THE ACCOUNTING REVIEW
Vol. 87, No. 4
2012
pp. 1135–1163

American Accounting Association
DOI: 10.2308/accr-10273

Lost in Translation: The Effects of Incentive Compensation on Strategy Surrogation

Jongwoon (Willie) Choi
University of Pittsburgh

Gary W. Hecht
University of Wisconsin–Madison

William B. Tayler
Brigham Young University

ABSTRACT: To facilitate managers' decision-making, firms develop strategic performance measurement systems that translate strategy into performance measures. Ideally, managers see measures for what they are—imperfect proxies for intangible strategic constructs. However, managers may fail to fully appreciate the fact that measures are merely representations of the strategic constructs, and act as though the measures are the construct of interest—a phenomenon we label *surrogation*. In this paper, we investigate whether and how the use of strategically linked performance measures for compensation purposes affects managers' propensity to exhibit surrogation. In accordance with the attribute substitution framework (Kahneman and Frederick 2002), we predict incentive compensation exacerbates surrogation, and that this effect is more prevalent when managers are compensated on a single measure of a strategic construct than when managers are compensated on multiple measures of a strategic construct. Via two experiments, we find support for these hypotheses. Our paper contributes to the literature on strategic performance measurement systems by highlighting the tendency of managers to use measures as surrogates for strategy. More

The authors appreciate the helpful comments and suggestions from Ranjani Krishnan (editor), Harry Evans (senior editor), two anonymous reviewers, Rob Bloomfield, Jan Bouwens, Eddy Cardinaels, Joel Demski, Tom Downen, Ron Guymon, Kathryn Kadous, Steve Kachelmeier, Bill Kinney, Lisa Koonce, Joan Luft, Kari Olsen, Lil Mills, Nick Seybert, Anuj Shah, Mike Shields, Monte Swain, Kristy Towry, Tom Vance, Greg Waymire, Alan Webb, Michael Williamson, anonymous reviewers at the AAA Annual Meeting, the Accounting, Behaviors, and Organizations Research Conference, and the Management Accounting Section Meeting, and workshop participants at The University of Amsterdam, Emory University, Illinois State University, Northeastern University, The University of Texas at Austin, Tilburg University, The University of Waterloo, University of Wisconsin–Madison, the American Accounting Association Annual Meeting, the Accounting, Behaviors, and Organizations Research Conference, Brigham Young University's Accounting Research Symposium, the European Accounting Association Annual Meeting, and the Management Accounting Section Midyear Meeting. We also appreciate the research assistance of Daniel Baum, Megan Buckley, Song Wang, and Ye Wang. This research benefited from a Dean's Research Grant from the Goizueta Business School at Emory University.

This paper was awarded the Best Manuscript Award at the 2010 Management Accounting Section Midyear Meeting.

Editor's note: Accepted by Ranjani Krishnan.

Submitted: July 2009
Accepted: February 2012
Published Online: July 2012

generally, we identify a by-product of contracting on imperfect performance measures not previously considered in extant literature, and establish when consideration of costs of this by-product are likely to be critical.

Keywords: *strategy surrogation; strategic performance measurement systems; incentive compensation; attribute substitution.*

I. INTRODUCTION

Strategic performance measurement systems facilitate managers' decision-making by translating strategy into performance measures. A critical feature of these measurement systems is that the strategically linked performance measures be transparent (Kaplan and Norton 1996), allowing managers to "see through" the measures back to the strategy. This transparency allows managers to infer the firm's desired course of action, gauge the appropriateness of the strategy, and adjust the strategy as deemed necessary. Ideally, managers see measures for what they are—imperfect proxies for intangible, over-arching strategic constructs. However, managers may fail to fully appreciate the fact that measures are merely representations of the true constructs of interest, ultimately acting as if the imperfect measures *are* the constructs of interest. This behavior—which we label *surrogation*—potentially hinders managers' ability to make appropriate strategy-related judgments and decisions. The purpose of this paper is to investigate managers' propensity to exhibit surrogation when using a strategic performance measurement system.

Surrogation occurs in a wide variety of circumstances. For example, university administrators accustomed to focusing on teaching evaluations during faculty reviews may lose sight of the strategic construct of "teaching effectiveness" and behave as though the evaluations are the construct of interest (Brickley and Zimmerman 2001; Flinn and Crumbley 2009). Similarly, a manager charged with "delighting the customer" when customer satisfaction surveys gauge strategic success may begin to see maximizing survey results as the strategy, and behave accordingly (Grizzle 2002). In a market setting, investors may use a firm's financial statements repeatedly, and in the process lose sight of the construct(s) of interest that these statements represent (Ijiri 1967, 1975).

Under some circumstances, the phenomenon of surrogation is likely to be less important. For instance, losing sight of strategic constructs that underlie performance measures may be acceptable for low-level employees who do not make strategic judgments and decisions. Additionally, the phenomenon of surrogation is moot when contracts are complete or when measures perfectly capture underlying constructs. However, complete contracts are rarely seen in practice (Williamson 1985), and firms often make trade-offs between contract *completeness* and contract *complexity*. This occurs because managers often face settings in which constructs are intangible, ill-defined, and complex, necessitating the use of imperfect measures of the constructs as part of a firm's strategic performance measurement system.

Many factors are likely to influence managers' propensity toward surrogation. This study focuses specifically on the use of measures as a basis of compensation. Specifically, we ask whether tying incentive compensation to one or more performance measures affects managers' propensity to use measures as surrogates for strategic constructs.

We rely on literature in psychology to develop our theory. In particular, we leverage Kahneman and Frederick's (2002) notion of attribute substitution, which occurs when an individual has a judgment to make regarding a complex "target attribute," and instead relies on a more easily accessible "heuristic attribute" in their judgment process. Using this framework, we develop two hypotheses regarding the influence of incentive compensation on a manager's surrogation of strategy. Our first hypothesis predicts that basing a portion of a manager's compensation on a single measure of a strategic construct exacerbates surrogation of that construct, relative to when the manager receives a fixed wage. Specifically, we argue that the incentive compensation directs a manager's focus toward

the compensated measure as a means of achieving higher compensation. Such attention-directing effects increase the accessibility of the compensated measure as a potential substitute for the construct of interest, increasing the likelihood that surrogation occurs, relative to when fixed-wage-compensation is used. Our second hypothesis predicts that incentive compensation based on multiple measures of a single strategic construct leads to less surrogation of that construct relative to when compensation is based on a single measure of that construct. In particular, we posit that managers are more likely to consider any single measure's capacity to represent the strategic construct when compensated on multiple measures of a strategic construct rather than on a single measure of that construct. Such consideration highlights the limitations of any one measure as a representation of the strategic construct, and according to attribute substitution theory, decreases the likelihood that managers use imperfect proxies as surrogates for the higher-level construct.

Investigating the effect of compensation on surrogation using archival or field data is challenging because of the difficulty of distinguishing surrogation from wealth-maximizing behavior. For example, a salesperson whose compensation is based on customer satisfaction survey scores may focus on maximizing customer satisfaction surveys both because the salesperson responds to incentives to maximize wealth *and* also because of surrogation whereby they view survey results as the strategic construct of interest. Determining where wealth maximization ends and where surrogation begins in such field settings is generally not possible. Therefore, we use an experiment to distinguish between surrogation and wealth maximization. To strengthen the test of our theory and hypotheses, we investigate our research question in a setting that, by necessity, does not map perfectly into the real-world settings in which we are most interested. This approach is similar to prior experimental work in accounting (Libby and Tan 1999; Tayler and Bloomfield 2011). Specifically, participants in our main experiment earn incentive compensation in one task, then move to a separate, similar task where incentive compensation is absent, allowing us to measure surrogation in isolation. In a follow-up experiment, we again induce surrogation via compensation manipulations in one task and measure surrogation in a separate task. However, in the second experiment we also provide compensation in the second task to help address alternative explanations for our results.

In the first task in our experiments, graduate business student participants in our study use the computer game *Spore* (Electronic Arts 2008; Lawton 2008) to design a virtual creature following a specified strategy. We provide four measures of a single strategic construct to participants in real-time as they design their creatures. We manipulate compensation tied to these measures at three levels: fixed-wage-compensation, incentive compensation based on a *single* measure of the creature's abilities, and incentive compensation based on *multiple* (three of the four) measures of the creature's abilities. After designing their creature, participants move to the next task where we measure their tendency to use measures as surrogates for the strategic construct. Each participant views a series of 18 virtual creatures and makes choices regarding the design of these creatures. We designed these choices to allow us to evaluate the extent to which incentive compensation influences participants' tendency to use compensated measures in the first task as surrogates for the strategy in the second task. Importantly, our main experiment does not tie these 18 design choices to compensation in any way, allowing us to view behavior absent the influence of incentive compensation. Based on these choices, we compute participants' "surrogation score," our key dependent variable.

We find evidence consistent with our hypotheses across both of our experiments. Participants compensated on a single measure of a strategic construct were more likely to exhibit strategy surrogation than participants who received a fixed wage. We observe this result even when participants had to incur an opportunity cost to make a choice reflecting surrogation. Further, we find evidence that participants compensated on multiple measures of a strategic construct were less likely to exhibit surrogation than participants compensated on a single measure of a strategic construct.

Our study contributes to the literature on strategic performance measurement systems by highlighting the tendency of managers to use measures as surrogates for strategy. While prior

research identifies the potential for such behavior to arise (Ijiri 1967, 1975), little evidence exists to help explain *why* managers use performance measures as surrogates for strategic constructs. Not only do we provide a theory-based explanation for such behavior, but also we document how an important factor—the number of measures on which managers are compensated—affects surrogation. We also contribute to literature relating to the adage "you get what you measure" (Kerr 1975; Goldratt 2004) and the concept of "performance measure myopia" (Lambert 2001, 38), providing evidence relating to why individuals may engage in such behavior even when they have no direct financial incentive to do so. More generally, we identify a by-product of contracting on imperfect performance measures not previously considered in extant literature (Bushman and Indjejikian 1993; Feltham and Xie 1994; Budde 2007), and establish when consideration of costs of this by-product are likely to be critical.

The remainder of this paper proceeds as follows. Section II provides the background and hypotheses. Section III describes our experiment and related procedures. Section IV presents results. Section V concludes.

