynski, & Solomon, 1982; Mezulis, Abramson, Hyde, & Hankin, 2004; Sedikides & Strube, 1995). Research also suggests that self-serving bias persists because it helps
individuals maintain positive self-esteem which in turn leads to greater happiness, more positive affect, and better functioning (for a review, see Mezulis et al. 2004).

Self-serving attributional bias, however, can also lower the quality of individuals’ judgments and decisions. In the present context, managers may inappropriately attribute negative events to external causes rather than to their own decisions. For example, if managers are involved in selecting a firm strategy that fails to achieve expected results over time, self-serving attributional bias may prevent them from attributing poor financial performance to the inappropriateness of their chosen strategy. Instead, they may attribute it to external factors such as unexpected competitive threats or even the ineffective implementation of the strategy by their subordinates. Tayler’s (2010) experimental results are consistent with this notion. He finds that managers who are involved in choosing strategic initiatives view those initiatives to be more successful than those who are not involved in the selection process.

To overcome self-serving attributional bias, psychologists have resorted to so-called “attributional therapy” (e.g., Noel et al. 1987) in which individuals are reminded of the importance of internal factors (e.g., effort) in causing the occurrence of negative events (e.g., academic failure). Noel et al. (1987) find that such attributional therapy significantly improves failing students’ academic performance. Consistent with this approach, our study proposes a simple decision aid (hereafter “attribution decision aid”) that serves as a prompt for managers to consider the importance of an internal factor, the quality of their chosen strategy, in explaining poor outcome performance. It is worth noting that Wong-On-Wing et al. (2007) used a similar attribution decision aid to reduce the effect of a different bias, the fundamental attribution error, in senior managers’ performance evaluations of middle managers that reported to them.6

Different from Wong-On-Wing et al. (2007), we use the attribution decision aid to reduce managers’ self-serving attributional bias in a strategy evaluation task. We expect that without the help of the decision aid, upper-level managers who are involved in designing the strategy will tend to attribute poor outcome performance to external factors (e.g., “My subordinates didn’t execute the strategy well”) rather than to internal factors (e.g., “I chose an inappropriate strategy given the current market condition”). Our expectation is that prompting managers to first assess the importance of strategy in determining outcome performance will help managers overcome any self-serving attributional bias and thereby increase the degree to which they think the strategy should be re-evaluated. We hypothesize the following:

\[ H1: \] Given SPMS results indicating a weak link between performance on driver and outcome measures, managers will perceive a higher need to re-examine the strategy underlying the SPMS when they are first required to assess the importance of strategy in determining inferior outcome performance than managers who are not required to do so.
Pattern Recognition/Hypothesis Generation and a Decomposition Decision Aid

According to judgment and decision-making research in auditing (e.g., Bedard & Biggs, 1991; Hammersley, 2006; O’Donnell & Perkins, 2011), auditors are often incapable of recognizing relationships among multiple pieces of information and/or of developing hypotheses about an underlying event that explains a recognized pattern of data. For example, Bedard and Biggs (1991) find that when performing analytical review, auditors tend to process one or two cues at a time rather than processing all important cues simultaneously. This prevents them from recognizing seeded patterns in the experimental instrument. Among auditors who did correctly recognize the seeded pattern, many did not develop viable hypotheses that could explain the pattern of results, which similarly interfered with their analytical review performance.

We argue that managers may encounter similar difficulties in recognizing problematic performance patterns in an SPMS and developing viable hypotheses that are associated with those patterns. The reasons are three-fold. First, the typical SPMS consists of a large set of measures (often more than 16 for a BSC) and it is hard for managers to attend to them and their results simultaneously (Chenhall, 2005). The inability to attend to all important cues simultaneously has been found to result in pattern recognition failures (Bedard & Biggs, 1991). Second, as for hypothesis generation, field evidence suggests that managers have the tendency to ignore the possibility that causal business models are questionable. Thus, even if they recognize the good-driver-poor-outcome performance pattern from an SPMS, managers may not consider an invalid underlying business model as a viable hypothesis. Third, even managers who are relatively familiar with an SPMS as a performance measurement tool may not have sufficient experience with using an SPMS as a strategy management tool. The strategic management function of the BSC, for example, is a secondary function that was gradually recognized years after the introduction of the BSC as a performance measurement tool (Kaplan & Norton, 1996). Theory suggests that lack of experience or expertise will interfere with managers developing knowledge structures useful for strategy evaluation (e.g., Hammersley, 2006; Rose, Rose, & McKay, 2007).

Given the abovementioned cognitive challenges managers may face in a strategy evaluation task, we propose a decision aid that decomposes the complex strategy evaluation decision into multiple judgment components (hereafter “decomposition decision aid”). In theory, judgment decomposition (Anderson, 1968, 1974; Kaplan, 1975; Raiffa, 1968) allows managers to make smaller judgments that do not require many items of information to be stored and processed simultaneously. By doing so, we can reduce working memory demanded and ease the task. Decomposition of complex judgments has been found to lead to more reliable and accurate judgments in the contexts of performance evaluation (e.g., Butler & Harvey, 1988; Jako & Murphy, 1990; Lyness & Cornelius, 1982), but not for strategy evaluation.
We propose to decompose the strategy evaluation decision into a series of smaller components: (1) managers select the pattern that best describes the SPMS performance, (2) the decomposition directs them to viable hypotheses associated with the identified performance patterns, and (3) managers assess the reasonableness of the assumption underlying the strategy. We expect that such judgment decomposition can reduce the difficulty of the strategy evaluation task and thereby improve the quality of the strategy evaluation judgment. Consequently, we predict managers who are prompted to decompose a complex judgment process into a set of manageable smaller components will be more likely to re-examine the strategy when SPMS results indicate a weak link between driver performance and outcome performance. Our hypothesis is as follows:

**H2:** Given SPMS results indicating a weak link between performance on driver and outcome measures, managers who are required to make decomposed decisions that lead to strategy evaluation will perceive a higher need to re-examine the strategy underlying the SPMS than managers who are not required to do so.

### STUDY 1

**Method**

**Experimental Task**

In this study, we use an experimental case adapted from Lipe and Salterio (2000) and Wong-On-Wing et al. (2007). It describes Classy, an SBU of ASL, Inc., a clothing retailer. Classy consists of a group of clothing stores targeting style-conscious and time-constrained professional women. Highly experienced store managers run the stores.

