



Marketing Science

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

Enduring Effects of Goal Achievement and Failure Within Customer Loyalty Programs: A Large-Scale Field Experiment

Yanwen Wang, Michael Lewis, Cynthia Cryder, Jim Sprigg

To cite this article:

Yanwen Wang, Michael Lewis, Cynthia Cryder, Jim Sprigg (2016) Enduring Effects of Goal Achievement and Failure Within Customer Loyalty Programs: A Large-Scale Field Experiment. Marketing Science 35(4):565-575. <https://doi.org/10.1287/mksc.2015.0966>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Enduring Effects of Goal Achievement and Failure Within Customer Loyalty Programs: A Large-Scale Field Experiment

Yanwen Wang

Leeds School of Business, University of Colorado Boulder, Boulder, Colorado 80309, yanwen.wang@colorado.edu

Michael Lewis

Emory University, Atlanta, Georgia 30322, mike.lewis@emory.edu

Cynthia Cryder

Washington University in St. Louis, St. Louis, Missouri 63130, cryder@wustl.edu

Jim Sprigg

Director of Database Marketing, IHG, Atlanta, Georgia 30346, jim.sprigg@ihg.com

This research investigates whether there are enduring effects of goal achievement and failure within customer loyalty promotion programs. We collaborated with a major hotel chain to launch a large scale field experiment involving 95,532 existing loyalty customers. We observed customers' hotel stays for eight months before the experiment, eight months during the experiment, and eight months after the experiment. Customers in the treatment group were asked to increase their hotel nights during the 8-month promotion by a set percentage relative to their baseline to receive a reward. Overall, the promotion led to increased purchasing in the post-promotion period. However, only 20% of customers successfully reached the goal whereas 80% missed the goal. We use a propensity score analysis to examine the distinct effects of goal achievement versus goal failure. Results show that goal attainment significantly increased post-promotion purchasing whereas goal failure significantly reduced post-promotion purchasing. Additionally, we use econometric methods to empirically test a behavioral theory of relationship-based reciprocity. We find that customers in a high status tier relationship, with the most invested in the firm, are most affected by goal failure whereas customers in a low status tier relationship, with the least invested in the firm, are most affected by goal success. Because the type of loyalty program described in this paper is widely used in a variety of industries the findings suggest that marketers should set reachable goals within loyalty promotion programs. Firms should be particularly cautious about the impact of goal failures for the firm's most loyal customers.

Data, as supplemental material, are available at <http://dx.doi.org/10.1287/mksc.2015.0966>.

Keywords: loyalty promotion program; goal achievement; goal failure; field experiment

History: Received: May 6, 2013; accepted: August 11, 2015; Preyas Desai served as the editor-in-chief and Barbara Kahn served as associate editor for this article. Published online in *Articles in Advance* January 21, 2016.

1. Introduction

Customer loyalty programs have become popular across a variety of industries. Two central dimensions of these programs include (1) thresholds that dictate success or failure for reaching reward-contingent goals, and (2) status segments based on levels of previous purchasing. While these dimensions are critical to loyalty program design (Breugelmans et al. 2015), there is little research on the consequences of receiving rewards (Dreze and Nunes 2011), the response of loyalty segments to reward programs, and (to our knowledge) no research about failing to receive rewards or the interplay among these factors.

The current research attempts to fill this gap. We report results of a large scale field experiment con-

ducted by a major hotel chain. The experiment systematically varied the thresholds required to earn extra reward points among two test groups and a control group. In addition, the experiment was conducted using members with different loyalty program status levels. Of particular significance, the study design afforded the opportunity to observe these customers in the eight months before the experiment, the eight months of the experiment, and for eight months following the experiment.

In addition to the direct effect of the promotion, we investigate the enduring impact of reward success and failure. This research question is difficult because reward success and failure are determined not through an experimental manipulation, but based on how consumers respond to the marketing intervention. To

address this type of selected response we use a combination of propensity scoring analysis and difference-in-differences modeling techniques in addition to control group comparisons to isolate the true impact of the loyalty-based treatments. We supplement this analysis with a panel Tobit model to further control for individual heterogeneity and to investigate the behavioral drivers of the long-term consequences of success and failure.

We find that failing to earn the promotional reward can significantly degrade customer relationship outcomes as evidenced by a reduction in hotel nights compared to the control group. Analogously, we observe that earning the reward enhances outcomes as indicated by a significant increase in hotel nights compared to the control group. Interestingly, we find that these effects interact with customer loyalty status level such that the most loyal customers (“Platinum”) are the least influenced by goal success but are the most influenced by goal failure.

Our conjecture is that these effects are driven by relationship-specific reciprocity norms; different customer segments (e.g., Platinum, Gold, and Club members) have different reciprocity expectations based on their history with the firm. For example, customers in a high tier relationship have invested heavily in the firm and thus might expect large returns; they therefore can be most affected by goal failure and subsequent denial of rewards. By contrast, customers in a lower tier have invested minimally in the firm and thus might expect few benefits; they therefore can be most affected by goal success and receipt of rewards.

2. Background and Hypothesis

Over the last two decades an academic literature focused on documenting and explaining consumer response to loyalty programs has grown. The early literature questioned the effectiveness of loyalty programs themselves (Dowling and Uncles 1997). A second wave of research focused on measurement (Lewis 2004) and psychological response to loyalty program design dimensions (Kivetz and Simonson 2002, Kivetz et al. 2006, Nunes and Dreze 2006, Roehm et al. 2002). The literature has begun to consider segment level responses (Liu 2007, Dreze and Nunes 2009).

Our research focuses on the effects of successful and unsuccessful goal attainment within a customer loyalty program. This is a challenging area of inquiry for several reasons. First, although archival data provides observations of reward achievement and failure, the lack of a control group unexposed to a reward opportunity means that it is difficult to make inferences about whether a drop or acceleration in buying is due to the reward event or if consumer demand changes are exogenous. Second, even controlled experiments

may be inadequate. While it is possible to design experiments that vary loyalty program availability and program dimensions, success or failure is an outcome of how consumers respond to the program rather than a direct result of the experimental manipulation. To address these challenges, we rely on multiple methods including a controlled field experiment (Montaguti et al. 2016, Tucker and Zhang 2010) and propensity scoring to isolate consequences of reward attainment and failure.