II. BACKGROUND AND HYPOTHESES

Strategic Performance Measurement Systems

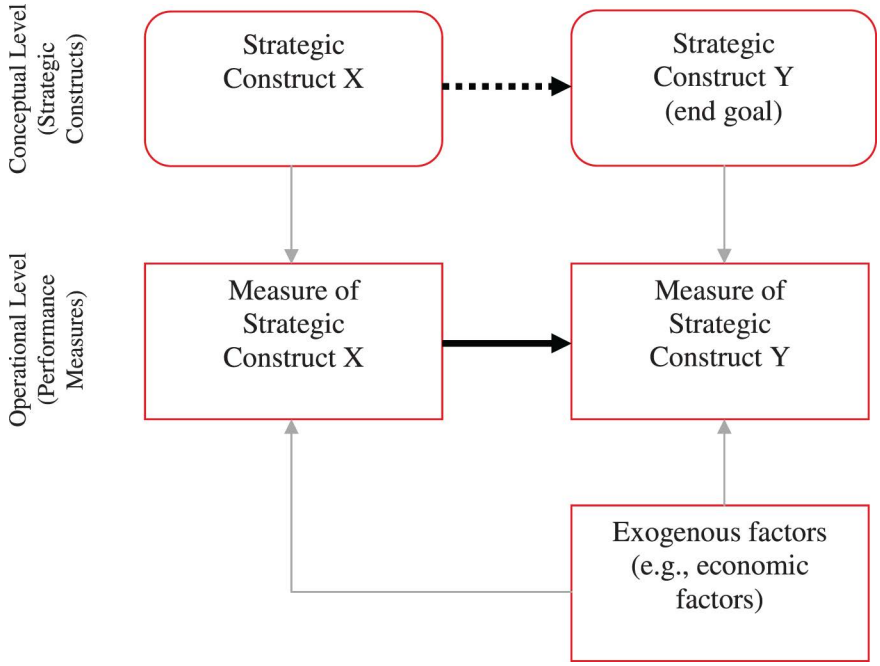
A fundamental purpose of accounting is to provide a means by which decision-makers can grapple with abstract and complex constructs of interest (Ijiri 1967, 1975). Ideally, decision-makers could work directly with these constructs. However, given the nature of these constructs, accounting measures are useful media through which managers make decisions related to the constructs.

This fundamental purpose of accounting is embodied in the context of firms' strategic performance measurement systems. Prior research uses various labels for strategic performance measurement systems, including performance measurement models (Malina et al. 2007) and balanced scorecards (Kaplan and Norton 1992). In this paper, we refer to any system with an explicit link between strategy and performance measurement as a strategic performance measurement system (Epstein and Manzoni 1997; Webb 2004). A key feature of these systems is the mapping of relationships among strategic constructs and performance measures. These outlines are commonly referred to as strategy maps (Kaplan and Norton 2000), value driver maps (Ittner and Larcker 2003), or causal chains of performance (Tayler 2010) (see Figure 1).[1]

Prior research highlights the importance of linking firm strategy and performance measures, noting at least three key functions of such alignment (Langfield-Smith 1997; Ittner and Larcker 1998, 2001; Chenhall 2003). First, translating strategy into a set of measures facilitates *communication* of strategy throughout the firm (Malina and Selto 2001, 2004). That is, performance measures signal to managers potential actions to take, and inform them of how their actions affect others within the firm, as well as the firm as a whole. Second, aligning strategy with performance measures enables *evaluation* of strategy. Performance measures provide tangible evidence that managers use to determine the extent to which the firm is accomplishing its strategy (Bromwich 1990), and to test and validate the strategy (Campbell et al. 2008). Third, tying measures to strategy facilitates the *development* of new strategies (Simons 1991, 1994). Performance measure information can help managers to discover new strategic constructs and/or new strategies. Key to these functions is the transparency of measures, which allows managers to "see through" the measures back to the strategy (Kaplan and Norton 1996).

[1] Strategic constructs, which define the purpose of actions taken, are often referred to as strategic objectives (Kaplan and Norton 1996). A collection of causally connected strategic constructs define the strategy for an organization.

FIGURE 1
Linking Strategy to Measures of Performance



Strategic performance measurement systems facilitate managers' actions by mapping relationships among strategic constructs and performance measures. Above is a simplified example of this mapping based, in part, on the predictive validity framework (Runkel and McGrath 1972; Libby 1981). Horizontal linkages at the *Conceptual Level* represent the hypothesized cause-and-effect relationships between strategic constructs (also called strategic objectives [Kaplan and Norton 1996]). These linkages form "strategy maps" (Kaplan and Norton 2000), also called "value driver maps" (Ittner and Larcker 2003) or "causal chains of performance" (Tayler 2010). Downward links between *Conceptual Level* strategic constructs and *Operational Level* measures reflect the role of measures as imperfect representations of higher-level strategic constructs. Horizontal linkages at the *Operational Level* represent testable relationships between measures (based on the hypothesized relationships of the associated strategic constructs at the *Conceptual Level*). Figure 1 also captures the potential effect of exogenous factors, such as economic conditions, that influence the realization of performance measures.

Surrogation

Notwithstanding the many benefits of aligning a firm's strategic constructs with performance measures, this process can also lead managers to focus on performance measures to the point that they lose sight of the strategic constructs that the measures represent. Although intended as proxies for higher-level constructs, Ijiri (1967, 1975) identifies the potential for measures to replace the construct as the focus of interest. Ijiri (1967, 1975) provides an example of how investors use a firm's accounting measures to learn about the firm's economic viability (the construct of interest).[2] When strategic

[2] A specific example of this phenomenon is seen in the relationship between accounting earnings and Hicksian income (Hicks 1939; Schipper and Vincent 2003). A firm's stakeholders rely on accounting earnings reported on the income statement to learn about the firm's Hicksian income. However, through repeated use of the earnings measure, investors may begin to act as though earnings, and not Hicksian income, is the construct of interest. Strong market responses to firms just missing earnings benchmarks (e.g., Burgstahler and Dichev 1997) is consistent with surrogation at the market level, and suggest that surrogation has non-trivial implications.

objectives are the constructs of interest, this phenomenon involves managers failing to fully appreciate the role of measures as mere representations of strategy, a phenomenon we label *surrogation*.

In some scenarios, surrogation is of little consequence. For example, if a construct is perfectly represented by a single measure or set of measures (Luft and Shields 2003; Farrell 2009), then the notion of surrogation is moot. However, because strategic constructs are typically ill-defined and complex, measures of constructs will be imperfect. That is, no single measure or set of measures perfectly captures the strategic construct in all scenarios to which the construct applies (Baker 1992, 2000; Bushman and Indjejikian 1993).

Surrogation may also be moot when complete contracts are in place. If a contract is designed such that a wealth-maximizing response yields optimal results in terms of the underlying constructs of interest, then employee beliefs about the true objective of the firm (maximize measure versus maximize construct) may be irrelevant. However, a growing body of work suggests that complete contracts are rarely seen in practice. Indeed, because of human limitations, firms must typically make trade-offs between contract completeness and contract complexity, such that the optimal contract is incomplete (Williamson 1985).

Notwithstanding these exceptions, surrogation is potentially detrimental in many scenarios, especially those in which the use of strategic performance measurement systems is most beneficial. For instance, suppose that a manager is charged with evaluating whether a customer-satisfaction focus increases financial performance. If the manager has used customer satisfaction survey scores as surrogates for customer satisfaction, and those scores suggest that satisfying customers does not improve financial results, then the manager may recommend moving away from a customer focus, rather than first evaluating the validity of the survey results. To the extent that the survey scores are an incomplete or imperfect proxy for the strategic construct of customer satisfaction, the decision to abandon the strategy could be suboptimal. Moreover, a manager who defines a strategic construct via an inherently imperfect measure may be less likely to recognize when the measure becomes obsolete, or may not perceive the benefits of a new measure that better captures the strategic construct in a changing business environment. Regardless of its "net" effect, the potential for the costs associated with surrogation to mitigate the effectiveness of the strategic performance measurement system highlights the need for a deeper understanding of the surrogation phenomenon.

Research in psychology provides a theoretical explanation for *why* managers use performance measures as surrogates for strategic constructs. In particular, Kahneman and Frederick (2002) discuss the notion of *attribute substitution*, which occurs when an individual has to make a judgment about a complex "target attribute," and instead relies on a more easily accessible "heuristic attribute" during the judgment process (see also Schkade and Kahneman 1998; Kahneman et al. 2006).[3]

Kahneman and Frederick (2002, 54) discuss three conditions necessary for attribute substitution to occur. First, the target attribute must be relatively inaccessible. Second, the heuristic attribute must be highly accessible. Third, the substitution of the target attribute with the heuristic attribute must not be consciously rejected. Strack et al.'s (1988) study highlights these three conditions. Strack et al. (1988) ask college students "How many dates did you have last

[3] Kahneman and Frederick (2002) characterize attribute substitution as a general framework that includes multiple heuristics, including representativeness (Tversky and Kahneman 1974), availability (Tversky and Kahneman 1973), and recognition (Gigerenzer and Goldstein 1996). While surrogation fits within the same attribute substitution framework as the heuristics above, it is also distinct from these other phenomena. Although we rely on the attribute substitution framework to develop our theory, we retain the surrogation label to differentiate our phenomenon of interest from those noted above, and to more consistently reflect Ijiri's (1967, 1975) original discussion of how accounting measures originally developed to represent a construct to eventually replace the construct itself.

month?" and "How happy are you with your life in general?" and manipulate the order of these questions between subjects. When students answered the happiness question first, the correlation between their answers to the two questions was negligible. However, when students answered the dating question first, the correlation jumped to 0.66. In this example, answering the happiness question first requires making a judgment about the relatively inaccessible attribute of overall happiness. When answering the happiness question second, students appeared to rely on the highly accessible and quantifiable number of dates they had last month to a greater extent than students answering the happiness question first. Though students may have been aware of their weighting of their dating success when assessing their overall happiness, their conscious processing did not reject this substitution.

In the context of strategic performance measurement systems, the first of the three necessary conditions for attribute substitution is usually satisfied because strategic constructs are typically abstract, ill-defined, and complex. In this study, we consider how incentive compensation, a feature common to strategic performance measurement systems, influences the degree to which the *second* and *third* attribute-substitution conditions are met.[4]

Incentive Compensation

A large body of literature addressing the incentives-performance relationship suggests that a variety of mechanisms determine how incentives affect performance (Bonner and Sprinkle 2002). While no one theory provides a comprehensive explanation for how incentives ultimately influence performance, this research suggests that incentive compensation has attention-directing effects, which influence an individual's experience with performance measures (Atkinson et al. 1997; Lambert 2001). In this study, we investigate how linking managers' compensation to performance measures affects their tendency to use measures as surrogates for strategic constructs.

Incentive Compensation on a Single Performance Measure of a Strategic Construct

Prior research shows that firms often use multiple performance measures to represent a single strategic construct (Banker et al. 2000; Chen et al. 2010) and must decide on the number of strategically linked performance measures on which to provide financial incentives. This decision is non-trivial because firms must balance potentially conflicting factors such as contract completeness and contract complexity in their decisions (Ittner and Larcker 1998). As such, firms often choose to base incentives on only a subset of the available measures.

We first consider the basic context in which incentive compensation depends on a single measure of a strategic construct. Psychology literature demonstrates that individuals have limited attention, and that task structure can affect the attention given to features of the environment (James 1890; Shiffrin 1988). Consistent with this research, we posit that incentive compensation directs managers' limited attention toward the compensated measure as the means by which managers' actions affect their compensation, thereby influencing their strategy-implementation experience (Frederickson et al. 1999). These attention-directing effects of incentive compensation heighten the accessibility of the heuristic attribute, which is here the compensated measure, thereby increasing the likelihood of meeting the second necessary condition for attribute substitution. Further, although managers may recognize their use of a heuristic attribute in their assessment, we expect at least some of them not to reject this substitution. Thus, relative to a scenario in which measures are not used for compensation purposes, incentive compensation on a single measure of a strategic

[4] Kaplan and Norton (1996, 219) advocate tying employee compensation to strategic performance measurement systems, but caution that doing so can be problematic when measures "are not perfect surrogates for the strategic objectives."

construct increases managers' propensity to exhibit surrogation. This leads to the following hypothesis:

H1: Basing compensation on a single measure of a strategic construct leads managers to use the measure as a surrogate for the strategic construct more than they would under a fixed-wage-compensation scheme.