The case materials indicate that Classy adopted a strategy three years ago, that was jointly designed by the SBU manager and corporate management. The main strategic goal is to grow sales of high-end, high-margin clothing lines. A key assumption of this strategy is that most of the customers are not price sensitive. The strategy consists of two themes: (1) investing in improving knowledge and skills of brand managers and (2) investing in training sales associates on how to provide a “perfect in-store shopping experience.” The participants receive a strategy map providing a graphical illustration of the assumed causal linkages within Classy’s strategy. Participants also receive a BSC for Classy showing its target and actual performance for the current year (see Table 1). Participants learn that the measures included on the BSC are carefully selected by the SBU manager himself/herself in consultation with store managers.

In the case materials, we purposely include factors that should help participants to recognize that the SBU strategy should be re-examined. First, the results relative to target presented on the BSC indicate a clear disconnect between driver and outcome measures. Specifically, as shown in Table 1, the BSC indicates good SBU performance (all actuals greater than targets) on measures in the learning and growth and internal business processes categories, but poor SBU performance (all actuals less than targets) on measures in the
customer and financial categories. Second, the materials note that the current year is three years after the strategy was first implemented and that the effectiveness and success of the strategy should be evident within two to three years. Third, to avoid perceptions that the data presented on the BSC are incorrect or otherwise unreliable, the case also states that an independent CPA firm provided assurance on the relevance of the BSC measures, the reasonableness and

### Table 1. Classy’s Balanced Scorecard – Targets and Actual for Current Year.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Target</th>
<th>Good Driver/Poor Outcome</th>
<th>Poor Driver/Poor Outcome (Study 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Actual</td>
<td>Actual % Better (Worse) Than Target</td>
</tr>
<tr>
<td><strong>Financial:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Sales margins</td>
<td>60%</td>
<td>48.00%</td>
<td>-20.00</td>
</tr>
<tr>
<td>(2) Sales growth in high-end product lines</td>
<td>15%</td>
<td>3.00%</td>
<td>-80.00</td>
</tr>
<tr>
<td>(3) Inventory turnover</td>
<td>6</td>
<td>5.4</td>
<td>-10.00</td>
</tr>
<tr>
<td>(4) Return on assets</td>
<td>15%</td>
<td>9.00%</td>
<td>-40.00</td>
</tr>
<tr>
<td><strong>Customer:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) % of sales in high-end product lines</td>
<td>70%</td>
<td>48.00%</td>
<td>-31.43</td>
</tr>
<tr>
<td>(2) Customer satisfaction rating</td>
<td>80%</td>
<td>65.00%</td>
<td>-18.75</td>
</tr>
<tr>
<td>(3) Sales/square meter of retail space</td>
<td>300,000</td>
<td>¥219,000</td>
<td>-27.00</td>
</tr>
<tr>
<td>(4) Repeat sales %</td>
<td>40.00%</td>
<td>30.00%</td>
<td>-25.00</td>
</tr>
<tr>
<td><strong>Internal process:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Brand recognition rating for high-end product lines</td>
<td>80%</td>
<td>95.00%</td>
<td>18.75</td>
</tr>
<tr>
<td>(2) Number of stock-outs</td>
<td>&lt;3 times</td>
<td>2.90</td>
<td>3.33</td>
</tr>
<tr>
<td>(3) “Mystery Shopper” audit rating</td>
<td>85%</td>
<td>97.00%</td>
<td>14.12</td>
</tr>
<tr>
<td>(4) Time to process customer returns</td>
<td>&lt;4 min</td>
<td>3.9 min</td>
<td>2.50</td>
</tr>
<tr>
<td><strong>Learning and Growth:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Employee satisfaction</td>
<td>80%</td>
<td>82.00%</td>
<td>2.50</td>
</tr>
<tr>
<td>(2) Hours of training invested in sales associates each year</td>
<td>80 hrs</td>
<td>96 hrs</td>
<td>20.00</td>
</tr>
<tr>
<td>(3) Store computerization</td>
<td>60%</td>
<td>60.00%</td>
<td>0.00</td>
</tr>
<tr>
<td>(4) Hours of training invested in brand managers each year</td>
<td>80 hrs</td>
<td>94 hrs</td>
<td>17.50</td>
</tr>
</tbody>
</table>
achievability of the target for each measure, and the reliability of the actual BSC results (as in Libby et al. 2004). Despite these design features, results from Study 1 suggest that without the help of the decision aids, participants on average did not think re-examining the strategy was more important than re-evaluating the subordinates. We suspect that in real business situations where problems with the strategy are less evident, recognizing the need to revisit the strategy would be even more difficult than demonstrated in our study.

**Design and Procedures**

We use a $2 \times 2$ (attribution decision aid present/absent) $\times$ (decomposition decision aid present/absent) between-subjects design. The attribution decision aid manipulation was adapted from Wong-On-Wing et al. (2007) and required the participants to allocate 100 points between two factors: “appropriateness of the strategy given the current market condition” and “employees’ and managers’ execution of the adopted strategy” to indicate the extent to which they believe each factor contributes to Classy’s outcome performance. The attribution decision aid is provided in Appendix 1.

The decomposition decision aid is designed specifically for this study to test $H2$ and is presented in Appendix 2. As the first step, four types of performance patterns are provided to the participants, and participants are asked to identify the performance pattern that can best describe Classy’s BSC results. Subsequently, two boxes direct managers to potential hypotheses that can explain the corresponding performance patterns (e.g., the assumptions on which the strategy was formulated may be questionable for pattern D). Lastly, the participants evaluate the reasonableness of the underlying strategic assumptions and estimate how likely strategic outcomes can be achieved if strategic drivers are successfully delivered.

**Dependent Variables**

The key dependent variable is participants’ perceived need to re-examine the strategy (Re-examine Strategy). The participants are asked to indicate on a scale of 0 (“not at all”) to 10 (“to a great extent”) the extent to which they would suggest to corporate management a re-evaluation of Classy’s strategy based on the BSC results. We also consider as a secondary dependent variable participants perceived need to re-evaluate their subordinates based on the BSC results on a scale of 0 (“not at all”) to 10 (“to a great extent”) (Re-examine Subordinates).