Use of field data creates a third challenge. Because the data are collected in the course of a commercial operation, we cannot collect psychological measures. Therefore, to investigate the underlying psychological drivers of response to reward success and failure, we rely on carefully designed econometric specifications.

Next we discuss several branches of literature that address response to dynamic incentive schemes. Specifically, we consider the consequences of goal success and failure, and how reciprocity expectations or reinforcement learning (or both) may contribute to goal success and failure outcomes.

2.1. Goals: Success and Failure

A host of psychological research documents the ability of goals and incentives to motivate effort (e.g., Gollwitzer 1999, Heath et al. 1999, Locke and Latham 1990, Jhang and Lynch 2015). Evidence within the domain of customer loyalty programs is consistent with this literature, finding, for example, that incentivized goals increase consumers’ purchase rates and wallet share (Lewis 2004, Kivetz et al. 2006, Taylor and Neslin 2005).

Research documenting behavior following goal success or failure is relatively scarce. However, one recent investigation found a positive effect of goal achievement such that goal achievement within ongoing customer loyalty programs allows consumers to learn about their own self-efficacy, thus increasing efforts toward future recurring rewards (Dreze and Nunes 2011). Research in marketing and psychology finds generally negative consequences of goal failure. Goal failure can result in subsequent worse performance (Soman and Cheema 2004); once a goal has been violated, further actions in violation of the goal can follow (Cochran and Tesser 1996).

In this investigation, we test whether similar effects of goal achievement and goal failure exist in a real one-time customer loyalty promotion. Based on previously documented effects of goal success and failure, we hypothesized the following:

HYPOTHESIS 1 (H1). *There will be a positive (negative) enduring effect on hotel stays for customers who succeed (fail) to receive loyalty program rewards.*

Although the previous literature described above leads to hypothesized positive effects of goal success

and negative effects of goal failure, note that the marketing literature shows some contrary findings within more applied research domains. Loyalty programs are not the only managerial practice that operate via dynamic incentives. For example, sales control or quota systems share similar characteristics. Customer loyalty programs provide benefits to customers who make a sufficient number of purchases within a specified time period. Similarly, outcome-based sales control systems provide benefits based on meeting sales targets within a specified time period (Anderson and Oliver 1987).

This literature has observed that these “outcome-based” control systems can have enduring negative effects on salespeople (Oliver and Anderson 1994) by placing a focus on incentives (Deci et al. 1999), and subsequently weakening organizational commitment and job satisfaction. Similar arguments exist in the sales promotion literature, contending that discounting may diminish brand equity because consumers attribute their purchase and consumption decisions to the promotion rather than the brand value (see Blattberg and Neslin 1990, Chapter 2).

Yet another possibility is that consumers will exhibit behavior that looks rational and strategic, yielding minimal change in long-term customer loyalty. A potential reaction to time constrained promotions is that consumers (or salespeople) change their behaviors to concentrate purchases (or sales) only during the time of the promotion to succeed in the promotion. This type of phenomena has been noted to the sales promotion literature (Blattberg and Neslin 1990) and the sales control literature (Oliver and Anderson 1994, Schwepker and Good 1999, Badovick et al. 1992, Brown et al. 2005). Purchase acceleration during the promotion period and subsequent post-promotion drops have been noted in the loyalty program literature as well (Lewis 2004). The specifications of Lewis (2004) and Kopalle et al. (2012) are consistent with the notion of consumers as rational forward-looking agents for whom reward success or failure have no impact on preference levels.

In sum, based on previous findings in psychology and marketing we anticipate that goal success will positively impact future purchasing and goal failure will adversely impact future purchasing. However, we acknowledge that other patterns, including negative overall effects due to decreased brand loyalty or no long-term effects due to strategic calculation, are also possible.

2.2. Process Exploration

We also examined two potential processes that are consistent with the hypothesized goal success and failure effects: (1) Relationship-based reciprocity, and (2) reinforcement learning.

2.2.1. Relationship-Based Reciprocity. Kumar and Shah (2004) suggest that rewards engender loyalty by establishing beneficial reciprocity between customers and the company. However, the most common type of loyalty program reward structure, i.e., contingent rewards, also allows denying rewards to customers who do not reach goal thresholds. Such goal failure and subsequent denial of rewards may trigger negative reciprocity between customers and the firm. Based on the basic premise of reciprocity (Fehr and Gächter 2000), we would expect that receiving rewards could trigger positive reciprocity and increased purchasing among customers, whereas withholding rewards could trigger negative reciprocity and decreased purchasing.

More specifically, these effects may be asymmetric within consumer segment levels. Because of heavy prior investment, the firm’s most loyal customers may view receipt of rewards as an appropriate return on their investment, resulting in minimal subsequent behavior change. At the same time, due to the same heavy prior investment, high-tier customers may view withholding rewards as a reciprocity violation, maximally influencing subsequent behavior change in the case of goal failure. By comparison, a firm’s least loyal consumers have little invested in the firm and may expect little, therefore, yielding positive reciprocity obligations when rewards are received. At the same time, because of the minimal prior investment, withholding rewards is unlikely to be viewed as a reciprocity violation. Based on this notion of relationship-specific reciprocity expectations, we hypothesized the following:

HYPOTHESIS 2A (H2A). Reward denial will be most influential for high tier (“Platinum” level) customers whereas reward receipt will be most influential for low tier (“Base” level) customers.

2.2.2. Reinforcement Learning. Recent evidence shows that goal achievement within loyalty programs exerts a positive influence on subsequent purchases by allowing consumers to learn about their own self-efficacy, thus increasing efforts toward future recurring goals (Dreze and Nunes 2011). In a one-time reward program such as the one in this investigation, however, subsequent rewards are not in play and therefore standard learning and application of that learning are unlikely to apply.