Incentive Compensation with Multiple Performance Measures of a Strategic Construct

Like compensation based on a single measure of a strategic construct, compensation based on multiple measures of a single strategic construct can also lead managers to view those measures as a means by which their actions determine their compensation. Thus, regardless of the number of compensated measures, a manager's attention is directed toward compensated measures, increasing the availability of these heuristic attributes. As with compensation based on a single measure, this heightened attention increases the likelihood of meeting the second necessary condition for attribute substitution. However, managers' attention toward, and experience with, multiple measures may be markedly different relative to when compensation is based on a single measure of a strategic construct. We focus on whether managers' experience with multiple measures will increase the likelihood that they will reject the substitution of the heuristic attribute for the target and, hence, fail to satisfy the third necessary condition for attribute substitution (Kahneman and Frederick 2002).

To the extent that multiple measures are relatively independent, managers compensated on multiple measures confront a greater number of action trade-offs that influence their compensation than managers compensated on a single measure (Wood 1986; Campbell 1988; Bonner 1994). To simplify the action-choice task (Payne et al. 1993) managers likely use similarity assessments, which facilitate object classification, concept formation, and related judgments and decisions (Tversky 1977; Fiske and Taylor 1991). Consistent with this theory, we posit that managers compensated on multiple measures of a single strategic construct will consider the similarities of these measures. In doing so, managers are likely to obtain greater awareness of the overall strategic construct and the extent to which the measures represent the strategic construct. This mechanism occurs in two ways. First, comparing the compensated measures increases the likelihood that managers will realize that all of the compensated measures represent the same strategic construct (Johnson 1988; Markman and Medin 1995; Shah and Oppenheimer 2011). Second, comparing the compensated measures highlights the limited ability of any one measure to completely represent the corresponding strategic construct (Zhang and Markman 2001). Thus, by performing these similarity assessments, managers will attend more to the representational role of the measures as proxies for a related strategic construct relative to a setting with a single compensated measure. This increases the likelihood that managers will reject the tendency to replace the strategic construct with performance measures.

Compensation on multiple measures could increase the tendency to use simplifying heuristics and thus increase the accessibility of heuristic attributes, which affects the second necessary condition for attribute substitution. However, such compensation also brings conscious operations into play that increase the likelihood that the substitution of the target attribute with the heuristic attribute will be rejected, reducing the likelihood that the third necessary condition for attribute substitution will be satisfied (Chaiken 1980; Schwarz 2004; Alter et al. 2007). Therefore, we hypothesize that managers compensated on multiple measures of a strategic construct are less likely to exhibit surrogation than those compensated on a single measure of a strategic construct.

H2: Basing compensation on multiple performance measures of a single strategic construct leads managers to use compensated measures as surrogates for the strategic construct less than they would under a single-measure compensation scheme.

III. METHOD

We tested our hypotheses in a laboratory experiment in which participants used computers to work through self-guided instructions and perform various tasks. Participants in our main experiment were 79 students recruited from graduate business courses at a highly ranked business school in the United States. On average, participants were 29 years old, and 61 percent were male.

Experimental Task

After an introduction to the study, participants designed a virtual creature using the *Creature Creator* module of the computer game *Spore* (Electronic Arts 2008; Lawton 2008). Instructions indicated that participants should use the following strategy for designing their creature:

> Design a creature that can socialize with other life forms such that it will become the dominant species on the planet.[5]

Figure 2 provides a sample screen similar to what participants viewed. At the outset of this first task in the experiment, participants received an endowment of 500 "DNA points," with which to purchase various body parts for their creature.[6] Each part added one or more abilities to the creature. Participants could see what abilities each part added before purchase, as well as measures of each ability in real-time as they added each part. When a part was added to the creature, the cost of the part was automatically deducted from the participant's DNA point endowment. If a participant removed a part from the creature, then the program immediately adjusted the creature's ability measures and refunded the participant's DNA points.

While designing their creature, participants saw numerous measures of their creature's abilities, including four measures tied to their creature's ability to "socialize," the strategic construct of interest.[7] These quantitative measures reflected the creature's ability to *sing*, *dance*, *charm*, and *pose* (each on a scale of 0 to 5). Instructions indicated that each of these abilities was equally important to their creature's ability to socialize.

All participants received compensation based on a post-experimental assessment of their creature's ability to achieve global domination via socialization. This compensation allowed us to hold constant, across all conditions, the incentive to design a creature in line with the provided strategy. We calculated participant compensation after the completion of the experimental session. A research assistant placed each participant's creature individually in a *Spore* environment, gave simple commands to the creature to befriend neighboring creatures, and recorded the time needed to befriend two creatures.[8] Within each experimental condition, we compared the performance of each

---

[5] The concepts of "socializing" and becoming "dominant" in this strategy statement may seem at odds with each other. To clarify, we provided participants with a general introduction to the game of *Spore*, which highlighted two strategies for becoming the dominant species: "You can become the dominant species by making friends with other species, or by overpowering them." We provided this information to participants to mitigate any confusion over features of the *Spore* software unrelated to the experimental task. To further ensure that participants understood the general *Spore* introduction and the specific instructions for the experimental task, we required all participants to correctly answer a question indicating that social abilities were most important to their creature's ability to achieve global domination in the task. Any confusion that may have persisted should work against our finding support for our directional predictions.

[6] This budget limit (500 DNA points) served to ensure that participants had enough resources to design a creature in line with the provided strategy, but not without making trade-offs. That is, given this resource constraint, participants had to choose among multiple ways of implementing the stated strategy.

[7] Measures of other abilities inherent to the *Spore* software, but unrelated to our experimental task, were also visible to participants. To avoid confusion, all available measures were described in the instructions. Availability of all measures was held constant across all experimental conditions.

[8] A creature's total social-ability level (i.e., the sum of a creature's *sing*, *dance*, *charm* and *pose* levels) is a key determinant of its time to befriend, as creatures with lower measure levels take longer to befriend other creatures.

FIGURE 2
Screenshot of Initial Creature Design

Above is a sample screenshot of the *Creature Creator* module of the computer game *Spore*. The left side of the screen displays various body parts that can be added to the creature. The upper-right corner of the screen displays various measures of the creature's abilities. The third row of boxes on the upper right displays the four social-ability measures: *sing* (level = 2), *dance* (level = 3), *charm* (level = 2), and *pose* (level = 3). During the experiment, these measures were constantly available and were adjusted real-time as participants added and removed parts from their creatures. "DNA points," used as currency to purchase parts, are displayed in the bottom-left corner of the screen (currently DNA = 45). Buttons at the top center of the screen allow participants to paint their creature and to see their creature move in a simple environment.

participant's creature to that of nine other randomly selected participants' creatures, and this relative performance ranking determined one portion of their compensation. The top performer in each group earned $20, the second best performer received $18, etc. Providing compensation tied directly to successful implementation of the strategy helped reinforce the equal importance of all four measures in minimizing the time to befriend two creatures. By heightening participants' awareness of all four social-ability measures, including this compensation works against finding evidence of surrogation in our experiment.

Note that neither "time to befriend" nor the related tournament-based compensation serve as our dependent variable. Although these variables may indirectly capture the extent to which participants use measures as surrogates for the strategy, they also reflect participants' desire to maximize wealth. Consequently, using these variables introduces a wealth-maximization confound

that would limit our ability to test the effect of incentives on surrogation. For this reason, we constructed an alternative dependent variable that does not suffer from this confound, which we describe following the discussion of our independent variable.

Independent Variable

In addition to the compensation tied to the participant's success at strategy implementation, some participants received incentive compensation tied directly to measures of their creature's social-ability. This compensation is our independent variable, which we manipulate at three levels: no social-ability compensation (*fixed-wage-compensation*), compensation tied to a single measure of the creature's social-ability (*single-measure-compensation*), and compensation tied to three of the four social-ability measures (*multiple-measure-compensation*). In the fixed-wage-compensation condition, participants received $10 regardless of their creature's social-ability measures. In the single-measure-compensation condition, participants received $2 per level of their creature's *sing* measure. For example, a participant who designed a creature with *sing* level 4 would receive $4 \times \$2 = \$8$. In the multiple-measure-compensation condition, participants received $1 per level of their creature's *sing*, *dance*, and *charm* measures. For example, a participant who designed a creature with *sing* level 3, *dance* level 2, and *charm* level 4, would receive $(3 \times \$1) + (2 \times \$1) + (4 \times \$1) = \$9$.

Dependent Variables

We elicited our primary dependent measure using a second task after participants finished designing their virtual creature. To elicit this dependent measure, participants viewed 18 successive pictures of other virtual creatures. With each of these creatures, participants saw two potential combinations of the four social-ability measures (*sing*, *dance*, *charm*, and *pose*). We labeled these combinations as "modification packages." As they viewed each creature, participants selected the modification package that they believed would best implement the stated strategy from the initial creature design task of the experiment. In addition, the instructions indicated that participants should assume that the DNA cost was equal across each pair of modification packages. Participants viewed each of the 18 creatures one at a time, and made their modification-package selections as they viewed each creature. Once participants selected a modification package and moved on to the next creature, they could not go back and change earlier modification-package selections. Figure 3 depicts an example of what participants saw as they made these selections.[9]

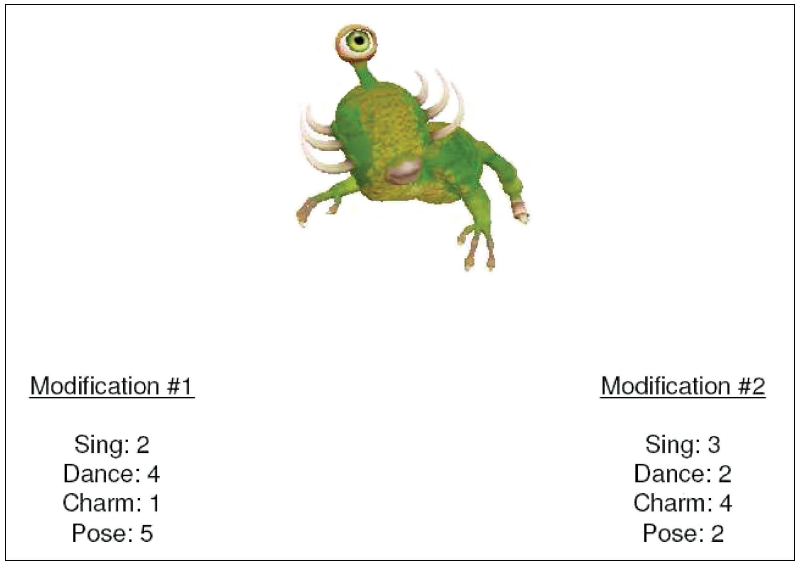
We designed modification-package pairings *ex ante* to allow us to use participants' selections to evaluate the extent to which our compensation manipulation influenced participants' tendency to use compensated measures as surrogates for strategy. To illustrate, consider the sample pairing presented in Figure 3, which includes the following two modification-package options:

Modification package #1: $sing = 2$, $dance = 4$, $charm = 1$, and $pose = 5$.
Modification package #2: $sing = 3$, $dance = 2$, $charm = 4$, and $pose = 2$.