**Participants**

Participants were 78 students enrolled in a graduate management accounting class in an MBA program at a Chinese university. 51 percent of the participants were male, and the average age of the participant group was 28.2 years. The participants had an average of 5.67 years of full-time work experience. Participants had been exposed in class to the fundamentals of the BSC, but none of them had extensive experience in the design and use of the BSC.
Participants completed the experiment as part of a class exercise and were not compensated. The case was originally written in English. Following Brislin (1970), it was translated into Chinese by one of the authors whose first language is Chinese, and then back-translated into English by a graduate student in accounting. Another author reconciled discrepancies between the back-translated version and the original version of the case. There were no significant problems in either the translation or back translation that could not be satisfactorily reconciled by the translators.

Results

Comprehension Checks

First, we examine whether our manipulation of performance pattern (good driver-poor outcome) was attended to by the participants. At the end of the experiment, we asked participants to assess Classy's performance on the financial measures, customer measures, internal process measures, and learning and growth measures, on separate 11-point Likert scales anchored at 0 (“extremely poor”) and 10 (“excellent”). Recall that per Classy’s BSC, all targets were achieved in the internal business process and learning & growth perspectives, but all targets were missed in the customer and financial perspectives.15

We find that the participants in general rated divisional performance relative to target on learning and growth (mean = 8.19, std. dev. = 1.21) and internal business process (mean = 7.90, std. dev. = 1.34) measures as positive (i.e., greater than the scale neutral point of 5) as expected. In addition, they rated the customer (mean = 4.21, std. dev. = 2.08) and financial (mean = 3.17, std. dev. = 1.64) measures as negative (i.e., less than the scale neutral point of 5) as expected. Results of a multivariate analysis of variance (MANOVA) also suggest that none of the scores in the four rated performance categories were significantly different across the four experimental conditions [Wilks’ Lambda (4, 71) F-ratio < 1.32, p > 0.27].

Given these questions were answered after the decision aid(s) had been presented, we also examine separately the mean performance ratings provided by those individuals in the control condition (i.e., those not presented with any decision aids). We find the participants in the control condition rated divisional performance relative to target on learning and growth (mean = 8.05, std. dev. = 1.03) and internal business process (mean = 7.53, std. dev. = 1.31) measures as positive as expected. In addition, they rated the customer (mean = 3.74, std. dev. = 1.20) and financial (mean = 3.11, std. dev. = 1.29) measures as negative as expected. These checks provide evidence that when asked to consider performance for each type of measure individually, participants interpreted the manipulation of the performance pattern as expected.

Participants were also asked to indicate on separate 11-point scales, the perceived realism (0: very unrealistic; 10: very realistic) and the level of difficulty (0: very easy; 10: very difficult) of the case. In general, they thought that the case was reasonably realistic (mean = 6.76, std. dev. = 1.91) and moderately difficult
There were no differences across experimental conditions in perceived realism or case difficulty (all $p > 0.53$).

### Descriptive Statistics

Means and standard deviations for Re-examine Strategy are presented in Table 2 (Panel A) and Fig. 1 (Panel A). As expected, the presence of either the attribution (mean = 8.25, std. dev. = 1.45) or the decomposition (mean = 8.47, std. dev. = 1.64) decision aid increases the participants’ average tendency to re-examine the strategy compared to the control condition where no decision aid is utilized (mean = 6.68, std. dev. = 2.56). The tendency to re-examine the strategy is the highest when both decision aids are provided together (mean = 8.75, std. dev. = 1.07). There was no significant difference in perceived need to re-examine the subordinates between the attribution and decomposition decision aids (mean = 7.00, std. dev. = 2.03) ($t = 0.46, p = 0.68$). This suggests that without the decision aids, participants may not know what the appropriate focus should be.

Table 2. Descriptive Statistics by Experimental Condition Study 1 ($n = 78$).

#### Panel A: Mean (Std. Dev.) for Re-examine Strategy*

<table>
<thead>
<tr>
<th>Attribution Decision Aid</th>
<th>Decomposition Decision Aid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>6.68 (2.56)</td>
</tr>
<tr>
<td>$n = 19$</td>
<td>$n = 19$</td>
</tr>
<tr>
<td>Present</td>
<td>8.25 (1.45)</td>
</tr>
<tr>
<td>$n = 20$</td>
<td>$n = 20$</td>
</tr>
</tbody>
</table>

#### Panel B: Mean (Std. Dev.) for Re-examine Subordinates**

| Absent                  | 7.00 (2.03) |
| $n = 19$                | 5.00 (3.33) |
| $n = 20$                | 4.80 (2.82) |

Notes: *Re-examine Strategy represents the extent to which participants indicated they would suggest that Classy re-examine its current strategy. Each scale ranged from 0 (not at all) through 10 (to a great extent).

**Re-examine Subordinates represents the extent to which participants indicated they thought they needed to re-evaluate their managers. Each scale ranged from 0 (not at all) through 10 (to a great extent).
Panel A: Mean for Re-examine Strategy by Condition

Panel B: Mean for Re-examine Subordinates by Condition

Fig. 1. Means by Experimental Condition – Study 1 (n = 78). Notes: Re-examine Strategy represents the extent to which participants indicated they would suggest that Classy re-examine its current strategy. Each scale ranged from 0 (not at all) through 10 (to a great extent). Re-examine Subordinates represents the extent to which participants indicated they thought they needed to re-evaluate their managers. Each scale ranged from 0 (not at all) through 10 (to a great extent).
**Hypothesis Tests**

We first conduct a two-way MANOVA with *Re-examine Strategy* and *Re-examine Subordinates* as dependent variables, and the two decision aids as independent variables. The results show significant multivariate main effects for both decision aids [Wilk’s Lambda $F(2, 73) = 3.89, p = 0.03$ for the attribution decision aid, and Wilk’s Lambda $F(2, 73) = 7.32, p = 0.01$ for the decomposition decision aid] and an insignificant multivariate interaction effect [Wilk’s Lambda $F(2, 73) = 2.00, p = 0.14$].