A specific type of learning, called reinforcement learning, could still take place, however, in such one-time reward programs. Reinforcement learning, a component of operant conditioning, occurs when people identify contingencies between behavior and positive or negative outcomes (Skinner 1938, 1969). When an association between a behavior and a positive outcome arises, the incidence of the original behavior will increase as a means to increase the positive outcome. If, over time, the reward is no longer received as a consequence

of the behavior, the association will weaken and the behavior will cease (Guttman 1953).

Within one-time reward programs, a reward provides positive reinforcement for purchasing, which could subsequently increase purchasing behavior. Over time, when the reward disappears, the association between reward and purchasing would weaken, and heightened purchasing behavior will end. This leads us to the following hypothesis:

HYPOTHESIS 2B (H2B). *Reward success will lead to increased purchasing, however, over time when rewards are no longer offered, purchasing will revert to previous levels.*

3. Field Experiment and Results

The foundation of our research is a field experiment conducted by a major international hotel chain. The experiment involved a short-term loyalty-based promotion offered within the structure of the normal loyalty program. For an 8-month period, loyalty program members in the treatment conditions received an opportunity to earn bonus points for meeting a variety of program conditions.

In our field experiment, a total of 48,270 existing loyalty members were randomly assigned to the treatment groups and provided with an opportunity to earn bonus points during an 8-month period. There were also 47,257 existing loyalty members in the control group. We observed customers' hotel stays during the eight months before the experiment, the eight months of the experiment, and the eight months following the experiment.

Our focus here is the impact of succeeding or failing to reach reward thresholds on subsequent customer response. The experiment involved three fully randomized experimental conditions, i.e., a control condition and two levels of threshold conditions. In the LOW threshold condition, customers were asked to increase their activity by a set percentage relative to the baseline period. In the HIGH threshold condition, customers were asked to increase their activity by doubling the percentage in the low condition.¹ Managerial insights were an important input to decide the threshold levels. Managers at the hotel chain believed that their average share of wallet was below 50% for the sample of members in the experiment. Therefore the LOW and HIGH threshold levels were chosen to not more than double baseline activity. For example, if the threshold rates were set at a 20% increase for the LOW condition and 40% in the HIGH condition, a customer staying 10 nights in the baseline period would need to stay 12 nights to qualify for bonus points in the LOW threshold condition or 14 nights if the customer was in the HIGH threshold condition.

¹ The firm has requested that we not disclose the specific threshold levels used in the experiment. We also orthogonally manipulated incentive magnitude, a factor that is omitted due to space constraints.

Table 1 Summary on Hotel Nights

	Control group	Low threshold	High threshold
No. of customers	47,257	24,082	24,188
Hotel nights			
8-months prior to the experiment	21.52 (26.90)	22.91 (28.15)	22.78 (26.70)
8-months of the experiment	19.11 (28.53)	21.92 (29.60)	21.59 (26.65)
8-months following the experiment	15.38 (24.54)	17.62 (25.82)	17.42 (25.99)
Revenues (\$)			
8-months prior to the experiment	2,010 (2,673)	2,193 (4,400)	2,157 (2,690)
8-months of the experiment	1,830 (2,729)	2,110 (2,904)	2,088 (2,682)
8-months following the experiment	1,607 (2,573)	1,849 (2,760)	1,839 (2,913)
Stays (#)			
8-months prior to the experiment	9.83 (10.48)	10.56 (10.79)	10.55 (10.80)
8-months of the experiment	9.05 (10.35)	10.50 (11.02)	10.50 (11.06)
8-months following the experiment	7.69 (9.72)	8.76 (10.03)	8.76 (9.97)
Difference between pre and post periods			
Hotel nights	−6.14*	−5.28**	−5.36**
Revenues (\$)	−403*	−343**	−318**
Stays (\$)	−2.14*	−1.81**	−1.79**

Note. There is a significant difference ($p < 0.05$) between the numbers with * and those with **.

Table 1 summarizes the consumption patterns for the control versus treatment groups in the before, within, and post-experiment period. Our first analysis was a difference-in-differences regression of hotel nights. The dependent variable was the hotel nights in the eight months following the experiment minus the nights in the eight months before the experiment. We found that the control group reduced hotel stays by an average of six nights after the experiment period, whereas customers in the treatment groups exhibited only a 5-night reduction, a significant difference ($p < 0.01$). We did not find significant differences between the low and high threshold treatments.

These results point to an initial conclusion that the loyalty promotion campaign has a positive enduring impact on customer consumption. However, Table 2 reports that only 27% of customers attained the goal in the low threshold group and only 21% did so in the high threshold group. Some prior research suggests negative consequences of goal failure for subsequent engagement. We therefore next attempted to separate the enduring effect of goal achievement versus failure within this loyalty program.

Table 2 Summary on Hotel Nights by Reward Success and Failure

	Low threshold		High threshold	
	Reward success	Reward failure	Reward success	Reward failure
No. of customers	6,700	17,382	5,086	19,102
% of customers	28	72	21	79
Hotel nights				
8-months pre	19.76 (24.64)	24.12 (29.31)	16.47 (19.70)	24.46 (28.03)
8-months within	37.77 (41.14)	15.81 (20.68)	36.99 (36.07)	17.49 (21.75)
8-months post	25.80 (33.62)	14.47 (21.27)	24.92 (35.93)	15.42 (22.20)
Revenues (\$)				
8-months pre	2,226 (2,462)	2,180 (4,948)	2,002 (2,712)	2,199 (2,683)
8-months within	3,610 (3,841)	1,532 (2,190)	3,540 (3,575)	1,701 (2,235)
8-months post	2,666 (3,386)	1,535 (2,405)	2,614 (4,320)	1,633 (2,362)
Stays (#)				
8-months pre	11.05 (10.05)	10.38 (11.06)	9.61 (8.94)	10.80 (11.23)
8-months within	16.28 (12.02)	8.27 (9.74)	15.70 (11.82)	9.12 (10.42)
8-months post	12.12 (10.90)	7.46 (9.36)	11.45 (10.50)	8.04 (9.70)
Difference between pre- and post-experiment periods				
	Control group	Reward success	Reward failure	
No. of customers	47,257	24,082	24,188	
Hotel nights	−6.14*	7.08**	−9.33***	
Revenues (\$)	−403*	515**	−604***	
Stays (#)	−2.14*	1.40**	−2.83***	

Note. There is a significant difference ($p < 0.05$) between the numbers of *, **, and ***.