Given this pairing, if a participant selects modification package #2, then this selection is consistent with the goal of maximizing *sing*, because *sing* is 2 in modification package #1 and is 3 in modification package #2. Thus, selecting modification package #2 is consistent with surrogation with *sing*. Selecting modification package #2 is also consistent with the goal of maximizing *sing*, *dance*, and *charm*, because the sum of those three abilities is 7 in modification package #1 and is 9

[9] To control for an order effect, we created four randomly determined orders for the 18 choices, and fully crossed these four order conditions with our three compensation conditions. The order in which choices are made does not alter our results. Thus, we ignore this variable in subsequent analyses and discussion.

FIGURE 3
Sample Modification-Package Choice



Above is a sample of one of the 18 modification-package pairings used to generate the primary dependent variable for the study. For each pairing, participants were shown a picture of a creature along with the two modification packages. Participants were asked to select the modification package that they believed, if added to the creature, would best enable them to implement the provided strategy. After making their selection, participants clicked "Next" to move on to the next creature and modification-package pairing.

in modification package #2. Thus, selecting modification package #2 is also consistent with surrogation with *sing*, *dance*, and *charm*.

To construct our primary dependent variable, we sum the number of selections a participant makes that are consistent with surrogation. In determining whether a selection is consistent with surrogation, we focus on behavior as it pertains to the social-ability measure(s) for which participants had previously earned compensation during the creature-design task of the experiment. Thus, for single-measure-compensation participants, we focus on their behavior as it pertains to the *sing* measure, and investigate whether participants' selections maximize *sing*. For multiple-measure-compensation participants, we focus on their behavior as it pertains to the *sing*, *dance*, and *charm* measures, and investigate whether participants' selections maximize the sum of *sing*, *dance*, and *charm*.

Although all participants saw the same 18 creatures and associated modification-package pairings, not all 18 pairings can be used to measure behavior consistent with surrogation. For example, a pairing in which *sing* is the same in both modification packages cannot be used to measure surrogation with *sing*. Similarly, a pairing in which the sum of *sing*, *dance*, and *charm* is the same in both modification packages cannot be used to measure surrogation with *sing*, *dance*, and *charm*. Thus, we divide the number of selections a participant makes that are consistent with surrogation in their experimental condition by the number of pairings that can be used to evaluate surrogation in that condition to come up with a participant's "comprehensive surrogation score."

TABLE 1

Modification-Package Pairings

	Surrogation-Consistent Modification Package[a]						Alternative Modification Package[a]								
Pairing	S	D	C	P	S+D+C	S+D+C+P	S	D	C	P	S+D+C	S+D+C+P	Pairing Used to Evaluate *sing* Surrogation?[b]	Pairing Used to Evaluate *sing*, *dance*, and *charm* Surrogation?[b]	Pairing Used to Evaluate Opportunity Cost?[c]
1	4	3	5	1	12	13	3	5	3	2	11	13	Yes	Yes	No
2	5	2	4	2	11	13	1	4	5	3	10	13	Yes	Yes	No
3	3	1	5	2	9	11	1	4	1	5	6	11	Yes	Yes	No
4	5	2	4	1	11	12	2	4	1	5	7	12	Yes	Yes	No
5	3	2	5	1	10	11	1	5	3	4	9	13	Yes	Yes	Yes
6	5	2	3	1	10	11	4	1	4	3	9	12	Yes	Yes	Yes
7	4	4	2	2	10	12	1	3	5	5	9	14	Yes	Yes	Yes
8	3	2	4	2	9	11	2	4	1	5	7	12	Yes	Yes	Yes
9	4	2	3	4	9	13	1	3	5	4	9	13	Yes	No	No
10	3	1	4	1	8	9	2	5	1	1	8	9	Yes	No	No
11	3	3	2	1	8	9	2	1	5	3	8	11	Yes	No	Yes
12	5	2	1	3	8	11	2	4	2	5	8	13	Yes	No	Yes
13	2	4	3	2	9	11	2	5	1	3	8	11	No	Yes	No
14	3	5	4	2	12	14	3	1	5	5	9	14	No	Yes	No
15	4	3	4	1	11	12	4	1	3	5	8	13	No	Yes	Yes
16	2	4	3	1	9	10	2	1	5	5	8	13	No	Yes	Yes
17[d]	5	1	4	1	10	11	4	3	2	1	9	10	No	No	NA
18[d]	3	3	2	2	8	10	3	4	1	3	8	11	No	No	NA

This table lists the 18 modification-package pairings provided to participants. Although pairings are numbered in the table, they were not numbered in the study, and were presented in four, predetermined, random orders (order is fully crossed with the compensation manipulation).

[a] The column labeled "Surrogation-Consistent Modification Package" gives the levels of *sing* (S), *dance* (D), *charm* (C), and *pose* (P) of the modification package within each pairing that represents an option consistent with surrogation with *sing*, surrogation with *sing*, *dance,* and *charm*, or both. The column labeled "Alternative Modification Package" gives the alternative to the surrogation-consistent option. Although not shown in the table, modification packages are labeled for each pairing as "Modification #1" or "Modification #2" (see Figure 3). In addition, the label on the surrogation-consistent modification package is randomly determined, and is held constant for all participants.

(continued on next page)

TABLE 1 (continued)

[b] The table also shows whether the pairing is used to evaluate surrogation with *sing* or surrogation with *sing*, *dance*, and *charm*. For example, in Pairing #9, a choice of the modification package listed in the "Surrogation-Consistent Modification Package" column is consistent with surrogation with *sing*, because it has a higher *sing* level than the alternative. However, this pairing is not used to evaluate surrogation with *sing*, *dance*, and *charm*, as the total of these three measures is equal across modification packages in this pairing. However, Pairing #5 can be used to evaluate both surrogation with *sing*, and with *sing*, *dance,* and *charm*, because one package is dominant on both *sing* and on the total of the *sing*, *dance,* and *charm* measures.

[c] The last column indicates whether there is an opportunity cost to selecting the surrogation-consistent Modification Package. *No opportunity cost* pairings are comprised of modification packages in which the total level of the four social-ability measures are equal between the two modification packages. Thus, a participant who chooses the modification package consistent with surrogation incurs no cost to do so. In contrast, *opportunity cost* pairings are comprised of modification packages in which the total level of the four social-ability measures are lower for the surrogation-consistent option than for the alternative.

[d] Pairings #17 and #18 are not used in any test of surrogation. These pairings are included in the experiment as distracters. For Pairing #17, the modification package with higher *sing*, and *sing*, *dance*, and *charm* levels also has a higher total across all four social-ability measures, confounding surrogation and optimal-card selection. For Pairing #18, both *sing* and the total of *sing*, *dance*, and *charm* are equal across packages, so there is no surrogation-consistent modification package for this pairing.

Twelve of the 18 pairings are relevant for evaluating participants' tendency to use *sing* as a surrogate for social-ability, the strategic construct. A different set of 12 pairings are relevant for evaluating participants' tendency to use *sing*, *dance*, and *charm* as surrogates for social-ability. Table 1 presents the 18 modification-package pairings, along with a breakdown of the relevant pairings for measuring behavior consistent with surrogation in each experimental condition.

Modification-package pairings differed according to the extent to which participants had to incur an opportunity cost to make a selection consistent with surrogation. We label pairings in which the total level of the four social-ability measures (*sing*, *dance*, *charm*, and *pose*) is equal between the two modification packages as *no opportunity cost* pairings. With these pairings, a participant selecting the modification package consistent with surrogation does not forgo a higher total social-ability (i.e., incurs no opportunity cost) to do so. In contrast, we label pairings in which the surrogation-consistent modification package has a lower total social-ability level as *opportunity cost* pairings. With these pairings, a participant selecting the surrogation-consistent modification package incurs an opportunity cost because a creature's total social-ability level is a key determinant of its ability to achieve the strategic objective of socializing with other creatures.[10] To illustrate, consider again the pairing shown in Figure 3. While modification package #2 maximizes the *sing* level because *sing* is 2 in modification package #1 and is 3 in modification package #2, modification package #1 has a higher total social-ability level because *sing*, *dance*, *charm*, and *pose* levels sum to 12 for modification package #1 and sum to 11 for modification package #2. Therefore, selecting the surrogation-consistent modification package in this example entails forgoing higher total social-ability, and is thus categorized as an opportunity cost pairing. The last column in Table 1 indicates the opportunity cost status of each of the surrogation-consistent modification package pairings.

In summary, we compute three measures of surrogation: *comprehensive*, *no opportunity cost*, and *opportunity cost*. As described above, we compute these three measures differently depending on how participants were compensated in the creature-design task. Specifically, we measure single-measure-compensation (multiple-measure-compensation) participants' propensity to use *sing* (*sing*, *dance*, and *charm*) as a surrogate for social-ability. We compute both versions of these measures for the fixed-wage-compensation participants to allow for baseline measures of surrogation against which the other conditions can be compared.

Session Timeline

As participants arrived at a session, we randomly assigned them to one of the three compensation conditions. Participants read computer-based instructions and watched a video tutorial on using the *Creature Creator* module of *Spore*. Participants then practiced building a virtual creature for 12 minutes. After practicing, participants read additional instructions that included information about participants' compensation and the strategy to be implemented, took a short quiz to ensure their understanding of the instructions, and then proceeded to design their virtual creatures. Next, participants viewed the 18 modification-package pairings, and made their selections for each pairing. Finally, participants completed a questionnaire to elicit demographic and process-related information. We debriefed and compensated participants after all sessions were completed.

[10] Our use of the label "opportunity cost" is not intended to convey that participants incurred a direct financial cost in choosing one of the modification packages. Rather, our intention in our main experiment is to differentiate pairings in which a surrogation-consistent selection forgoes something of value in executing the strategy (i.e., a creature with a higher total social-ability level).

Additional Discussion of the Experimental Design

A number of issues relating to our experimental design warrant further discussion. First, each of the four social-ability measures completely captures non-overlapping dimensions of the social-ability construct. While at odds with our contention that no single measure or set of measures perfectly captures the strategic construct across all real-world settings in which the construct applies, this design feature allows for a clean test of our hypotheses. Importantly, this design choice highlights the incomplete nature of the compensation contract in all three experimental conditions. A complete contract in our setting would place positive and equal incentive weights on all four social-ability measures. However, as discussed earlier, our compensation schemes are meant to reflect real-world settings in which incomplete contracts are common (Williamson 1985).