We next use a 2(attribution decision aid present/absent) × 2 (decomposition decision aid present/absent) ANOVA with *Re-examine Strategy* as the dependent variable to test our hypotheses. As indicated in Table 3 (Panel A), we find a significant main effect for both the attribution decision aid ($F = 5.36, p = 0.02$) and the decomposition decision aid ($F = 8.28, p < 0.01$). Follow-up cell mean contrast tests indicate that the average score on *Re-examine Strategy* increased when the attribution decision aid was used alone ($t = 2.82, p < 0.01$) and when the decomposition decision aid was used alone ($t = 3.22, p < 0.01$). These results together support $H1$ and $H2$ that the introduction of each decision aid would increase the managers’ tendency to re-examine the existing strategy.

As shown in Table 3 (Panel A), the interaction between the attribution and decomposition manipulations is not significant ($F = 2.63, p = 0.11$). Considering also the pattern shown in Fig. 1 (Panel A), it appears that the two decision aids are not fully substitutes, that is, the tendency to *Re-examine Strategy* increases more when both decision aids are used together than when alone.

### Table 3. Effects of Decision Aids on Re-examine Strategy and Re-examine Subordinates Study 1 ($n = 78$).

<table>
<thead>
<tr>
<th>Panel A: ANOVA for Re-examine Strategy*</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor</strong></td>
<td>df</td>
<td>Sum of Squares</td>
<td>$F$</td>
</tr>
<tr>
<td>Attribution decision aid</td>
<td>1</td>
<td>16.53</td>
<td>5.36</td>
</tr>
<tr>
<td>Decomposition decision aid</td>
<td>1</td>
<td>25.54</td>
<td>8.28</td>
</tr>
<tr>
<td>Attribution × decomposition decision aid</td>
<td>1</td>
<td>8.10</td>
<td>2.63</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ANOVA on Re-examine Subordinates**</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attribution decision aid</strong></td>
<td>1</td>
<td>13.26</td>
<td>1.90</td>
</tr>
<tr>
<td><strong>Decomposition decision aid</strong></td>
<td>1</td>
<td>36.84</td>
<td>5.28</td>
</tr>
<tr>
<td><strong>Attribution × decomposition decision aid</strong></td>
<td>1</td>
<td>7.61</td>
<td>1.09</td>
</tr>
<tr>
<td>Error</td>
<td>74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: *Re-examine Strategy* represents the extent to which participants indicated they would suggest that Classy re-examine its current strategy. Each scale ranged from 0 (not at all) through 10 (to a great extent).

**Re-examine Subordinates** represents the extent to which participants indicated they thought they needed to re-evaluate their managers. Each scale ranged from 0 (not at all) through 10 (to a great extent). All *p*-values are two-tailed.
each is used independently of the other. This is not surprising given that the biases the two decision aids are designed to overcome are of different natures (i.e., motivational vs cognitive).

**Supplementary Analyses**

Descriptive statistics for Re-examine Subordinates are presented in Table 2 (Panel B) and Fig. 1 (Panel B). Compared to the control condition where no decision aid is provided (mean = 7.00, std. dev. = 2.03), the presence of either the attribution (mean = 5.55, std. dev. = 2.19) or the decomposition (mean = 5.00, std. dev. = 3.33) decision aid reduces the participants’ average tendency to re-examine their subordinates in response to the relatively poor outcome performance.

Although not hypothesized, we also test the effects of the two decision aids on participants’ tendency to re-examine their subordinates. As indicated in Table 3 (Panel B), we find a significant main effect for the decomposition decision aid ($F = 5.28$, $p < 0.03$) but the main effect for the attribution decision aid is not significant ($F = 1.90$, $p = 0.17$).

**STUDY 2**

*Hypothesis Development*

In Study 2, we investigate two specific issues that arose based on our examination of the effectiveness of our attribution and decomposition decision aids in Study 1. First, we change the performance pattern. Specifically, we examine the effectiveness of the decision aids under a poor driver-poor outcome performance pattern. This pattern provides evidence that there is a positive covariation between performance on drivers and outcomes (Einhorn & Hogarth, 1986) (and thus the strategy may not be the culprit for poor performance). The literature on the effect of covariation on judgment suggests that individuals may completely ignore such evidence (e.g., Jenkins & Ward, 1965). Thus, when presented with a poor driver-poor outcome performance pattern, managers may not gauge the soundness of the strategy any differently than they would when presented with a good driver-poor outcome performance pattern (as in Study 1). With the help of the attribution and decomposition decision aids we propose, however, managers’ tendency to re-examine the strategy should be enhanced under the good driver-poor outcome pattern (as shown in Study 1) while such tendency should be reduced under the poor driver-poor outcome pattern. Second, since managers may be subject to both motivational and cognitive biases when using SPMS for strategy evaluation, as indicated in Study 1, we explore the effectiveness of the joint decision aids under the two different performance patterns. To achieve these research objectives, we propose $H3$ as follows:

$H3$: Managers who use both the attribution and decomposition decision aids will perceive a greater need to re-examine the strategy when SPMS results indicate a weak (rather than strong) link between performance on driver and outcome measures.
Participants

The participants in Study 2 are middle managers from publicly traded companies located in China. A total of 59 participating managers were from an electronics manufacturer and a commercial bank. One of these organizations had already adopted the BSC prior to our study and the other organization was considering the adoption of the BSC at the time of our study. These middle managers attended a one-day executive education program on performance measurement systems. They completed the experiment at the beginning of the training session so that they were not exposed to teaching materials that may bias their judgment. They volunteered to participate in the study and were not compensated for participation. A total of 69 percent of the participants were male, and the average age was 36.5 years. The participants had an average of 15 years of full-time work experience.

Whether experience with using BSC necessarily leads to better strategy evaluation judgment is an empirical question. On one hand, prior research on pattern recognition (e.g., Baron & Ensley, 2006; Hahn & Chatter, 1997; Whittlesea, 1997) shows that experienced managers may have stronger pattern recognition abilities than MBA students. This is because, with experience, managers may develop prototypes or templates of performance patterns, which may help them recognizing problematic performance patterns in our experiment. If this is true, there should be a small or even insignificant effect of the decision aids among experienced managers. On the other hand, however, evidence suggests that experience alone does not necessarily lead to more accurate judgments especially when no useful feedback is provided (see Kleinmuntz, 1990). In addition, if more experienced managers are more confident in their knowledge, it can potentially decrease their reliance on decision aids, which in turn will decrease their judgment quality (Whitecotton, 1996).