3.1. Propensity Score Analysis

A complicating factor in our analysis is that, in addition to the effects of the experiment, we are primarily interested in the long-term effects of hitting versus missing the reward. The issue is that our variable of interest (success or failure) is selectively determined by consumers' behavior rather than being a direct result of experimental manipulation. The before-and-after experimental design and difference-in-differences approach helps to control for customer heterogeneity. To further address self-selection concerns in this context we use a propensity scoring procedure.

Propensity scoring is designed to help us understand how pre-experimental behaviors can be used to predict reward success. This information can then be applied to the control group to predict the likelihood that a control group member would have earned a reward if placed in the treatment. This estimate is then used in an analysis of post-reward behavior. This analysis is designed to control for the self-selection that occurs in terms of customer response to treatment conditions.

Our propensity-based analysis has three steps. The first step involves a logistic regression of reward achievement using customers in the treatment conditions. In this model we regress reward success against customer activity levels and loyalty status observed in the eight months before the experiment. The second step uses the coefficients from the logistic regression to score the members of the control group. These propensity scores define the likelihood that each customer

would achieve a loyalty-based reward if included in the treatment. The third step involves econometric analyses of post-program activity levels. The propensity scores are included as explanatory variables to account for self-selection driven consumption effects due to the experimental treatments.

Table 3 reports the results of the first step logistic regression. The logistic regression uses a binary indicator of reward attainment as the dependent variable. Hotel nights in the pre-period and dummy variables for Gold and Platinum status are used as the key explanatory factors. The complete model includes these primary factors and interactions between pre-period nights and loyalty status and an estimate of each customer's trend in terms of stays during the pre-period. Each customer's trend is estimated with a Bayesian regression for the eight months of the pre-period experiment. This variable is intended to pick up natural trends in use rates. Key results from the model include significant positive coefficients associated with night stays in the pre-period and positive significant effects for trend (individual time trend slope) and trend squared. These results are especially salient as they capture the relationship between natural (exogenous) upward trends in demand and reward achievement.

We use the coefficients from Model 3 to score each member of the control group. We then construct two samples for further analysis. The first sample includes the control group and members of the treatment that earned a reward. The second sample includes the

Table 3 Estimates of Propensity Score of Reward Success vs. Failure

	Model 1	Model 2	Model 3
Intercept	–1.417 (0.028)	–1.673 (0.044)	–1.725 (0.046)
Member status (gold)	0.625 (0.033)	0.936 (0.054)	0.896 (0.055)
Member status (platinum)	1.106 (0.036)	1.373 (0.052)	1.289 (0.056)
Hotel nights	–0.022 (0.001)	0.024 (0.006)	0.031 (0.006)
Member status (gold) × Hotel nights	/	–0.051 (0.006)	–0.058 (0.007)
Member status (platinum) × Hotel nights	/	–0.047 (0.006)	–0.056 (0.006)
Each customer's hotel night trend	/	/	3.378 (0.161)
Each customer's hotel night trend ²	/	/	29.862 (1.260)
Log likelihood	–26,065	–26,021	–25,720
No. of observations	48,270	48,270	48,270

Notes. (a) Results from a logistic regression with “reward success = 1” versus “reward success = 0” as dependent variable; (b) Only customers who were treated with the loyalty promotion experiment are included in this analysis; (c) Numbers in bold have the significance level < 0.05; (d) The member status refers to the member status at the beginning of the promotion campaign; (e) The hotel nights refer to the 8-month hotel nights before the experiment; (f) Each customer's hotel night trend was estimated using the 8-month hotel nights data with a Bayesian regression on time trend.

control group and the members of the experiment that failed to earn a reward. We then executed several analyses using the two populations. Specifically, we conducted difference-in-differences analyses of hotel nights, revenues, and stays for the post- versus pre-periods. Tables 4 and 5 report the results of these analyses for specifications that include the propensity score or not, for the success and failure plus control populations.

Table 4 shows results for the combined reward success and control group sample. The propensity score acts as a control for the likelihood that the consumer's natural activity level would have earned a loyalty-based reward even in the absence of the treatment. A key comparison across the models that include and do not include the propensity score is therefore the difference in the “Treatment” effects. Given the nature of the sample, being a member of the treatment group for the population analyzed in Table 4 is equivalent to having earned a reward. In the models that include the propensity score the treatment coefficients are smaller for each dependent measure than the values in the models without the scores. For example, the estimated impact of reward achievement is 8.9 incremental nights over the 8-month post-experiment observation period. By comparison, in the model that does not include the propensity score the estimate is 12.9 nights. Controlling for the propensity of consumers to naturally expand consumption levels allows for the decomposition of growth due to reward success versus exogenous change in demand. In terms of other metrics, we also estimate that reward success leads to an incremental \$765 in revenue and about three stays during the 8-month post period.

Table 5 repeats the analysis for the combined control and reward failure populations. In this case controlling for propensity to earn rewards helps us separate natural changes in consumption levels that lead to underestimation of reward failure effects. The difference-in-differences analysis with propensity scores indicates that reward failure leads to a reduction of 5.5 nights and \$377 in revenue over the 8-month post period.