Second, we provide no compensation for participants' modification-package selections. Single- or multiple-measure-based compensation similar to that provided to participants during the creature-design task of the experiment would introduce a "wealth maximization" confound, precluding us from observing the phenomenon of interest. Thus, we use an experimental setting that is admittedly different from many real-world settings involving incentives. Indeed, although real-world settings that parallel our experimental setting likely exist (e.g., see Malina and Selto 2004), our experimental setting is not meant to map into any specific real-world setting. Rather, our intent is to create a setting that yields a valid measure of surrogation and a clean test of our hypotheses. However, the results of our main study speak directly to settings where managers must make strategic decisions without having compensation tied directly to every aspect of decision-making.[11,12] Further, to address alternative explanations for our results, we conduct an additional experiment in which we compensate participants for strategically aligned modification-package selections. We discuss this follow-up experiment and associated results in Section IV.

Third, the single-measure-compensation scheme and the multiple-measure-compensation scheme differ not only in terms of the number of compensated measures (our manipulation of interest), but also in terms of the representativeness of the compensated measures. In general, the better the set of measures used for decision making represents strategy, the better the strategic decisions will be. However, our theory focuses on the effect of the number of compensated measures on *surrogation*, and not on the increase in the optimality of strategy-related decisions induced by greater representativeness. Thus, if we chose to measure surrogation behavior indirectly

[11] One potential concern is that, absent incentives, participants in all conditions may exert minimal effort when making their modification-package selections. Notably, low effort works against finding support for our directional predictions, as low effort increases noise in our dependent measures across all conditions. Additionally, if participants exerted minimal effort and selected modification packages randomly, then they would select the surrogation-consistent option about half the time. In contrast, high effort implies a lower likelihood that participants select the surrogation-consistent option. In all three conditions, participants selected the surrogation-consistent option significantly *less* than half the time ($p < 0.01$ in all cases). Furthermore, within each condition, participants chose the surrogation-consistent option significantly more often for *no opportunity cost* pairings than for *opportunity cost* pairings ($p < 0.01$ in all cases), providing additional evidence of an effortful evaluation in this task.

[12] Relatedly, Holmstrom and Milgrom (1991) analyze a model wherein an employee is charged with performing multiple tasks. They show that compensating on only a subset of tasks induces the employee to allocate attention to the compensated tasks at the expense of those tasks that are not compensated, resulting in unbalanced task performance. Our analysis differs from Holmstrom and Milgrom (1991) in that compensation in our study is also expected to lead to an unbalanced interpretation of the strategic construct with the number of compensated measures moderating this effect. Further, our primary dependent variable goes beyond wealth-maximizing behavior to capture surrogation-related behavior induced by the attention-directing effects of incentive compensation. Thus, our study complements Holmstrom and Milgrom (1991) by suggesting additional, but separate, effects of tying compensation to a subset of available measures.

via a measure of decision optimality, then disentangling the effects of multiple-measure-compensation and the representativeness of compensated measures would not be possible (because both are expected to influence decision optimality). However, our dependent variable captures surrogation independent of decision optimality by measuring the tendency to ignore non-compensated measure(s). By constructing our dependent variable in this way, the greater degree of representativeness of the compensated measures in the multiple-measure-compensation condition works against finding support for our hypotheses. Specifically, surrogation in the single-measure-compensation condition involves attending to a smaller proportion of social-ability than surrogation in the multiple-measure-compensation condition. Consequently, to the extent that representativeness of the compensated measures influences participants' judgments and decisions, participants in the single-measure-compensation condition are more likely to recognize that the compensated measure is insufficient in terms of its representativeness, decreasing participants' propensity to use the compensated measures as surrogates for social-ability.

Fourth, our use of the computer game *Spore* for our experimental instrument helps protect against the possibility that participants' priors regarding appropriate surrogates for strategic constructs affected their judgments and decisions, because the setting is far removed from most participants' practical and educational experiences. Importantly, however, the setting still allows participants to "get their hands dirty" implementing the provided strategy and to deal with various associated measures.[13]

Finally, we took several steps to maximize the internal validity of our study. To ensure that the strategy statement was salient, we explicitly stated the strategy on two occasions, the latter of which occurred immediately before participants designed their creature. We also included in the task instructions an explicit statement that all four measures of social-ability are equally important in implementing the strategy and achieving global domination. Further, participants had to demonstrate thorough understanding of the various aspects of our research design via a quiz.[14]

IV. RESULTS

We begin with an analysis related to the effectiveness of our incentive compensation manipulation. To assess whether participants performed the initial creature design task consistently with the provided incentives, we asked each participant to "indicate how much attention you gave to the [*sing*, *dance*, *charm*, and *pose*] abilities, by allocating 100 points among the four abilities. Allocate more points to abilities to which you gave more attention." We compared the mean number of points allocated either to *sing*, or to *sing*, *dance*, and *charm* across conditions (reported in Table 2). Not surprisingly, single-measure-compensation participants allocated more points (i.e., paid more attention) to the *sing* measure ($n = 26$, mean = 35.1 percent) than fixed-wage-compensation participants ($n = 27$, mean = 22.4 percent) ($t = 3.44$, $p < 0.01$).[15] Likewise,

[13] None of the participants reported any prior *Spore* experience. Further, experience playing video games is not a significant predictor of any of the results in our paper.

[14] Two quiz questions addressed the strategy statement and the equal importance of each ability measure. One multiple-choice question read, "Which of the following will help your creature achieve global domination most rapidly?" Options were "The ability to fly," "Social abilities," "Attack abilities," and "The ability to jump." Participants could not proceed until they gave the correct response ("Social abilities"). The other question read, "In *Spore*, which of the following is most important in determining your creature's ability to make friends with other creatures?" Options were *sing*, *dance*, *charm*, *pose*, and "All socializing abilities are equally important in determining a creature's ability to make friends." Again, participants could not continue until they gave the correct responses ("All socializing abilities are equally important in determining a creature's ability to make friends.").

[15] We report one-sided p-values for tests of directional predictions. Otherwise, we report two-sided p-values.

multiple-measure-compensation participants allocated more points to the *sing*, *dance*, and *charm* measures (n = 26, mean = 86.2 percent) than fixed-wage-compensation participants (mean 81.1 percent) (t = 1.98, p = 0.03).[16,17] Collectively, these results suggest that our compensation manipulation was successful.[18] We now turn to our hypothesis tests and related supplemental analyses.

Primary Tests of Hypothesis 1

H1 predicts that compensation on a single measure of a strategic construct increases managers' propensity to use a measure as a surrogate for the strategic construct relative to a fixed-wage-compensation scheme. To test this hypothesis, we examine fixed-wage-compensation and single-measure-compensation participants' surrogation scores. H1 predicts that, relative to fixed-wage-compensation participants, single-measure-compensation participants are more likely to choose modification packages with higher levels of the *sing* measure, at the expense of other, equally important, social-ability measures. Such behavior would be consistent with surrogation.

To test H1, we first calculate participants' *comprehensive* surrogation scores—the percentage of each participant's modification-package choices that are consistent with surrogation with *sing* as described in Section III. Next, we compare the mean *comprehensive* surrogation score between fixed-wage-compensation and single-measure-compensation conditions. At the operational level, H1 is supported if single-measure-compensation participants' surrogation scores are higher than fixed-wage-compensation participants' surrogation scores. Participants' mean *comprehensive* surrogation score is 0.59 in the single-measure-compensation condition, but is only 0.39 in the fixed-wage-compensation condition (see Table 3, Panel A). As we report in Table 3, Panel B, the difference of 0.20 in *comprehensive* surrogation scores across these two conditions is statistically significant (t = 4.56, p < 0.01).

These results are robust to distinguishing surrogation scores in terms of whether there is an opportunity cost associated with selecting a surrogation-consistent modification package. Panel A of Table 3 presents participants' *opportunity cost* and *no opportunity cost* surrogation scores. As reported in Panel B of Table 3, mean *opportunity cost* and *no opportunity cost* surrogation scores

---

[16] Participants in the single-measure-compensation condition designed creatures with a higher level of *sing* (mean = 3.81) than participants in the fixed-wage-compensation condition (mean = 2.96) (t = 2.98, p < 0.01). However, participants in the multiple-measure-compensation condition (mean = 8.88) did not design creatures with higher combined levels of *sing*, *dance*, and *charm* than participants in the fixed-wage-compensation condition (mean = 8.41) (t = 0.80, ns), though the means are directionally consistent with the expected influence of incentives. This latter result likely stems from parameters inherent to the *Spore* software that lead to a small difference between the sum of the *sing*, *dance*, and *charm* levels for a participant maximizing on those measures and a participant maximizing on all four social-ability measures. This constraint is one reason why we elicited participants' self-reported point (i.e., attention) allocations.

[17] Although our hope in using an abstract environment was to limit preconceived notions about appropriate surrogates for strategy, participants nonetheless judged *charm* to be especially important for their creatures' social-ability. On average, participants allocated more points to *charm* than any other social-ability. This works against our finding results, as it adds noise to every condition. It specifically works against our finding support for H2, as an emphasis on *charm* would increase, not decrease, the tendency to use measures as surrogates for the strategic construct in our multiple-measure-compensation condition.

[18] Participants' total compensation did not differ across conditions (F = 0.97, ns). We designed our experiment to maximize the likelihood that, on average, creature design compensation and the additional compensation were approximately equal. The proportion of total compensation stemming from the creature design compensation for the fixed-wage-compensation, single-measure-compensation, and multiple-measure-compensation conditions were 50.5 percent, 43 percent, and 48 percent, respectively. The lower proportion in the single-measure-compensation condition stems from the fact that the mean creature-design compensation was higher in both the multiple-measure-compensation (mean = $8.88) and the fixed-wage-compensation (mean = $10.00) conditions than in the single-measure-compensation condition (mean = $7.62).

TABLE 2

Responsiveness to Incentive Compensation Manipulation

Panel A: Mean Creature Design Point Allocations[a] (Standard Deviation)

	Compensation Condition[b]		
Social-Ability Measures	**Fixed-Wage Compensation (n = 27)**	**Single-Measure Compensation (n = 26)**	**Multiple-Measure Compensation (n = 26)**
Sing	22.4	35.1	26.7
	(7.6)	(17.6)	(7.3)
Dance	23.8	17.9	26.5
	(8.6)	(7.0)	(9.6)
Charm	34.9	34.5	33.0
	(16.9)	(19.1)	(14.5)
Pose	19.2	12.5	13.9
	(8.9)	(7.5)	(9.5)

Panel B: Statistical Tests

	Estimate	**t**	**p-value[c]**
Points Allocated to *sing*: Fixed-Wage versus Single-Measure Compensation	12.7	3.44	< 0.01
Points Allocated to *sing*, *dance*, and *charm*: Fixed-Wage versus Multiple-Measure Compensation	5.1	1.98	0.03

[a] *Creature Design Point Allocations* represent participants' allocations of exactly 100 points across the four social-ability measures—*sing*, *dance*, *charm*, and *pose*—when asked to "indicate how much attention you gave to the [*sing*, *dance*, *charm*, and *pose*] abilities, by allocating 100 points among the four abilities. Allocate more points to abilities to which you gave more attention."