Design and Procedures

We employed a 2 (performance pattern: good driver-poor outcome/poor driver-poor outcome) by 2 (decision aids: present/absent) between-subjects design. Half of the participants were presented with a BSC including the good driver-poor outcome performance pattern (as in Study 1) and the other half were presented with a BSC reflecting a poor driver-poor outcome performance pattern. The poor driver-poor outcome pattern was created by varying the “actual” and “actual % better/worse than target” columns of Classy’s BSC (see Table 1).

Half of the participants were presented with the joint decision aids as in the joint condition in Study 1 while the other half were not provided with any decision aid. Participants read the same case used in Study 1 and their task again was to assess the extent to which they would re-examine the strategy (Re-examine Strategy) and re-examine their subordinates (Re-examine Subordinates) (scale anchored at 0 = not at all through 10 = To a great extent).
Results

Manipulation Checks
As a check on the manipulation of performance pattern, we asked participants to assess Classy’s performance on each of the learning & growth, internal business processes, customer, and financial measures on separate 11-point Likert scales anchored at 0 (extremely poor) and 10 (excellent). In the good driver-poor outcome condition, the participants indicated that performance on the learning & growth and internal business processes measures were higher than on customer and financial measures ($p < 0.01$). There was no difference in performance ratings between the group that received the decision aids and the group that did not (Wilks’ Lambda $F(4, 24) = 0.25, p = 0.91$). In the poor driver-poor outcome condition, participants indicated that performance on all measures was about equal to the scale midpoint of five. Again, performance ratings were not different whether the decision aids were received (Wilks’ Lambda $F(4, 25) = 0.10, p = 0.98$). Between the good driver-poor outcome and the poor driver-poor outcome conditions, the performance on the learning & growth ($t = 7.64, p < 0.01$) and internal business processes ($t = 6.93, p < 0.01$) measures were significantly different whereas that on customer ($t = 1.44, p = 0.15$) and financial ($t = 1.33, p = 0.19$) measures are not. This supports the effectiveness of the manipulation of performance pattern.

Descriptive Statistics
Means and standard deviations for Re-examine Strategy2 are presented in Table 4 (Panel A) and Fig. 2 (Panel A). Under the good driver-poor outcome performance pattern, the joint decision aids raised participants’ tendency to re-examine the strategy when the decision aids were present (mean = 8.79, std. dev. = 1.25) as compared to when the decision aids were absent (mean = 7.00, std. dev. = 2.95). This is consistent with the findings of Study 1. In contrast, under the poor driver-poor outcome pattern, participants’ tendency to re-examine the strategy when decision aids were present (mean = 5.79, std. dev. = 3.09) was reduced compared to when decision aids were absent (mean = 7.00, std. dev. = 2.76).

Test of Hypothesis 3
We conduct a two-way MANOVA with Re-examine Strategy2 and Re-examine Subordinates2 as dependent variables, and the two manipulated variables as independent variables. The results show a significant multivariate main effect for performance pattern (Wilk’s Lambda $F(2, 54) = 3.37, p = 0.04$) and a significant multivariate interaction effect (Wilk’s Lambda $F(2, 54) = 3.09, p = 0.05$).

Next, we perform two separate ANOVAs. As indicated in Table 5 (Panel A), we find a significant main effect of performance pattern ($F = 4.78, p = 0.03$) and a significant interaction between performance pattern and decision aids ($F = 4.78, p = 0.03$) on Re-examine Strategy2. Cell mean contrast tests show that under the good driver-poor outcome performance pattern, the joint decision aids significantly raised participants’ tendency to re-examine the strategy ($t =$}

An Examination of Motivational and Cognitive Biases
Under the poor driver-poor outcome pattern, the joint decision aids decreased such tendency, but not significantly (t = 1.26, p = 0.21).23 Consistent with our prediction, while the mean of Re-examine Strategy2 was equal across performance patterns when the decision aids were absent, the means of Re-examine Strategy2 when the decision aids were present were quite different from one another (t = 3.12, p < 0.01). Specifically, in the presence of the decision aids, managers had a greater tendency to re-examine the strategy under the good driver-poor outcome performance pattern (mean = 8.79, std. dev. = 1.25), but not under the poor driver-poor outcome performance pattern (mean = 5.79, std. dev. = 3.09).

Supplementary Analyses
Though not hypothesized, we also examine the effects on participants’ tendency to re-examine their subordinate managers. Descriptive statistics for Re-examine Subordinates2 are presented in Table 4 (Panel B) and Fig. 2 (Panel B). Under the good driver-poor outcome performance pattern, participants’ tendency to re-examine their subordinates decreased from 5.67 (std. dev. = 2.02) when the decision aids were absent to 4.29 (std. dev. = 2.73) when the decision aids were

---

**Table 4.** Descriptive Statistics by Experimental Condition Study 2 (n = 59).

<table>
<thead>
<tr>
<th>Performance Pattern</th>
<th>Joint Decision Aids</th>
<th>Absent</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good driver-poor outcome</td>
<td>7.00 (2.95)</td>
<td>8.79 (1.25)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 15</td>
<td>n = 14</td>
<td></td>
</tr>
<tr>
<td>Poor driver-poor outcome</td>
<td>7.00 (2.76)</td>
<td>5.79 (3.09)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 16</td>
<td>n = 14</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Mean (Std. Dev.) for Re-examine Subordinates2

<table>
<thead>
<tr>
<th>Performance Pattern</th>
<th>Joint Decision Aids</th>
<th>Absent</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good driver-poor outcome</td>
<td>5.67 (2.02)</td>
<td>4.29 (2.73)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 15</td>
<td>n = 14</td>
<td></td>
</tr>
<tr>
<td>Poor driver-poor outcome</td>
<td>5.81 (2.93)</td>
<td>5.93 (3.00)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 16</td>
<td>n = 14</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *Re-examine Strategy2 represents the extent to which participants indicated they would suggest that Classy re-examine its current strategy. Each scale ranged from 0 (not at all) through 10 (to a great extent). **Re-examine Subordinates2 represents the extent to which participants indicated they thought they needed to re-evaluate their managers. Each scale ranged from 0 (not at all) through 10 (to a great extent). Performance pattern: In the good driver-poor outcome condition, the BSC shows that actual performance exceeds target in learning & growth and internal process perspectives, but actual performance is lower than target in the customer and financial perspectives. In the poor driver-poor outcome condition, the BSC shows that actual performance is lower than target across all four BSC perspectives.
Panel A: Mean for Re-examine Strategy2 by Condition