Table 4 Effect of Loyalty Program of Reward Success

	Post and pre difference (NOT controlling for selection)			Post and pre difference (controlling for selection)		
	Nights	Revenue (\$)	Stays	Nights	Revenue (\$)	Stays
Intercept	–6.140 (0.118)	–403.00 (11.99)	–2.136 (0.035)	–22.835 (0.289)	–888.69 (30.41)	–3.333 (0.090)
Treatment	12.870 (0.445)	879.94 (45.37)	3.148 (0.134)	8.913 (0.435)	764.82 (124.02)	2.864 (0.135)
Propensity score	/	/	/	74.059 (1.180)	2,155 (124.02)	5.314 (0.366)
High hurdle	2.426 (0.475)	173.22 (48.49)	0.770 (0.143)	1.853 (0.460)	156.54 (48.38)	0.729 (0.143)
Medium incentive	–0.965 (0.545)	–56.53 (55.67)	0.154 (0.164)	–0.739 (0.528)	–49.97 (55.53)	0.170 (0.164)
High incentive	–1.388 (0.601)	–66.06 (61.30)	0.006 (0.181)	–1.262 (0.582)	–62.41 (61.15)	0.015 (0.180)
Adjusted R-square (%)	4.15	1.96	3.32	10.14	2.45	3.66
Observations		59,043			59,043	

Note. Numbers in bold are significant at < 0.05 level; only customers in the control group (47,258) and those with reward success (11,786) are included in the analysis.

Table 5 Effect of Loyalty Program of Reward Failure

	Post and pre difference (NOT controlling for selection)			Post and pre difference (controlling for selection)		
	Nights	Revenue (\$)	Stays	Nights	Revenue (\$)	Stays
Intercept	-6.140 (0.113)	-0.403 (13.93)	-2.136 (0.035)	-27.659 (0.236)	-1,191 (30.76)	-4.002 (0.077)
Treatment	-4.859 (0.269)	-352.44 (33.19)	-1.117 (0.083)	-5.527 (0.253)	-376.89 (33.04)	-1.175 (0.083)
Propensity score	/	/	/	95.459 (0.934)	3,494.02 (121.82)	8.281 (0.305)
High hurdle	0.614 (0.257)	79.55 (31.75)	0.158 (0.079)	0.480 (0.242)	74.63 (31.60)	0.146 (0.079)
Medium incentive	2.869 (0.297)	230.57 (36.73)	0.735 (0.091)	2.810 (0.280)	228.40 (36.55)	0.730 (0.092)
High incentive	1.377 (0.327)	117.14 (40.40)	0.323 (0.101)	1.760 (0.308)	131.16 (40.21)	0.356 (0.101)
Adjusted <i>R</i> -square (%)	0.52	0.15	0.28	11.56	1.13	1.15
Observations		83,741			59,043	

Note. Numbers in bold are significant at < 0.05 level; only customers in the control group (47,258) and those with reward failure (36,484) are included in the analysis.

3.2. Model-Based Results

The preceding results highlight the differential impact of goal achievement and failure on subsequent consumer purchasing. In this section we empirically test a behavioral theory of relationship-based reciprocity. Specifically, we address the question of whether and how customers in different relationship tiers react differently to goal achievement and failure.

For this test we use a subsample of the customer population. Specifically, we limit the sample to customers that remain in the same loyalty segment during the data collection period. The logic of this approach is that changes in relationship status tiers can complicate interpretation of the results. For example, a customer who achieves a goal and simultaneously improves her loyalty status presents an empirical challenge since any incremental stays might be attributable to goal achievement, tier improvement or an interaction between the two events. Note that restricting the sample in the preceding manner results in a more conservative test. Because achievement is correlated with status improvement and failure with status decline, this sample removes a possible source of bias that would tend to overpredict our hypotheses. This results in more conservative tests of whether customers in a high tier relationship would be more affected by goal failure than those in a low tier relationship, and whether customers in a low tier relationship would be more affected by goal achievement than those in a high tier relationship.

Given that the experiment occurs over an 8-month period and that we have data on member's activities for the eight months preceding and following the experiment, we leveraged the panel structure of the data

to control for observed and unobserved heterogeneity in the monthly hotel night consumption. We estimated a random-effect Tobit model using monthly hotel night consumption as the dependent variable. The Tobit model is used since hotel night consumption is left censored at zero. We assume that there is a latent variable N_{it}^* that is a function of the observed monthly hotel night consumption, N_{it}

$$N_{it} = \begin{cases} N_{it}^* & \text{if } N_{it}^* > 0 \\ 0 & \text{elsewhere.} \end{cases} \quad (1)$$

Term N_{it}^* follows as:

$$\begin{aligned} N_{it}^* = & \beta_0 + \beta_1 Gold_i + \beta_2 Pltn_i + \beta_3 Hit_{it} + \beta_4 Hit_{it} \times Gold_i \\ & + \beta_5 Hit_{it} \times Pltn_i + \beta_6 Hit_{it} \times Hit_{it} \times Gold_i \\ & + \beta_8 Hit_{it} \times Pltn_i + \beta_9 Miss_{it} + \beta_{10} Miss_{it} \times Gold_i \\ & + \beta_{11} Miss_{it} \times Pltn_i + \beta_{12} Miss_{it} \times Hit_{it} \times Gold_i \\ & + \beta_{14} Miss_{it} \times Pltn_i + \beta_{15} IncentM_i \\ & + \beta_{16} IncentH_i + \beta_{17} LogPrice_{it} + \beta_{18} T_{1t} + \beta_{19} T_{2t} \\ & + \beta_{20} T_{2t} \times HurdleL_i + \beta_{21} T_{2t} \times HurdleH_i + \beta_{22} T_{2t} \\ & \times IncentM_i + \beta_{23} T_{2t} \times IncentH_i + \beta_{24} GoalDist_{it} \\ & + \beta_{25} TierDist_{it} + u_i + \varepsilon_{it}, \end{aligned} \quad (2)$$

where β_0 represents the average monthly hotel nights for customers in the control group in the eight months following the experiment; $Gold_i$ and $Pltn_i$ are indicators of program status. The three terms Hit_{it} , $Hit_{it} \times Gold_i$, and $Hit_{it} \times Pltn_i$ capture the difference in average monthly hotel nights between goal-achieving customers and those in the control group for the baseline (Club), Gold,

and Platinum members, respectively. The next three terms, $HitH_{it}$, $HitH_{it} \times Gold_i$, and $HitH_{it} \times Pltn_i$ capture the difference between the low and high hurdle goal-achievement customers. Similarly, the three terms, $Miss_{it}$, $Miss_{it} \times Gold_i$, and $Miss_{it} \times Pltn_i$ capture the difference in average monthly hotel nights between goal-failure customers and those in the control group for the baseline (Club), Gold, and Platinum members, respectively. The $MissH_{it}$, $MissH_{it} \times Gold_i$, and $MissH_{it} \times Pltn_i$ terms account for the potential differences between the low and high hurdle goal-failure customers. The effect of incentive levels in the post-experiment period are accounted for with the two terms $IncentM_i$ and $IncentH_i$ (medium and high incentives). We also include $LogPrice_{it}$, which is a weighted average of monthly hotel prices. The weights are individual-specific based on each customer's share of nights across all of the hotels. In addition, we include T_{1t} to indicate the eight months before the experiment, and T_{2t} as the eight months of the experiment.