[b] This represents compensation tied directly to social-ability measures of the creatures initially designed by participants. We manipulated this source of compensation at three levels: no social-ability compensation (*fixed-wage-compensation*), performance-based compensation based on a single measure of the creature's social abilities (*single-measure-compensation*), and performance-based compensation based on three of the four measures of the creature's social abilities (*multiple-measure-compensation*). In the fixed-wage-compensation condition, participants received $10 regardless of the state of their creature's social-ability measures. In the single-measure-compensation condition, participants received $2 per level of their creature's *sing* measure (e.g., a level 4 *sing* measure resulted in compensation of $8). In the multiple-measure-compensation condition, participants received $1 per level of their creature's *sing*, *dance*, and *charm* measures (e.g., a creature with *sing* level 3, *dance* level 2, and *charm* level 5 resulted in compensation of $10).

[c] All p-values are presented on a one-tailed basis, given the expected influence of incentive compensation.

are higher in the single-measure-compensation condition than in the fixed-wage-compensation condition (*opportunity cost*: $t = 2.49$, $p = 0.01$; *no opportunity cost*: $t = 4.15$, $p < 0.01$).[19]

Collectively, these results support our hypothesis that compensation on a single measure of a strategic construct increases managers' propensity to use the compensated performance measures as

[19] As discussed earlier, our *opportunity cost* surrogation scores measure participants' tendency to sacrifice a high overall social-ability for higher levels of the compensated measures. However, participants could also have viewed low variance among social abilities as desirable. As such, opportunity cost surrogation may reflect sacrificing low social-ability variance to maximize the compensated measures. For both hypotheses, untabulated results investigating this possibility are inferentially identical to those of our main analyses.

TABLE 3

Tests of Hypotheses

Panel A: Mean Surrogation Scores (Standard Deviation)

	Compensation Condition[a]		
Surrogation Scores	**Fixed-Wage Compensation (n = 27)**	**Single-Measure Compensation (n = 26)**	**Multiple-Measure Compensation (n = 26)**
Surrogation Score: Comprehensive[b]	0.39[e]	0.59	0.46
	(0.12)	(0.19)	(0.25)
Surrogation Score: No Opportunity Cost[c]	0.62[e]	0.84	0.63
	(0.20)	(0.18)	(0.26)
Surrogation Score: Opportunity Cost[d]	0.15[e]	0.34	0.29
	(0.23)	(0.32)	(0.31)

Panel B: Hypothesis Tests

	Estimate	**t**	**p-value[f]**
H1: Fixed Wage versus Single-Measure			
Comprehensive	0.20	4.56	< 0.01
No Opportunity Cost	0.22	4.15	< 0.01
Opportunity Cost	0.19	2.49	0.01
H2: Single-Measure versus Multiple-Measure			
Comprehensive	−0.13	2.15	0.02
No Opportunity Cost	−0.21	3.38	< 0.01
Opportunity Cost	−0.05	0.58	ns

[a] Compensation condition represents compensation tied directly to social-ability measures of the creatures initially designed by participants. We manipulated this source of compensation at three levels: no social-ability compensation (*fixed-wage-compensation*), performance-based compensation based on a single measure of the creature's social abilities (*single-measure-compensation*), and performance-based compensation based on three of the four measures of the creature's social abilities (*multiple-measure-compensation*). In the fixed-wage-compensation condition, participants received $10 regardless of the state of their creature's social-ability measures. In the single-measure-compensation condition, participants received $2 per level of their creature's *sing* measure (e.g., a level 4 *sing* measure resulted in compensation of $8). In the multiple-measure-compensation condition, participants received $1 per level of their creature's *sing*, *dance*, and *charm* measures (e.g., a creature with *sing* level 3, *dance* level 2, and *charm* level 5 resulted in compensation of $10).

[b] *Comprehensive* surrogation scores are the percentage of participants' modification-package choices that are consistent with surrogation with *sing* or with *sing*, *dance*, and *charm*, depending on the compensation condition.

[c] *No Opportunity Cost* surrogation scores are the percentage of participants' modification-package choices that are consistent with surrogation with *sing* or with *sing*, *dance*, and *charm*, depending on the compensation condition, when the total of all four social-ability measure levels does not differ between modification packages.

[d] *Opportunity Cost* surrogation scores are the percentage of participants' modification-package choices consistent with surrogation with *sing* or with *sing*, *dance*, and *charm*, depending on the compensation condition, given that such a choice means forgoing a higher total level of the four social-ability measures.

[e] Because H1 requires a comparison of surrogation with *sing* between the fixed-wage-compensation and single-measure-compensation conditions, the surrogation scores for the fixed-wage-compensation condition for this table are based on surrogation with the *sing* measure.

[f] All p-values are presented on a one-tailed basis, given the directional predictions of H1 and H2.

a surrogate for the strategic construct. Especially interesting is the robustness of these results when considering choices where surrogating with *sing* entails incurring an opportunity cost. Further, these results cannot be explained by a simple propensity toward wealth-maximization, because participants were not paid to maximize a subset of the available measures during this task of the experiment (see Section III).

Primary Tests of Hypothesis 2

H2 predicts that compensation on multiple measures of a strategic construct leads managers to use the compensated measures as surrogates for the strategic construct less than if they are compensated on a single measure. Operationally, this suggests that participants in the multiple-measure-compensation condition will exhibit lower surrogation scores than participants in the single-measure-compensation condition.

In calculating surrogation scores, we focus on the social-ability measures for which participants received compensation while designing their initial virtual creature. Thus, for single-measure-compensation participants, we use the same surrogation scores used to test H1, which are computed based on their surrogation with *sing*. For multiple-measure-compensation participants, we compute surrogation scores based on choices related to the sum of the *sing*, *dance*, and *charm* measure levels. Participants' mean *comprehensive* surrogation score is 0.59 in the single-measure-compensation condition, but is only 0.46 in the multiple-measure-compensation condition (see Table 3, Panel A). As we report in Table 3, Panel B, the difference of 0.13 is significant ($t = 2.15$, $p = 0.02$). These results are robust when considering the *no opportunity cost* surrogation score ($mean_{single} = 0.84$; $mean_{multiple} = 0.63$; $t = 3.38$, $p < 0.01$). With respect to the *opportunity cost* surrogation score, however, the results are not statistically significant ($mean_{single} = 0.34$; $mean_{multiple} = 0.29$; $t = 0.58$, ns). Importantly, the comparison of *no opportunity cost* surrogation scores provides the most conservative test of H2. That is, choices consistent with surrogation are *most* likely to be made when surrogation is *not* costly. Thus, our results suggest that multiple-measure-compensation participants are less likely than single-measure-compensation participants to make such choices, but only in the absence of opportunity costs. These results thereby provide some evidence consistent with H2.

A concern regarding our tests of H2 is that these tests entail using a different subset of modification-package pairings across compensation conditions. That is, the *comprehensive*, *no opportunity cost*, and *opportunity cost* surrogation scores are computed using different subsets of modification-package pairings across the single-measure- and multiple-measure-compensation conditions. To control for this inherent lack of consistency, we also test H2 using pairings that are used to compute surrogation scores for both of these compensation conditions (Pairings #1–8 in Table 1). Based on these alternative measures, untabulated results for the *comprehensive* and *no opportunity cost* surrogation scores are inferentially identical to those initially reported. For the alternative *opportunity cost* surrogation scores, however, we find that the difference between multiple-measure-compensation participants (mean = 0.28) and single-measure-compensation participants (mean = 0.40) is marginally significant ($t = 1.35$, $p = 0.09$). Thus, results holding the modification-package pairings constant across conditions provide additional support for H2.

Supplementary Analyses

As a follow-up analysis, we compare the surrogation scores of participants in the fixed-wage-compensation and the multiple-measure-compensation conditions. To serve as a relevant comparison with multiple-measure-compensation participants in this analysis, we calculate the fixed-wage-compensation condition surrogation scores based on modification-package choices that

TABLE 4

Comparison of Fixed-Wage and Multiple-Measure-Compensation Conditions

Panel A: Mean Surrogation Scores (Standard Deviation)

	Compensation Condition[a]	
Surrogation Scores	**Fixed-Wage Compensation (n = 27)**	**Multiple-Measure Compensation (n = 26)**
Surrogation Score: Comprehensive[b]	0.40[e]	0.46
	(0.14)	(0.25)
Surrogation Score: No Opportunity Cost[c]	0.63[e]	0.63
	(0.23)	(0.26)
Surrogation Score: Opportunity Cost[d]	0.16[e]	0.29
	(0.23)	(0.31)

Panel B: Statistical Tests

Fixed-Wage versus Multiple-Measure	**Estimate**	**t**	**p-value[f]**
Comprehensive	0.06	1.16	ns
Low Opportunity Cost	0.00	0.02	ns
High Opportunity Cost	0.13	1.70	0.10

[a] This represents compensation tied directly to social-ability measures of the creatures initially designed by participants. We manipulated this source of compensation at three levels: no social-ability compensation (*fixed-wage-compensation*), performance-based compensation based on a single measure of the creature's social abilities (*single-measure-compensation*), and performance-based compensation based on three of the four measures of the creature's social abilities (*multiple-measure-compensation*). In the fixed-wage-compensation condition, participants received $10 regardless of the state of their creature's social-ability measures. In the single-measure-compensation condition, participants received $2 per level of their creature's *sing* measure (e.g., a level 4 *sing* measure resulted in compensation of $8). In the multiple-measure-compensation condition, participants received $1 per level of their creature's *sing*, *dance*, and *charm* measures (e.g., a creature with *sing* level 3, *dance* level 2, and *charm* level 5 resulted in compensation of $10).

[b] *Comprehensive* surrogation scores in this table are the percentage of participants' modification-package choices that are consistent with surrogation with *sing*, *dance*, and *charm*.

[c] *No Opportunity Cost* surrogation scores in this table are the percentage of participants' modification-package choices that are consistent with surrogation with *sing*, *dance*, and *charm* when the total of all four social-ability measure levels does not differ between modification packages.

[d] *Opportunity Cost* surrogation scores in this table are the percentage of participants' modification-package choices consistent with surrogation with *sing*, *dance*, and *charm* given that such a choice means forgoing a higher total level of the four social-ability measures.

[e] To compare surrogation with *sing*, *dance*, and *charm* between the fixed-wage-compensation and multiple-measure-compensation conditions, the surrogation scores for the fixed-wage-compensation condition in this table are based on surrogation with *sing*, *dance*, and *charm*.

[f] All p-values are presented on a two-tailed basis, given the lack of directional predictions.

capture surrogation with *sing*, *dance*, and *charm* (i.e., the compensated measures in the multiple-measure-compensation condition).

Participants' mean *comprehensive* surrogation score is 0.46 in the multiple-measure-compensation condition, and is 0.40 in the fixed-wage-compensation condition (see Table 4, Panel A). As we report in Table 4, Panel B, the difference of 0.06 is not significant (t = 1.16, ns, two-tailed). This result is similar for the *no opportunity cost* surrogation score ($\text{mean}_{\text{fixed-wage}}$ =

0.63; $mean_{multiple} = 0.63$; $t = 0.02$, ns, two-tailed). With respect to the *opportunity cost* surrogation score, the difference between the fixed-wage-compensation condition (mean = 0.16) and the multiple-measure-compensation condition (mean = 0.29) is marginally significant ($t = 1.70$, $p = 0.10$, two-tailed). These findings are consistent with the primary tests of H2, where both the *comprehensive* and *no opportunity cost* surrogation scores decreased when going from single- to multiple-measure-compensation, but the *opportunity cost* surrogation score did not reflect a significant decrease.