Panel B: Mean for Re-examine Subordinates2 by Condition

Fig. 2. Means by Experimental Condition – Study 2 (n = 59). Notes: Re-examine Strategy2 represents the extent to which participants indicated they would suggest that Classy re-examine its current strategy. Re-examine Subordinates2 represents the extent to which participants indicated they thought they needed to re-evaluate their managers. Each scale ranged from 0 (not at all) through 10 (to a great extent). The performance pattern factor is manipulated by varying the “actual” and “actual % better/worse than target” columns of Classy’s BSC (see Table 1). Specifically, in the Good driver-poor outcome condition, the BSC shows that actual performance exceeds target in the learning & growth and internal process perspectives, but it is lower than target in the customer and financial perspectives. In the poor driver-poor outcome condition, the BSC shows that actual performance is lower than target across all four BSC perspectives.
present. Under the poor driver-poor outcome pattern, the tendency to re-evaluate their subordinates was similar whether or not the joint decision aids were utilized.

As shown in Table 5 (Panel B), both the main and interaction effects are insignificant for Re-examine Subordinates2. However, cell mean contrast tests and Fig. 2 (Panel B) show that the mean of Re-examine Subordinates2 was marginally significantly different across performance patterns when the decision aids were present ($t = 1.67, p = 0.10$) but not when they were absent. In other words, only with the help of the decision aids did the participants gauge their subordinates’ performance differently under different performance patterns.

### DISCUSSION

With traditional performance measurement systems, performance on strategic driver and outcome measures are often separated and involve different time horizons. It is therefore hard to determine whether disappointing financial outcomes

#### Table 5. Results for Joint Decision Aids under Different Patterns of Performance Study 2 ($n = 59$).

<table>
<thead>
<tr>
<th>Panel A: ANOVA on Re-examine Strategy2*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>Performance pattern</td>
</tr>
<tr>
<td>Decision aids</td>
</tr>
<tr>
<td>Performance pattern × decision aids</td>
</tr>
<tr>
<td>Error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: ANOVA on Re-examine Subordinates2**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>Performance pattern</td>
</tr>
<tr>
<td>Decision aids</td>
</tr>
<tr>
<td>Performance pattern × decision aids</td>
</tr>
<tr>
<td>Error</td>
</tr>
</tbody>
</table>

Notes: *Re-examine Strategy2 represents the extent to which participants indicated they would suggest that Classy re-examine its current strategy. Each scale ranged from 0 (not at all) through 10 (to a great extent).

**Re-examine Subordinates2 represents the extent to which participants indicated they thought they needed to re-evaluate their managers. Each scale ranged from 0 (not at all) through 10 (to a great extent).

Performance pattern: Manipulated by varying the “actual” and “actual % better/worse than target” columns of Classy’s BSC (see Table 1). Specifically, in the good driver-poor outcome condition, the BSC shows that actual performance exceeds target in the learning & growth and internal process perspectives, but it is lower than target in the customer and financial perspectives. In the poor driver-poor outcome condition, the BSC shows that actual performance is lower than target across all four BSC perspectives.

Decision aids are either absent or present. In the absent condition, participants do not receive any decision aids. In the present condition, participants first receive the decomposition decision aid and then the attribution decision aid.

All $p$-values are two-tailed.
are due to the poor execution of the strategy or the ineffectiveness of the strategy. In contrast, the multiple performance measures included in an SPMS that are selected to represent causally linked strategic drivers and outcomes enable one to infer the effectiveness of the strategy more easily. This feature of an SPMS is expected to improve the effectiveness of strategy evaluation and strategic learning (Chenhall, 2005; Kaplan & Norton, 2000, 2001, 2008). However, even though information about strategy effectiveness is readily available in an organization’s SPMS, managers may not fully employ such information in their judgments due to potential motivational as well as cognitive biases.

This chapter presents the results of two studies examining the effectiveness of two decision aids designed to increase managers’ tendency to employ information that is readily available in the firm’s SPMS to evaluate the effectiveness of the firm’s current strategy. In Study 1, a sample of MBA students acted as SBU managers who were involved in strategy design. They were presented with a BSC indicating that the SBU performed well on driver measures but poorly on outcome measures. We find that when participants were prompted to reconsider the possibility that an incorrect strategy could be affecting outcome performance (i.e., used the attribution decision aid) or to complete a series of decomposed judgment tasks designed to facilitate the processes of pattern recognition and hypothesis generation (i.e., used the decomposition decision aid), the participants expressed a stronger tendency to re-examine the strategy as compared to participants not provided with these prompts.

Study 2 involved more experienced middle managers. We explored the effect on managerial judgment of the decision aids used together under two different performance patterns, that is, when the SBU performed well on driver measures but poorly on outcome measures (as in Study 1) and when the SBU performed poorly on both driver and outcome measures. While the first performance pattern suggests a potential problem with the underlying strategy, the second pattern does not. We found that the use of the decision aids (used together) appropriately increased managers’ tendency to re-examine the underlying strategy under the good driver-poor outcome performance pattern (which replicates results in Study 1), but not under the poor driver-poor outcome pattern. We also found that only with the use of the decision aids were managers able to make different judgments about strategy effectiveness under the two different performance patterns.

As with any study of this type, our experiments have their limitations. First, our participants’ work experience was gained in China. We note that due to their cultural background, our Chinese participants may differ from Westerners in their response to our experimental conditions in at least two ways. First, recent research in cross-cultural psychology (e.g., Choi, Dalal, Kim-Prieto, & Park, 2003; Nisbett, Peng, Choi, & Norenzayan, 2001) and in accounting (Wong-On-Wing & Lui, 2007, 2013), finds that East Asians tend to think more holistically and consider more information when making an attribution than their Western counterparts. Consistent with that research, we expect that our Chinese participants are more likely than Westerners to attend to the BSC’s multiple components and consequently are less susceptible to the self-serving
Attributional bias. Second, while our participants did not have the choice to use or ignore the decision aids, prior research indicates when given a choice, Chinese participants may be more willing to rely on decision aids than their North-American counterparts. For example, Arnold, Clark, Collier, Leech, and Sutton (2005) show that accountants in a high power-distance Chinese culture (Singapore) rely more on a knowledge-based system, which is supported by senior management, than those in a low power-distance culture (Australia). Thus, future research testing the generalizability of our findings among managers from other cultures and in settings where managers have a choice to use the decision aid or not is warranted.