We also include terms that reflect the dynamic nature of the reward program since both the experiment

rewards and loyalty status eligibility require customers to achieve a certain number of hotel nights within a limited time window. We include $GoalDist_{it}$ and $TierDist_{it}$ to capture separate types of points pressure. The two terms are measured on a percentage scale to indicate the distance towards the loyalty promotion goal or the next level loyalty status. We also include a random effect u_i to capture unobserved individual-specific heterogeneity. Term ε_{it} is assumed to follow independently identical normal distributions.

Table 6 reports the estimation results. As specified in Equation (2), the results address whether customers in a high tier relationship are more affected by goal failure than those in a low tier relationship, and whether customers in a low tier relationship are more affected by goal achievement than those in a high tier relationship.

There are three patterns worth noting in the results. First, there is a clear distinction in monthly hotel nights for the baseline (Club), Gold, and Platinum customers in the control group in the eight months following the experiment. For example, Platinum customers in the

Table 6 Monthly Hotel Nights Tobit Model Results

	Estimate (SE)	p-value	Diff between low and high hurdle	
Intercept	1.876 (0.311)	<0.001		
Gold _i	2.953 (0.034)	<0.001		
Platinum _i	6.886 (0.032)	<0.001		
Hit Low Hurdle _{it}	1.936 (0.171)	<0.001		
Hit Low Hurdle _{it} × Gold _i	−1.239 (0.187)	<0.001		
Hit Low Hurdle _{it} × Platinum _i	−1.909 (0.177)	<0.001		
Hit High Hurdle _{it}	1.819 (0.181)	<0.001	−0.116 (0.245)	0.635
Hit High Hurdle _{it} × Gold _i	−1.222 (0.201)	<0.001	0.017 (0.270)	0.951
Hit High Hurdle _{it} × Platinum _i	−1.976 (0.189)	<0.001	−0.067 (0.255)	0.793
Miss Low Hurdle _{it}	−0.150 (0.056)	<0.001		
Miss Low Hurdle _{it} × Gold _i	−0.894 (0.071)	<0.001		
Miss Low Hurdle _{it} × Platinum _i	−0.589 (0.065)	<0.001		
Miss High Hurdle _{it}	−0.110 (0.056)	0.049	0.040 (0.068)	0.556
Miss High Hurdle _{it} × Gold _i	−0.760 (0.070)	<0.001	0.134 (0.087)	0.122
Miss High Hurdle _{it} × Platinum _i	−0.457 (0.064)	<0.001	0.132 (0.079)	0.096
LogPrice _{it}	−1.047 (0.063)	<0.001		
GoalDist _{it}	−0.555 (0.029)	<0.001		
TierDist _{it}	−1.505 (0.017)	<0.001		
Medium Incentive _i	−0.106 (0.033)	<0.001		
High Incentive _i	−0.026 (0.033)	0.450		
T _{1t}	1.401 (0.011)	<0.001		
T _{2t}	0.469 (0.014)	<0.001		
T _{2t} × Low Hurdle _i	0.345 (0.029)	<0.001		
T _{2t} × High Hurdle _i	0.362 (0.030)	<0.001		
T _{2t} × Medium Incentive _i	0.106 (0.027)	<0.001		
T _{2t} × High Incentive _i	0.121 (0.030)	<0.001		
Sigma u_i	2.250 (0.008)	<0.001		
Sigma ε_{it}	4.523 (0.004)	<0.001		
No. of customers	70,874			
No. of observations	1,700,976			
Log-likelihood	−3,085,612			

Note. There are 813,981 left-censored observations and 886,995 uncensored observations.

control group stay 6.9 more nights than baseline (Club) members in the control group.

In terms of the enduring effects of reward success, baseline (Club) customers who hit the experiment goals stayed 1.8 or 1.9 more nights per month depending on the threshold level in the eight months following the experiment. Gold members also significantly increase stays following a success but this group's increase is significantly lower than the base members at approximately 0.6 nights per month ($p < 0.01$). By contrast, for Platinum members who hit the threshold there is no positive impact on the monthly hotel nights in the eight months following the experiments.

In terms of reward failures, we find significant negative effects across all loyalty tiers. Baseline customers who missed the experiment goal reduced stays by between 0.1 and 0.15 nights per month depending on the threshold required ($p < 0.01$). This decrease is significantly larger for the goal-failure Gold (0.8 nights per month; $p < 0.01$) and Platinum (0.5 nights per month; $p < 0.01$).

The patterns are consistent with the premise of relationship-based reciprocity: Customers in a high tier relationship are more affected by goal failure and less affected by goal achievement than are customers in a low tier relationship.

The results also show some additional patterns, which are not the main focus of this research. For example, results confirm a behavioral pattern of points pressure documented in Nunes and Dreze (2006) and Kivetz et al. (2006) with the significantly negative coefficients associated with $GoalDist_{it}$ and $TierDist_{it}$. As customers move closer towards a goal, they exert more effort. Furthermore, the fraction of variance due to unobserved individual heterogeneity (random effect) is approximately 20%.