To further explore the surrogation phenomenon, we investigate participants' responses to a post-experimental question eliciting their allocation of attention among the four measures of social-ability while making modification-package choices. Similar to the manipulation-check question regarding the creature-design task described earlier, we asked each participant to "indicate how much attention you gave to the [*sing*, *dance*, *charm*, and *pose*] abilities, by allocating 100 points among the four abilities, [allocating] more points to abilities to which you gave more attention" when making their modification-package selections. Our findings, reported in Table 5, suggest that single-measure-compensation participants allocated more attention to the *sing* measure when they selected modification packages than did fixed-wage-compensation participants (mean = 23.7 percent) ($t = 2.52$, $p < 0.01$). Importantly, multiple-measure-compensation participants *did not* allocate significantly more points to the *sing*, *dance*, and *charm* measures (mean = 83.7 percent) than did the fixed-wage-compensation participants (mean 80.0 percent) ($t = 1.57$, ns, two-tailed). This latter finding is particularly notable given evidence reported earlier that participants in the multiple-measure-compensation condition reported allocating *more* attention to the *sing*, *dance*, and *charm* measures than fixed-wage-compensation participants during the initial creature-design task, for which multiple-measure-compensation participants earned incentive compensation. This suggests that while participants in the multiple-measures-compensation setting responded in a wealth-maximizing manner to their incentives in the creature-design task, these incentives did not lead to a significant increase in surrogation for these participants.

Follow-Up Experiment

As discussed earlier, an important aspect of the internal validity of our main experiment relates to our choice not to compensate participants for their modification-package selections. While appropriate for our research purpose, this choice raises questions related to external validity. Therefore, we ran a follow-up experiment to help address this question. Specifically, we investigate whether our results hold if participants are compensated for their modification-package selections.

Our follow-up experiment was identical to our main experiment, with three exceptions. First, the follow-up experiment was 1×2, focusing only on the single-measure-compensation and the multiple-measure-compensation conditions, in which we expected surrogation to be most prevalent. Second, participants earned compensation for their modification-package selections. We informed participants that this additional compensation would be based on their ability to implement the strategy, and that this compensation would be determined in a similar manner to the compensation that participants earned for their creature's ability to achieve global domination. Specifically, each participant's performance during the modification-package selection task was compared to that of nine other randomly selected participants within the same experimental condition. The best performer within each ten-person group received $20, the next best performer received $18, etc. Third, we added a post-experimental question assessing participants' understanding of the compensation they earned for their modification-package selections. Participants answered a multiple choice question indicating whether their compensation for their modification-package selections was determined "Based on [their] ability to implement the strategy," "Based on [their] ability to maximize *sing*," or "Based on [their] ability to maximize *sing*, *dance*, and *charm*."

TABLE 5

Modification-Package Choice Point Allocations

Panel A: Mean Modification-Package Choice Point Allocations[a] (Standard Deviation)

	Compensation Condition[b]		
Social-Ability Measures	**Fixed-Wage Compensation (n = 27)**	**Single-Measure Compensation (n = 26)**	**Multiple-Measure Compensation (n = 26)**
Sing	23.7	33.5	22.2
	(6.4)	(19.2)	(10.3)
Dance	24.3	18.6	23.0
	(7.6)	(8.7)	(13.7)
Charm	32.0	33.5	38.5
	(14.1)	(20.5)	(24.8)
Pose	20.4	14.4	16.3
	(7.7)	(9.0)	(9.4)

Panel B: Statistical Tests

	Estimate	**t**	**p-value[c]**
Points Allocated to *sing*: Fixed-Wage versus Single-Measure Compensation	9.8	2.52	< 0.01
Points Allocated to *sing*, *dance*, and *charm*: Fixed-Wage versus Multiple-Measure Compensation	3.7	1.57	ns

[a] *Modification-Package Choice Point Allocations* represent participants' allocations of exactly 100 points across the four social-ability measures—*sing*, *dance*, *charm*, and *pose*—when asked to "indicate how much attention you gave to the [*sing*, *dance*, *charm*, and *pose*] abilities, by allocating 100 points among the four abilities, [allocating] more points to abilities to which you gave more attention" when making modification-package selections.

[b] This represents compensation tied directly to social-ability measures of the creatures initially designed by participants. We manipulated this source of compensation at three levels: no social-ability compensation (*fixed-wage-compensation*), performance-based compensation based on a single measure of the creature's social abilities (*single-measure-compensation*), and performance-based compensation based on three of the four measures of the creature's social abilities (*multiple-measure-compensation*). In the fixed-wage-compensation condition, participants received $10 regardless of the state of their creature's social-ability measures. In the single-measure-compensation condition, participants received $2 per level of their creature's *sing* measure (e.g., a level 4 *sing* measure resulted in compensation of $8). In the multiple-measure-compensation condition, participants received $1 per level of their creature's *sing*, *dance*, and *charm* measures (e.g., a creature with *sing* level 3, *dance* level 2, and *charm* level 5 resulted in compensation of $10).

[c] All p-values are presented on a one-tailed basis, except for comparisons involving the fixed-wage-compensation and multiple-measure-compensation conditions, which are presented on a two-tailed basis.

Table 6 presents the results of the additional experiment. As shown in Panel A, participants' mean *comprehensive* surrogation score is 0.61 (n = 24) in the single-measure-compensation condition, but is only 0.46 (n = 26) in the multiple-measure-compensation condition. As we report in Panel B, the difference of 0.15 is statistically significant ($t = 2.31$, $p < 0.01$). These results are robust when considering the *no opportunity cost* surrogation score ($mean_{single} = 0.84$; $mean_{multiple} = 0.66$; $t = 2.70$, $p < 0.01$). When considering the *opportunity cost* surrogation score, the difference between conditions is marginally significant ($mean_{single} = 0.38$; $mean_{multiple} = 0.25$; $t = 1.35$, $p = 0.09$). When we drop from our analyses six participants who incorrectly answered the question in the post-experimental questionnaire about the basis of their modification-package-selection

TABLE 6

Additional Experiment

Panel A: Mean Surrogation Scores (Standard Deviation)

	Compensation Condition[a]	
Surrogation Scores	**Single-Measure Compensation (n = 24)**	**Multiple-Measure Compensation (n = 26)**
Surrogation Score: Comprehensive[b]	0.61	0.46
	(0.24)	(0.23)
Surrogation Score: No Opportunity Cost[c]	0.84	0.66
	(0.22)	(0.26)
Surrogation Score: Opportunity Cost[d]	0.38	0.25
	(0.34)	(0.32)

Panel B: Statistical Tests

Single-Measure versus Multiple-Measure	**Estimate**	**t**	**p-value[e]**
Comprehensive	0.15	2.31	0.01
No Opportunity Cost	0.21	2.70	0.01
Opportunity Cost	0.13	1.35	0.09

[a] This represents compensation tied directly to social-ability measures of the creatures initially designed by participants. We manipulated this source of compensation at two levels: performance-based compensation based on a single measure of the creature's social abilities (*single-measure-compensation*) and performance-based compensation based on three of the four measures of the creature's social abilities (*multiple-measure-compensation*). In the single-measure-compensation condition, participants received $2 per level of their creature's *sing* measure (e.g., a level 4 *sing* measure resulted in compensation of $8). In the multiple-measure-compensation condition, participants received $1 per level of their creature's *sing*, *dance*, and *charm* measures (e.g., a creature with *sing* level 3, *dance* level 2, and *charm* level 5 resulted in compensation of $10).

[b] *Comprehensive* surrogation scores are the percentage of participants' modification-package choices that are consistent with surrogation with *sing* or with *sing*, *dance*, and *charm*, depending on the compensation condition.

[c] *No Opportunity Cost* surrogation scores are the percentage of participants' modification-package choices that are consistent with surrogation with *sing* or with *sing*, *dance*, and *charm*, depending on the compensation condition, when the total of all four social-ability measure levels does not differ between modification packages.

[d] *Opportunity Cost* surrogation scores are the percentage of participants' modification-package choices consistent with surrogation with *sing* or with *sing*, *dance*, and *charm*, depending on the compensation condition, given that such a choice means forgoing a higher total level of the four social-ability measures.

[e] All p-values are presented on a one-tailed basis, given the directional predictions implied by our theory.

compensation, all results are inferentially identical, except that the *opportunity cost* results become more significant ($t = 2.14$, $p = 0.02$). These results support our theory and corroborate the findings from our main experiment.

V. CONCLUSION

This paper investigates the influence of incentive compensation on managers' propensity to use compensated measures as surrogates for strategic constructs. Our findings suggest that compensating managers based on a single measure of a strategic construct increases their propensity to use the compensated measure as a surrogate for the construct. Our results further suggest that this propensity to use compensated measures as surrogates for strategy is less

pronounced when compensating managers on multiple measures of a strategic construct. Our paper contributes to a variety of research streams, including literature on firms' use of strategic performance measurement systems. Specifically, we highlight managers' tendency to use measures as surrogates for strategic constructs as a potentially detrimental response to performance measure information. Additionally, our study helps develop academic and practitioner understanding of factors that influence the extent to which firms incur this potential cost. Thus, our study is important to practicing managers and management accountants responsible for developing and implementing strategic performance measurement systems.

Notably, by finding support for both of our hypotheses, we are able to rule out a number of alternative explanations for our findings. For instance, some participants may have been confused by the incomplete compensation contracts, and believed that compensated measures were more important to their creature's ability to socialize. Such a belief would lead participants to make modification-package selections consistent with, but not driven by, surrogation. Another potential alternative explanation for results consistent with H1 is that some participants may have simply tried to make modification-package-pairing selections that were consistent with their behavior in the creature design task (Abelson et al. 1968). Motivated reasoning (Kunda 1990; Tayler 2010; Luft et al. 2011) and other psychology-based theories could also explain H1 results. Participants might identify reasons why surrogation-consistent modification packages, which also happen to be consistent with participants' emphasis when designing their creature, are superior. However, these alternative explanations cannot explain the full pattern of our findings. In particular, because they would predict high surrogation across both incentive compensation conditions, they cannot explain the increased surrogation of the strategic construct when compensation is tied to a single measure, and the *decreased* surrogation of the strategic construct when compensation is tied to multiple measures.

Our paper is subject to multiple limitations. First, while we view the rich, but non-business oriented setting as a strength of our design, the setting raises questions of boundary conditions and generalizability. For instance, we develop and test hypotheses referring to compensation based on a "single measure" or based on "multiple measures." While we operationalize these conditions using specific numbers of measures, we appreciate the continuous nature of the number of possible measures that may be included in a strategic performance measurement system. The number of measures at which surrogation begins to be mitigated is likely to depend on a variety of factors that are beyond the scope of this paper. Finally, while we draw on a vast, well-established literature relating to attention, our supplemental measures of attention are self-reported and potentially very noisy. Although we believe the goals of our study necessitate this approach, future research may benefit from utilizing more precise measures of attention.