Second, in our study of the combined effect of the two decision aids, the pattern recognition aid was always presented first followed by the attribution aid. Whether similar results would obtain if the order of the aids was reversed is an empirical question. Although we have no theoretical reason to believe that the order of presentation would matter, it is worth examining both the isolated effects of the two decision aids and differences that may be caused by a different order of presentation of the aids with more experienced managers.

Our results suggest that both novice and more experienced managers may face difficulties in strategy evaluation in an SPMS setting. In particular, we show that even when the SPMS results clearly indicate a disconnect between driver and outcome performance, managers on average inadequately consider the need to re-examine the strategy, possibly due to self-serving attributional bias and information-processing limitations. After observing that most firms do not establish and validate business causal models, Ittner and Larcker (2003, p. 89) blame managers’ “laziness or thoughtlessness” for such negligence. Our results suggest that cognitive limitations may not be the only culprit; motivational biases such as self-serving attribution may have also contributed to the problem.

Our study also examines ways of improving strategy evaluation. Specifically, we show that judgment improves when we offer decision aids allowing managers to overcome the challenges they face in evaluating strategy in the presence of an SPMS. Our results imply that managers in practice would benefit from the application of simple and relatively inexpensive tools such as the ones utilized in this study when making strategy evaluation judgments based on SPMS results.

NOTES

1. A decision aid is defined as a tool to help the user of the aid solve a problem by presenting the decision maker with some type of imbedded information (Wheeler, Arunachalam, & Murthy, 2011).

2. Motivational biases are caused by conscious choices to ignore important information in decision making to protect one’s self-image or maintain one’s self-esteem. Cognitive biases are typically subconscious and are caused by the way individuals process information (Pyszczynski & Greenberg, 1987).

3. These participants were similar in age and years of work experience to the MBA student participants in several of the previous studies of the BSC and managerial judgment that have used the Lipe and Salterio (2000) case materials (e.g., Banker et al. 2004; Libby et al. 2004).
4. Another viable hypothesis that can explain this performance pattern is measurement error; that is, the strategic assumptions may be correct, but the metrics used may be a poor fit. We control for potential perceptions of measurement error by informing participants that an independent CPA firm provided assurance on the relevance and reliability of the BSC measures similar to Libby et al. (2004).

5. Prior research suggests that appropriately designed decision aids can help decision makers overcome biases and information-processing deficiencies (for a review, see Rose, 2002). For example, Kleinmuntz (1990) demonstrates that a decision aid can improve judgment performance when individuals are required to combine many cues to make an overall judgment while Butler (1985) shows that a decision aid can help to focus an auditor’s attention on base rate information that would otherwise be ignored in sampling risk problems. More recently, Wheeler and Arunachalam (2008) demonstrate that an appropriately designed decision aid can reduce confirmation bias in the judgment of tax professionals. Since the nature of decision error should determine the appropriate strategy for improving judgment (Fischhoff, 1982), we propose decision aids to address two specific limitations that may be encountered by managers when using BSC results to evaluate the effectiveness of firm strategy.

6. Fundamental attribution error refers to observers drawing inferences about an actor’s disposition from observed behaviors when the behavior can be entirely explained by the situation (Weiner, 1986).

7. Experiential learning and expertise development usually precedes effective pattern recognition and hypothesis generation (Bonner & Walker, 1994; Libby & Luft, 1993). Hammersley (2006), for example, finds that industry specialist auditors can interpret and fill in particular patterns of misstatements that occur often in their industry of specialization while auditors working outside of their industry specialization cannot recognize the important implications of even complete patterns presented to them.

8. This design choice is based on prior studies that illustrate managers’ judgments improve when presented with causal strategy maps (Banker et al. 2004; Banker, Chang, & Pizzini, 2011; Cheng & Humphreys, 2012; Vera-Muñoz et al. 2007).

9. Case materials are available from the authors.

10. Note that along a continuous cause-effect chain, any performance measure can be a driver of its “down-stream” measures as well as an outcome of its “up-stream” measures. Following Wong-On-Wing et al. (2007), we however adopt a simple classification scheme in the case and label the first two perspectives of the BSC as drivers and the last two perspectives of the BSC as outcomes. The Transworld Auto case (Narayanan & Brem, 2010) provides an illustrative pedagogical example of a BSC indicating a similar good driver-poor outcome performance pattern.

11. We developed and pilot-tested two versions of the decomposition decision aid before we settled on the one used in the current study. Initial versions were less directive in that they simply cued participants to consider the underlying relationships between performance measures in the SPMS. As in prior empirical studies in auditing, it was very difficult to get participants to recognize the seeded pattern using more subtle manipulations (e.g., O’Donnell & Perkins, 2011). Therefore, we designed the decomposition decision aid that decomposes the complex strategy evaluation decision into three specific smaller judgments. Although directive, we believe that the three steps included in the decomposition decision aid can be easily adapted to suit different firms’ decision environments and strategies.

12. To further explore the validity of our Re-examine Strategy measure, we correlated the measure with several other related questions: the extent to which participants believe that Classy’s strategy would be successful in the future (reverse coded), how risky they considered it to be for Classy to continue implementing its current strategy, and the extent to which they would recommend that the current strategy be adopted for additional stores acquired in the future (reverse coded). All correlations were greater than 0.58 and significant (p < 0.001). We re-ran all statistical tests reported in the results section using a
dependent variable made up of a composite of those questions and the results were similar.

13. This measure was also correlated with how participants would evaluate their managers’ (as a whole) overall performance (0: very poor; 10: very good) (reverse coded). This correlation was 0.35 and significant ($p < 0.01$).

14. Evidence suggests that Chinese companies have been particularly quick to adopt the BSC. A recent survey conducted in China (Sheng et al., 2008, also see Xiong & Su, 2008) suggests that 53% of respondents had implemented the BSC to various degrees and among those BSC adopting firms, 56% had used the BSC as a strategic management system.