4. Alternative Explanations

In this section we extend the previous model specification to test for the possibility of reinforcement learning. As explicated in the development of H2B, if reinforcement learning is the driver of post-promotion behavior then we would expect to see the following: A reward would cause an increase in demand trajectory while a failure to earn a reward might cause a decrease. However, by contrast to the relationship-oriented explanation proposed, demand would gradually revert to its old trajectory following the event.

To investigate the possibility of reinforcement learning we test whether demand levels gradually revert to previous trajectory following the success or failure event. We do this by adding three terms to the previous panel Tobit specification. In particular we add a time trend that begins after the promotion and interactions between the time trend and binary indicators of reward success (HIT) and failure (MISS).

Table 7 provides estimation results for the model with the three additional terms. For our question of interest the two key terms are the interactions of hitting and missing with the post-program time trend. A reinforcement learning explanation would suggest that the interaction would be negative for the HIT \times Trend interaction and positive for the MISS \times Trend interaction. We find positive coefficients associated with both interactions. This implies that we only see reversion to previous levels following a miss. For the reward achievers the positive coefficient suggests a positive feedback effect. Based on these findings the data does not support H2B. At the same time, we still find patterns similar to those in Table 6, which are consistent with the premise of relationship-based reciprocity in H2A.

We acknowledge that our investigation of reinforcement learning is somewhat different than previous studies. In our case, the focal loyalty promotion does not continue following reward success or failure. However, while the focal promotion ends, the loyalty program itself is ongoing. Our reinforcement conjectures are based on the notion that consumers learn about their capabilities within the short-term promotion and that these learnings drive response to the permanent program.

5. Conclusion

In this paper we examine the post-promotion effects of a loyalty program promotion conducted by a major hotel company. Although such frequent buyer programs are intended to foster customer loyalty and increase metrics such as retention and customer lifetime value, these programs often include reward structures that may have negative effects on subsequent customer purchasing. We hypothesize that successfully earning or failing to achieve a loyalty-based reward will have enduring effects on purchasing, and further speculate that these effects could vary based on the previous nature of the customer relationship.

Although participation in the program led to an overall positive effect on subsequent purchasing, the effects of goal success or goal failure diverge once separated. Over the eight months following the program we estimate that reward success leads to an incremental \$765 in revenues for a customer who earns rewards. However, we also estimate that failing to earn a reward results in a reduction in revenues of about \$375. Given these results, firms need to carefully design loyalty-based promotions to balance success and failure rates.

We also find asymmetric effects of success and failure within loyalty tiers. In terms of success, we find that the positive long-term effects are negatively related to program status. Econometrically, we identify significantly larger long-term positive consequences for

Table 7 Post Promotion Reset Patterns

	Estimate (SE)	p-value	Diff between low and high hurdle	
Intercept	6.770 (0.313)	<0.001		
Gold _i	2.915 (0.034)	<0.001		
Platinum _i	6.811 (0.032)	<0.001		
T _{3t} – Time trend	–0.248 (0.004)	<0.001		
T _{3t} – Time trend × Hit _i	0.037 (0.006)	<0.001		
T _{3t} – Time trend × Miss _i	0.024 (0.004)	<0.001		
Hit Low Hurdle _{it}	1.840 (0.168)	<0.001		
Hit Low Hurdle _{it} × Gold _i	–1.202 (0.184)	<0.001		
Hit Low Hurdle _{it} × Platinum _i	–1.846 (0.175)	<0.001		
Hit High Hurdle _{it}	1.745 (0.179)	<0.001	–0.095 (0.242)	0.694
Hit High Hurdle _{it} × Gold _i	–1.214 (0.198)	<0.001	–0.012 (0.266)	0.964
Hit High Hurdle _{it} × Platinum _i	–1.923 (0.198)	<0.001	–0.077 (0.251)	0.759
Miss Low Hurdle _{it}	–0.179 (0.056)	<0.001		
Miss Low Hurdle _{it} × Gold _i	–0.880 (0.070)	<0.001		
Miss Low Hurdle _{it} × Platinum _i	–0.578 (0.064)	<0.001		
Miss High Hurdle _{it}	–0.142 (0.056)	0.011	0.037 (0.067)	0.575
Miss High Hurdle _{it} × Gold _i	–0.749 (0.069)	<0.001	0.131 (0.086)	0.125
Miss High Hurdle _{it} × Platinum _i	–0.448 (0.063)	<0.001	0.130 (0.078)	0.096
LogPrice _{it}	–1.843 (0.063)	<0.001		
GoalDist _{it}	–0.581 (0.029)	<0.001		
TierDist _{it}	–1.711 (0.017)	<0.001		
Medium Incentive _i	0.108 (0.027)	<0.001		
High Incentive _i	0.121 (0.030)	<0.001		
T _{1t}	0.556 (0.016)	<0.001		
T _{2t}	–0.407 (0.018)	<0.001		
T _{2t} × Low Hurdle _i	0.404 (0.030)	<0.001		
T _{2t} × High Hurdle _i	0.420 (0.031)	0.499		
T _{2t} × Medium Incentive _i	0.108 (0.027)	<0.001		
T _{2t} × High Incentive _i	0.121 (0.030)	<0.001		
Sigma u _i	2.215 (0.008)	<0.001		
Sigma e _{it}	4.517 (0.004)	<0.001		
No. of customers	70,874			
No. of observations	1,700,976			
Log-likelihood	–3,082,845			

Note. There are 813,981 left-censored observations and 886,995 uncensored observations.

base members than for Gold or Platinum members. In fact, we do not observe a significant positive impact of success on Platinum members. By contrast, the negative impact of program failure is significant for all groups but the impact on Gold and Platinum members is significantly more negative than for the base members.

The current study is based on results of a large scale field experiment conducted by a major international hotel company. As such, the results have significant ecological validity. However, the nature of the experiment also involves limitations. Given that the experiment was conducted with actual customers, data collection was purposively nonintrusive and little burden was imposed on subjects. This is relevant given that the academic communities want to go beyond the documentation of enduring behavioral changes to investigate the underlying drivers of the long-term effects.