While our focus in this paper has been the effects of compensation on strategy surrogation, future research could investigate other institutional factors that inhibit or exacerbate surrogation. Prior research identifies various measure attributes (e.g., congruence, precision, and sensitivity) that affect the optimal weighting of performance measure in the compensation contract. These attributes may also influence managers' propensity to use measures as surrogates for a strategic construct (Feltham and Xie 1994; Krishnan et al. 2005). In addition, managers' involvement in the choice of strategy may influence their propensity toward surrogation. Moreover, the extent to which surrogation is lower under multiple-measure-compensation than under single-measure-compensation may depend on the relationship among the multiple measures (e.g., extent of overlap, causal relationships). Finally, future research could explore specific implications of surrogation, such as how and under what circumstances costs of surrogation arise (e.g., managers' failure to update obsolete measures, sub-optimal strategic decisions), how the representativeness of a given measure influences the cost of surrogation with that measure, and whether strategy surrogation affects managers' evaluations of subordinates. Research addressing such issues would contribute greatly to

our understanding of the surrogation phenomenon, as well as the influence of surrogation on the effectiveness of strategic performance measurement systems.

REFERENCES

Abelson, R. P., E. Aronson, W. J. McGuire, T. Newcomb, M. J. Rosenberg, and P. H. Tannenbaum, eds. 1968. *Theories of Cognitive Consistency: A Sourcebook*. Chicago, IL: Rand McNally.

Alter, A. L., D. M. Oppenheimer, N. Epley, and R. N. Eyre. 2007. Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General* 136 (4): 569–576.

Atkinson, A. A., J. H. Waterhouse, and R. Wells. 1997. A stakeholder approach to strategic performance measurement. *Sloan Management Review* 38 (3): 25–37.

Baker, G. 1992. Incentive contracts and performance measurement. *Journal of Political Economy* 100 (3): 598–614.

Baker, G. 2000. The use of performance measures in incentive contracting. *American Economic Review* 90 (2): 415–420.

Banker, R. D., G. Potter, and D. Srinivasan. 2000. An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review* 75 (1): 65–92.

Bonner, S. E. 1994. A model of the effects of audit task complexity. *Accounting, Organizations and Society* 19 (3): 213–234.

Bonner, S. E., and G. B. Sprinkle. 2002. The effect of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27 (4-5): 303–345.

Brickley, J. A., and J. L. Zimmerman. 2001. Changing incentives in a multitask environment: Evidence from a top-tier business school. *Journal of Corporate Finance* 7: 367–396.

Bromwich, M. 1990. The case for strategic management accounting: The role of accounting information for strategy in competitive markets. *Accounting, Organizations and Society* 15 (1-2): 27–46.

Budde, J. 2007. Performance measure congruity and the balanced scorecard. *Journal of Accounting Research* 45 (3): 515–539.

Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (1): 99–126.

Bushman, R. M., and R. J. Indjejikian. 1993. Stewardship value of "distorted" accounting disclosures. *The Accounting Review* 68 (4): 765–782.

Campbell, D. J. 1988. Task complexity: A review and analysis. *The Academy of Management Review* 13 (1): 40–52.

Campbell, D., S. M. Datar, S. L. Kulp, and V. G. Narayanan. 2008. *Testing Strategy with Multiple Performance Measures: Evidence from a Balanced Scorecard at Store 24*. Working Paper, Harvard University.

Chaiken, S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* 39 (5): 752–766.

Chen, C. X., M. Martin, and K. A. Merchant. 2010. *The Effect of Measurement Timing on the Predictive Abilities of Customer Satisfaction Measures*. Working paper, University of Illinois at Urbana–Champaign, Arizona State University, and University of Southern California.

Chenhall, R. H. 2003. Management control systems design within its organizational context: Findings from contingency-based research and directions for the future. *Accounting, Organizations and Society* 28: 127–168.

Electronic Arts. 2008. *Spore*. Redwood City, CA: Electronic Arts, Inc.

Epstein, M., and J. F. Manzoni. 1997. The balanced scorecard and tableau de bord: Translating strategy into action. *Management Accounting* 79 (2): 28–36.

Farrell, A. M. 2009. *Effects of Multiple Performance Measures, Multicollinearity, and Task Structure on Individuals' Judgment Performance*. Working paper: Miami University of Ohio.

Feltham, G. A., and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69 (3): 429–453.

Fiske, S. T., and S. E. Taylor. 1991. *Social Cognition*. New York, NY: McGraw-Hill.

Flinn, R. E., and D. L. Crumbley. 2009. *Measure Learning Rather Than Satisfaction in Higher Education*. Sarasota, FL: American Accounting Association.

Frederickson, J. R., S. A. Peffer, and J. Pratt. 1999. Performance evaluation judgments: Effects of prior experience under different performance evaluation schemes and feedback frequencies. *Journal of Accounting Research* 37 (1): 151–165.

Gigerenzer, G., and D. G. Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychology Review* 103: 650–669.

Goldratt, E. M. 2004. *The Goal: A Process of Ongoing Improvement*. Burlington, VT: Gower.

Grizzle, G. A. 2002. Performance measurement and dysfunction: The dark side of quantifying work. *Public Performance and Management Review* 25 (4): 363–369.

Hicks, J. 1939. *Value and Capital*. Oxford, U.K.: University Press.

Holmstrom, B., and P. Milgrom. 1991. Multitask principal agent analyses—Incentive contracts, asset ownership, and job design. *Journal of Law Economics & Organization* 7: 24–52.

Ijiri, Y. 1967. *The Foundations of Accounting Measurement: A Mathematical, Economic, and Behavioral Inquiry*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Ijiri, Y. 1975. *Theory of Accounting Measurement*. Sarasota, FL: American Accounting Association.

Ittner, C. D., and D. F. Larcker. 1998. Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research* 10: 205–238.

Ittner, C. D., and D. F. Larcker. 2001. Assessing empirical research in managerial accounting: A value-based management perspective. *Journal of Accounting and Economics* 32 (1-3): 349–410.

Ittner, C. D., and D. F. Larcker. 2003. Coming up short on nonfinancial performance measurement. *Harvard Business Review* (November): 88–95.

James, W. 1890. *The Principle of Psychology*. New York, NY: Henry Holt.

Johnson, M. D. 1988. Comparability and hierarchical processing in multialternative choice. *Journal of Consumer Research* 15 (3): 303–314.

Kahneman, D., and S. Frederick. 2002. Representiveness revisited: Attribute substitution in intuitive judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman. New York, NY: Cambridge University Press.

Kahneman, D., A. B. Krueger, D. Schkade, N. Schwarz, and A. A. Stone. 2006. Would you be happier if you were richer? A focusing illusion. *Science* 312 (5782): 1908–1910.

Kaplan, R. S., and D. P. Norton. 1992. The Balanced Scorecard—Measures that drive performance. *Harvard Business Review* 70 (1): 71–79.

Kaplan, R. S., and D. P. Norton. 1996. *The Balanced Scorecard: Translating Strategy Into Action*. Boston, MA: Harvard Business School Press.

Kaplan, R. S., and D. P. Norton. 2000. Having trouble with your strategy? Then map it. *Harvard Business Review* 78 (5): 167–176.

Kerr, S. 1975. On the folly of rewarding A, while hoping for B. *Academy of Management Journal* 18 (4): 769–783.

Krishnan, R., J. L. Luft, and M. D. Shields. 2005. Effects of accounting-method choices on subjective performance-measure weighting decisions: Experimental evidence on precision and error covariance. *The Accounting Review* 80 (4): 1163–1192.

Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480–498.

Lambert, R. A. 2001. Contracting theory and accounting. *Journal of Accounting and Economics* 32 (1-3): 3–87.

Langfield-Smith, K. 1997. Management control systems and strategy: A critical review. *Accounting, Organizations and Society* 22 (2): 207–232.

Lawton, C. 2008. Electronic Arts bets big on new game. *Wall Street Journal* (September 2): B1.

Libby, R. 1981. *Accounting and Human Information Processing: Theory and Applications*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Libby, R., and H.-T. Tan. 1999. Analysts' reactions to warnings of negative earnings surprises. *Journal of Accounting Research* 37 (2): 415–435.

Luft, J., and M. D. Shields. 2003. Mapping management accounting: Graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations, and Society* 28 (2-3): 169–249.

Luft, J., M. D. Shields, and T. F. Thomas. 2011. *Performance Measures, Motivated Reasoning, and Subjective Performance Evaluation*. Working paper, Michigan State University.

Malina, M. A., and F. H. Selto. 2001. Communicating and controlling strategy: An empirical study of the effectiveness of the Balanced Scorecard. *Journal of Management Accounting Research* 13: 47–90.

Malina, M. A., and F. H. Selto. 2004. Choice and change of measures in performance measurement models. *Management Accounting Research* 15: 441–489.

Malina, M. A., H. S. O. Nørreklit, and F. H. Selto. 2007. Relations among measures, climate of control, and performance measurement models. *Contemporary Accounting Research* 24 (3): 935–982.

Markman, A. B., and D. L. Medin. 1995. Similarity and alignment in choice. *Organizational Behavior and Human Decision Processes* 63 (2): 117–130.

Payne, J. W., J. R. Bettman, and E. J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge, U.K.: Cambridge University Press.

Runkel, P. J., and J. E. McGrath. 1972. *Research on Human Behavior: A Systematic Guide to Method*. New York, NY: Holt, Rinehart, and Winston, Inc.

Schipper, K., and L. Vincent. 2003. Earnings quality. *Accounting Horizons* 17 (Supplement): 97–110.

Schkade, D. A., and D. Kahneman. 1998. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science* 9:340–346.

Schwarz, N. 2004. Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology* 14 (4): 332–348.

Shah, A. K., and D. M. Oppenheimer. 2011. Grouping information for judgments. *Journal of Experimental Psychology: General* 140: 1–13.

Shiffrin, R. M. 1988. Attention. In *Stevens' Handbook of Experimental Psychology*, edited by R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce. New York, NY: Wiley.

Simons, R. 1991. Strategic orientation and top management attention to control systems. *Strategic Management Journal* 12: 49–62.

Simons, R. 1994. How new top managers use control systems as levers of strategic renewal. *Strategic Management Journal* 15: 169–189.

Strack, F., L. L. Martin, and N. Schwarz. 1988. Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology* 18: 429–442.

Tayler, W. B. 2010. The Balanced Scorecard as a strategy-evaluation tool: The effects of responsibility and causal-chain focus. *The Accounting Review* 85 (3): 1095–1117.

Tayler, W. B., and R. J. Bloomfield. 2011. Norms, conformity, and controls. *Journal of Accounting Research* 49 (3): 753–790.

Tversky, A., and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5: 207–232.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1130.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84 (4): 327–352.

Webb, R. A. 2004. Managers' commitment to the goals contained in a strategic performance measurement system. *Contemporary Accounting Research* 21 (4): 925–958.

Williamson, O. E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York, NY: Free Press.

Wood, R. E. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes* 37: 60–82.

Zhang, S., and A. B. Markman. 2001. Processing product unique features: Alignability and involvement in preference construction. *Journal of Consumer Psychology* 11 (1): 13–27.