15. We also note that the performance is good across all “driver” performance measures but is superb on the strategy-linked measures (such as training and brand recognition). This should convey to the participants that the SBU employees have exerted extra effort to implement the new strategy (rather than simply working hard in all domains).

16. The perceived difficulty level does not differ between experimental conditions. We note that this does not indicate that the decomposition decision aid is not useful to decision maker’s judgment on the task. In most cases, cognitive biases are subconscious and often due to not exerting enough cognitive effort. Therefore, it is unlikely that participants in the conditions without the decomposition decision aid would find the case any more difficult than those in the condition with the decomposition decision aid.

17. When the decision aids are presented jointly, we always present the pattern recognition aid followed by the attribution aid.

18. No significant differences in the responses were found between the managers from the two firms, so we combined the data to test $H3$.

19. Their full-time work experience related to various areas including finance, banking, or investing (45 percent), general management or personnel (29 percent), research and development (27 percent), accounting, auditing, or taxation (16 percent), marketing or sales (16 percent), engineering (13 percent), information systems (11 percent), and others (4 percent).

20. Note that in the poor driver-poor outcome condition, the performance is below target across all driver measures but is especially poor on the strategy-linked measures. This conveys to the participants that the SBU employees have not implemented the strategy effectively (rather than simply slacking off in all aspects of their job).

21. As in Study 1, we correlated responses to these questions with several other related questions. The correlations between the Re-examine Strategy2 question and the other related questions were again greater than 0.58 and significant ($p < 0.001$). The correlation between Re-evaluate Subordinates2 and their assessment of their managers’ (as whole) overall performance (reversed) was 0.42 and significant ($p < 0.01$). We re-ran all statistical tests using dependent variables made up of composites of these questions. Results were similar.

22. While we expected participants in the poor driver-poor outcome conditions would rate performance on each of the measures lower than five (the scale midpoint), we do note that the mean ratings on the learning & growth (mean = 4.79, std. dev. = 2.34) and internal business processes (mean = 5.70, std. dev. = 2.00) measures were lower than the mean ratings on these measures (mean = 8.85, std. dev. = 1.18 for learning & growth and mean = 8.85, std. dev. = 1.24 for internal business processes) in the good driver-poor outcome condition.

23. Given that these managers were significantly more experienced than the MBA students participating in Study 1, we ran an additional ANOVA including years of work experience as a covariate. The work experience covariate was insignificant ($F = 2.35, p = 0.132$ for Re-examine Strategy2, and $F = 0.95, p = 0.34$ for Re-examine Subordinates2) in this analysis while the interaction between decision aids and performance pattern remained significant ($F = 3.52, p = 0.07$ for Re-examine Strategy2, and $F = 5.39, p = 0.03$ for Re-examine Subordinates2). Moreover, the patterns of results are similar to those illustrated in Fig. 2. Thus, reported results continue to hold even after controlling for the effect of work experience.
ACKNOWLEDGMENTS

The authors thank M. Lipe and S. Salterio for sharing their research instrument. We have received helpful comments from Anne Farrell, Tim Miller, Kim Sawers, Bill Tayler, Alan Webb, reviewers, and participants of the 2012 AAA management accounting mid-year meeting and 2015 EAA annual congress, and workshop participants at Chinese University of Hong Kong, Sun Yat-sen University, and Jinan University.

Bernard Wong-On-Wing and Dan Yang acknowledge the financial support of the Ministry of Education of China through the 111 Project at SWUFE (Project Number B18043: Innovation and Talents Base of Financial Security and Development), as well as the financial support of the National Natural Science Foundation of China (Project Number 71620107005: Research on Capital Market Trading System and Stability).

REFERENCES


## APPENDIX 1 : ATTRIBUTION DECISION AID

Allocate 100 points between the following two factors to indicate the extent to which you believe each contributed to the actual strategic outcomes. Allocate more points to the factor that you believe contributed more to the actual strategic outcomes. You may allocate any number of points from 0 to 100 to either factor. Make sure that the total adds up to 100 points.

- Appropriateness of the strategy given the current market condition
- Employees’ and managers’ execution of the adopted strategy

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriateness of the strategy given the current market condition</td>
<td></td>
</tr>
<tr>
<td>Employees’ and managers’ execution of the adopted strategy</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>100 points</td>
</tr>
</tbody>
</table>

**An Examination of Motivational and Cognitive Biases**
APPENDIX 2 : DECOMPOSITION DECISION AID

According to experts, one should use the BSC to evaluate the following causal linkage: that is, the causal linkages between strategic drivers (e.g., measures of learning & growth and internal process) and strategic outcomes (e.g., targets on financial measures) indicated in the strategy map (e.g., see Fig. 1). This evaluation will allow one to identify potential problems with the execution of the strategy and/or the quality of the strategy. Refer to the BSC in Table 2 and complete the following steps:

Step 1: Compare the Actual Results with the Targets on the BSC (see Table 2):
Which one of the following patterns (A, B, C or D) best describes the SBU’s performance? (circle one)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were targets for strategic <strong>drivers</strong> achieved?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Were targets for strategic <strong>outcomes</strong> achieved?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

**Pattern A and Pattern B indicate that the causal relationship between strategic drivers and outcomes is strong. This suggests that:**
1. The assumptions based on which the strategy was formulated may be reasonable.
2. Achievement of strategic outcomes appears to depend on employees’ and managers’ performance on the driver measures.

**Pattern C and Pattern D indicate that the causal relationship between strategic drivers and outcomes is weak. This suggests that:**
1. The assumptions based on which the strategy was formulated may be questionable.
2. Achievement of strategic outcomes does not appear to depend on employees’ and managers’ performance on the driver measures.

Step 2:
Based on the BSC, to what extent do you believe that the key assumption of Classy’s current strategy (i.e., customers are not price sensitive) is valid? (circle one)

Not at all valid 0----1----2----3----4----5----6----7----8----9----10 To a great extent valid

Based on your above assessment of the validity of Classy’s key assumption, to what extent do you believe that Classy’s employees and managers can achieve the strategic outcomes if they achieve the targets for strategic **drivers**? (circle one)

Do Not Believe so at all 0----1----2----3----4----5----6----7----8----9----10 To a great extent Believe so