Future research could be designed to focus more on underlying cognitive processes. In the current study we

are limited to logical inferences based on econometric results. It may be useful to design research that is less constrained by commercial objectives and better situated for collecting data related to underlying consumer decision processes. Collection of psychological measures may be useful for reaching and more precisely determining the mechanisms through which reward success or failure impact customer relationships across different consumer segments.

In sum, the current findings highlight a managerial dilemma related to the type of promotion under study. The promotion itself drives incremental revenue and seems to have positive or negligible effects on base members depending on whether the reward is earned. The promotion, therefore, would seem to be of great value for managing the “base” segment of customers. However, the promotion also presents potential problems for the firm’s most valuable customers when there is a possibility that rewards will be withheld. A possible solution is to only offer such programs to

base level consumers. That said, however, the dilemma exists that excluding the highest level program tiers from a promotion might have an adverse effect on the firm's relationships with its most valuable customers. Future investigations should explore ways that firms can tailor loyalty promotions to meet the diverse needs of multiple customer segments.

Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/mksc.2015.0966>.

Acknowledgments

The authors are grateful to the editor, associate editor, and two anonymous reviewers for their helpful comments.

References

- Anderson E, Oliver RL (1987) Perspectives on behavior-based versus outcome-based salesforce control systems. *J. Marketing* 51(4): 76–88.
- Badovick GJ, Hadaway FJ, Kaminski PF (1992) Attributions and emotions: The effects on salesperson motivation after successful vs. unsuccessful quota performance. *J. Personal Selling Sales Management* 12(3):1–11.
- Blattberg RC, Neslin SA (1990) *Sales Promotion: Concepts, Methods and Strategies* (Prentice-Hall, Englewood Cliffs, NJ).
- Breugelmans E, Bijmolt TH, Zhang J, Basso LJ, Dorotic M, Kopalle P, Wunderlich NV (2015) Advancing research on loyalty programs: A future research agenda. *Marketing Lett.* 26(2):1–13.
- Brown SP, Evans KR, Mantrala MK, Challagalla G (2005) Adapting motivation, control, and compensation research to a new environment. *J. Personal Selling Sales Management* 25(2):155–167.
- Cochran W, Tesser A (1996) The “what the hell” effect: Some effects of goal proximity and goal framing on performance. Martin LL, Tesser A, eds. *Striving and Feeling: Interactions Among Goals, Affect, and Self-Regulation* (Erlbaum, Mahwah, NJ), 99–120.
- Deci EL, Koestner R, Ryan RM (1999) A meta-analysis review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psych. Bull.* 125:627–668.
- Dowling GR, Uncles M (1997) Do customer loyalty programs really work? *Sloan Management Rev.* 38(4):71–82.
- Dreze X, Nunes JC (2009) Feeling superior: The impact of loyalty program structure on consumers' perceptions of status. *J. Consumer Res.* 35(6):890–950.
- Dreze X, Nunes JC (2011) Recurring goals and learning: The impact of successful reward attainment on purchase behavior. *J. Marketing Res.* 48(2):268–281.
- Fehr E, Gächter S (2000) Fairness and retaliation: The economics of reciprocity. *J. Econom. Perspect.* 14(3):159–181.
- Gollwitzer PM (1999) Implementation intentions: Strong effects of simple plans. *Amer. Psych.* 54(7):493–503.
- Guttman N (1953) Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. Experiment. Psych.* 46(4):213–224.
- Heath C, Richard PL, George W (1999) Goals as reference points. *Cognitive Psych.* 38(1):79–109.
- Jhang JH, Lynch JG (2015) Pardon the interruption: Goal proximity, perceived spare time, and impatience. *J. Consumer Res.* 41(5): 1267–1283.
- Kivetz R, Simonson I (2002) Earning the right to indulge: Effort as a determinant of customer preference. *J. Marketing Res.* 39(2): 155–170.
- Kivetz R, Urminsky O, Zheng Y (2006) The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *J. Marketing Res.* 43(1):39–58.
- Kopalle P, Sun Y, Neslin S, Sun B, Swaminathan V (2012) The joint impact of frequency reward and customer tier components of loyalty programs. *Marketing Sci.* 31(2):216–235.
- Kumar VK, Shah D (2004) Building and sustaining profitable customer loyalty for the 21st century. *J. Retailing* 80(4):317–329.
- Lewis M (2004) The influence of loyalty programs and short-term promotions on customer retention. *J. Marketing Res.* 41(3): 281–292.
- Liu Y (2007) The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *J. Marketing* 71(3):19–35.
- Locke EA, Latham GP (1990) *A Theory of Goal Setting and Task Performance* (Prentice-Hall, Englewood Cliffs, NJ).
- Montaguti E, Neslin SA, Valentini S (2016) Can marketing campaigns induce multichannel buying and more profitable customers? A field experiment. *Marketing Sci.* 35(2):201–217.
- Nunes JC, Dreze X (2006) The endowed progress effect: How artificial advancement increases effort. *J. Consumer Res.* 32(4): 504–512.
- Oliver RL, Anderson E (1994) An empirical test of the consequences of behavior- and outcome-based sales control systems. *J. Marketing* 58(4):53–67.
- Roehm M, Ellen BP, Harper AR (2002) Designing loyalty-building programs for packaged goods brands. *J. Marketing Res.* 39(2): 202–213.
- Schweperker CH, Good DJ (1999) The impact of sales quotas on moral judgment in the financial services industry. *J. Services Marketing* 13(1):38–58.
- Skinner BF (1938) *The Behavior of Organisms* (Appleton-Century-Crofts, New York).
- Skinner BF (1969) Contingency management in the classroom. *Education* 90(2):93–100.
- Soman D, Cheema A (2004) When goals are counterproductive: The effects of violation of a behavioral goal on subsequent performance. *J. Consumer Res.* 31(1):52–62.
- Taylor GA, Neslin SA (2005) The current and future sales impact of a retail frequency reward program. *J. Retailing* 81(4):293–305.
- Tucker C, Zhang J (2010) Growing two-sided networks by advertising the user base: A field experiment. *Marketing Sci.* 29(5): 805–814.